



Journal of Undergraduate Research International

Volume 2, Issue 1, March 2026
ISSN 2961-4473



Journal of Undergraduate Research International

Volume 2, Issue 1, March 2026

Research Articles

-  **Optimal Design of a Setpoint Filter and Fractional-Order PID Controller for Unstable Launch Vehicles Using the Zebra Optimization Algorithm**

Yazan Zwawi and Fadi Alyoussef
(Aerospace Engineering)

1
-  **Search-and-Track Missions with Limited Communication in Unmanned Aerial Vehicle Swarms: A Centralized Leader – Follower Study**

Jinan Ali Shajiri and Mohammed Basheer Mohiuddin
(Aerospace Engineering/Control and Instrument Engineering)

8
-  **Embedded Kalman Filter-Based GNSS/INS Fusion for Robust Planetary Rover Navigation**

Ahmed Bashammakh and Atif Mahmood
(Aerospace Engineering/Control and Instrument Engineering)

14
-  **Fractional-Order Proportional-Integral-Derivative Control for Stable Close Proximity Flight in Multirotor Unmanned Aerial Vehicle Swarms**

Jassim Mohammed Alsayegh and Ghulam E Mustafa Abro
(Aerospace Engineering/Control and Instrument Engineering)

20
-  **Design and Comparative Analysis of a Ka-band Microstrip Patch Antenna Array for Deep-Space Exploration: Rovers Communications**

Hussah Abdulrahman Aldawish and Bazilah Baharom
(Aerospace Engineering/ Physics)

29
-  **A hydrodynamics quantification of tortuosity and residence time of TPMS based bone scaffold**

Btoul S. Alkhalil and Ahmed S. Dalaq
(Bioengineering)

36
-  **Hydrothermally Synthesized Ni/Co-Doped MoS₂/MoO₃ Nanosheets for NO₂ Sensing: Experimental and DFT Insights**

Zahra Ali Alharbi, Retno Dwi Astuti, Md Saif Hasan, and Md Delowar Hossain
(Chemistry/Materials Science and Engineering)

45
-  **Design and Characterization of Green Cementitious Composites for Structural Health Monitoring**

Mohammed Al Abdulmohsin and Asad Hanif
(Civil Engineering)

58
-  **Arabic News Dataset Collection and Generation for Deepfake-Text Detection**

Khulud Alotaibi, Amal Sunba, and Tarek Helmy
(Computer Science)

67
-  **Arabic Review Dataset for Deepfake Text Detection: Collection and Generation**

Malath Alhashem, Amal Sunba, and Tarek Helmy
(Computer Science)

76

- Tumor Classification on the Bone Tumor X-ray Radiograph Dataset : Cross Validation, Contrast-Limited Adaptive Histogram Equalization Ablation, and Explainability Assessment** 84
Saad Bin Waqas and Muhammad Shahid Jabbar
 (Computer Science)
- Fuzzy-Based Sliding Mode Control for Quadrotor Trajectory Tracking** 95
Mansour Abdullah Nasrullah and Syed Muhammad Amrr
 (Control and Instrument Engineering)
- Resource-Constrained Event Scheduling: An Optimization Approach** 105
Shimaa Alwadai and Yasser Almoghathawi
 (Industrial Engineering)
- Blocking Humidity Ingression in FAPbI₃ Thin Film for Perovskite Solar Cell Application** 113
Ghadah Khalid Mobaireek, Muhammed P.U. Haris, and Amir Al-Ahmed
 (Material Science and Engineering)
- Low-Temperature Electrowinning of Iron from Hematite Suspension in Sodium Hydroxide Solution: Toward Sustainable Ironmaking** 119
Mohammed A. Bahamdan and Nestor K. Ankah
 (Material science and Engineering)
- CO₂ Capture using Dual-Amine Functionalized MCM-41 Nanoparticles Synthesized via Silane Condensation** 127
Maymounah Alshahrani and Ahmed Wasel A Alsmail
 (Material science and Engineering)
- Platinum-loaded Molybdenum Phosphide Dispersed on Carbon Nanotubes for Improved Hydrogen Evolution Performance** 135
Abdullah Alismail and Mohammad Qamar
 (Material science and Engineering)
- Hybrid Statistical-Machine Learning Framework for Evaluating Drivers of Power Sector Decarbonization** 141
Maryam Y. Alkhazraj and Mohamed Elsamadony
 (Mechanical Engineering)
- Elastic Property Prediction of Stochastic Gyroid TPMS Lattices Using Simulation based Machine Learning** 152
Abdullah Alduaydi and Ammar Alsheghri
 (Mechanical Engineering)
- Influence of Ba-Site Substitution on the Structural Properties of TI-2223 Superconductor** 163
Hawra Tuhaifa and Khalil A. Ziq
 (Physics/Materials Science and Engineering)

Optimal Design of a Setpoint Filter and Fractional-Order PID Controller for Unstable Launch Vehicles Using the Zebra Optimization Algorithm

Yazan Zwawi and Fadi Alyoussef*

Cite <https://doi.org/10.64589/juri/218954>

Submitted: November 18, 2025 Revised: February 16, 2026 Accepted: February 23, 2026

ABSTRACT

Attitude control of an unstable launch vehicle (ULV) is crucial for maintaining an accurate spatial orientation and ensuring a precise location for communication, maneuvering, and scientific research. Classical proportional-integral-derivative (PID) controllers are frequently used to control spacecraft attitude owing to their simplicity. However, their robustness and agility in response to dynamic space conditions and disturbances do not meet the required design standards. To improve the system stability, in this study, we propose a fractional-order PID (FOPID) controller with a setpoint filter, which is optimally tuned using the zebra optimization algorithm (ZOA). The performance of the proposed FOPID-ZOA controller is compared with that of a traditional PID controller. The proposed optimization strategy is applied offline to tune the controller parameters, which remain constant during simulation tests. The simulation results show that the proposed approach outperforms the classical PID in terms of robustness, accuracy, disturbance rejection, and overall performance. Notably, compared with the traditional PID controller, the proposed controller requires a significantly lower control voltage during the transient response. Thus, the peak voltage is reduced from 20 V to 1 V (or approximately 95% less), which in turn reduces energy consumption and improves system stability. Therefore, the proposed approach represents a promising solution for systems exhibiting ULV-like dynamics, particularly where energy efficiency is a primary concern.

Keywords: attitude control of unstable launch vehicle, PID, fractional-order PID, zebra optimization algorithm

1. INTRODUCTION

Solving the control problem of unstable launch vehicles (ULV) is crucial for successful space exploration missions, satellite communications, and general Earth observations. Attitude control confirms that the space vehicle preserves its desired pitch orientation using actuators such as reaction wheels or thrusters, which allow correction of the measured pitch angle. Several important publications have focused on controlling ULV since the 1960s. In such systems, the center of gravity lies behind the neutral point, which is an inherent characteristic of ULV¹. Consequently, the angles of attack must be prevented from diverging. In addition, the achievable closed-loop performance is limited by actuator constraints and phase delays¹. Thus, a robust control system is necessary to control the ULV. Classical ULV control design generally linearizes the ULV dynamics around an operating point using the small-perturbation theory. Also, traditional proportional-integral-derivative (PID) controllers are still dominant in attitude ULV control^{1,2}, but they cannot effectively handle nonlinearities, uncertainties, and external disturbances. Additionally, PID controllers generate response with large overshoots in integrating systems³⁻⁶. Notably, the pitch control dynamic of the ULV is an integrating process. A large overshoot implies that

the controller uses more control energy, which can potentially cause instability, considering the physical limitations of the actuators^{7,8}. Therefore, a more robust control strategy for the ULVs should be developed.

Fractional-order PID (FOPID) controllers are robust alternatives to classical PID controllers⁶. FOPID increases the controller degree of freedom by adding the fractional orders λ and μ as extra design parameters. This imparts engineers the capability to select parameters for improved stability margins. The right tuning of FOPID parameters reduces overshoot and decreases control effort, helping avoid actuator limitations. FOPID handles PID's disadvantage for higher-order plants where PID increases unmodeled dynamics. FOPID is particularly useful for unstable systems with long time delays due to its flexibility. These features confirm that compared with standard PID, FOPID offers robust performance and better balance among disturbance elimination, noise mitigation, and overshoot^{4,6,9-11}.

Optimization algorithms are used to compute system parameters as well as to improve system stability¹²⁻¹⁴. Thus, in this study, we suggest using an FOPID controller, tuned by the zebra optimization algorithm (ZOA), to improve the ULV attitude control accuracy and stability. The limited utilization of fractional-order

controllers in ULV systems, which can be identified by integrating unstable dynamics, is the primary research requirement of this study. The control design of a ULV is challenging owing to these characteristics. Therefore, a robust control solution was required. The contributions of this study are as follows:

1. The classical PID design methods for ULV reported in^{1,2} has a large overshoot and slow disturbance rejection. However, the proposed FOPID with a setpoint filter handles these advantages, reduces overshoot, and improves disturbance rejection. Also, the proposed method is simple and can be easily implemented in real time.
2. The ZOA was used to optimally adjust the proposed FOPID and setpoint filter gains to reduce the control engineer's time allocated for the tuning process.

The remainder of this paper is structured as follows. The attitude dynamics model of the ULV is presented in Section 2; the suggested controller is described in Section 3; the outcomes of the simulation and performance evaluation are discussed in Section 4; and the conclusions are presented in Section 5.

2. UNSTABLE ATTITUDE DYNAMICS MODEL OF UNSTABLE LAUNCH VEHICLE

The behavior of engineering systems (Figure 1), including unmanned aerial vehicles, satellite attitude control systems, and missile guidance systems, can be represented as integrating systems with an integrator in their dynamics. This control problem provides a benchmark for numerous real-world applications and is of great practical interest. From theoretical and industrial perspectives, investigating reliable control mechanisms for such systems is crucial. The proposed control approach can be applied to other systems with comparable dynamic structures, demonstrating its wider applicability for various control applications.

In this study, the ULV system is considered an integrating system and is inherently unstable. The small-perturbation theory was used to model the ULV dynamics. The model is represented

by the following set of equations¹:

$$\begin{aligned}\frac{d^2\theta}{dt^2} &= m_q \frac{d\theta}{dt} + m_\alpha + m_{\delta_e} \delta_e \\ \frac{d\gamma}{dt} &= c_\alpha \alpha + c_{\delta_e} \delta_e \\ \theta &= \gamma + \alpha \\ q &= \frac{d\theta}{dt}\end{aligned}\quad (1)$$

where α , γ , θ , and q denote the angle of attack, flight-path angle, pitch angle, and pitch rate, respectively, and δ_e is the elevator deflection. The pitching-moment derivatives m_q , m_α , and m_{δ_e} , and the lift slopes c_α and c_{δ_e} characterize the vehicle's stability and control effectiveness; their values are obtained from aerodynamic analysis or system identification, such that the dynamics of the launch vehicle are captured over the operating points. The elevator is driven by a control voltage δ through an actuator that is modeled as a pure time delay:

$$\delta_e = \delta e^{-Ls} \quad (2)$$

where s is the Laplace variable, and L is the delay constant reflecting the actuator bandwidth and other lags. L has a pronounced impact on both the transient response and achievable control performance. Combining equation (1) with equation (2) and performing the standard linearization and algebraic reduction yields the longitudinal (pitch-axis) transfer function of the statically unstable launch vehicle, as presented below¹:

$$\frac{\theta(s)}{\delta(s)} = \frac{k_m (T_n s + 1)}{s(\tau s - 1)(T_d s + 1)} e^{-Ls} \quad (3)$$

where k_m is the static gain, T_n and T_d are lead-lag time constants, and τ denotes the time constant corresponding to the unstable aerodynamic pole. Equation (3) shows a mathematical model with an integrator and an unstable pole, increasing the difficulty of control design^{5,15-17}.

3. PROPOSED SETPOINT FILTER AND FRACTIONAL PID CONTROLLER

The proposed FOPID, along with the setpoint filter, is mathematically defined as follows:

$$G(s)_{\text{FOPID}} = k_p + \frac{k_i}{s^\lambda} + k_d s^\mu \quad (4)$$

$$G(s)_{\text{setpoint-filter}} = \frac{1}{k_f s + 1} \quad (5)$$

where k_p , k_i , k_d , and k_f are the proportional, integral, derivative, and filter gains, respectively; these are positive design parameters. λ and μ are the fractional-order terms in the range $[0, 2]$. Figure 2 illustrates the proposed controller.

In this study, the parameters k_p , k_i , k_d , λ , μ , and k_f are computed using ZOA, which will be explained in the next subsection. The optimization algorithm was executed offline, and the resulting optimal parameters were stored and used directly in the controller without real-time updating.



Figure 1. Examples of integrating systems

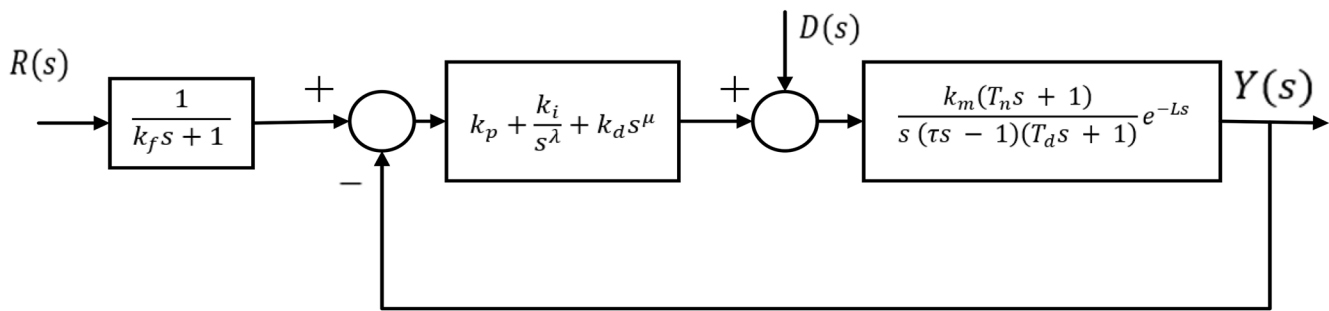


Figure 2. Proposed controller

4. ZEBRA OPTIMIZATION ALGORITHM

Zebras are native to eastern and southern Africa and are commonly classified into three species: Grevy's, plain, and mountain zebra. Plain zebras usually inhabit grassy ecosystems; mountain zebras are found in mountainous and hilly regions, while Grevy's zebra occupy dry grasslands and deserts¹⁸. Zebras feed on grass, branches, leaves, bark, and other plant material and usually live in herds. They are protected from flies and predators by their characteristic black-and-white striped patterns, mostly distributed along the body and neck. Zebras exhibit strong defensive behaviors to protect themselves from predators when they feel threatened. Their behavior can be broadly divided into two stages: foraging and protective responses to predators. Herd movement is usually guided by a lead zebra¹⁹.

Lead zebras typically direct herding movements. Each "zebra" in the swarm-based ZOA is an autonomous search agent traversing a virtual plain, which is the problem's decision space. A full candidate solution or a particular assignment for every decision variable is encoded by the location of the zebra. The optimal population comprises the herd as a whole. To promote wide exploration, the zebras were initially randomly dispersed inside the permitted boundaries. Equation 6 specifies the ZOA population matrix¹⁹.

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix}_{N \times M} = \begin{bmatrix} x_{(1,1)} & \dots & x_{(1,j)} & \dots & x_{(1,m)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{(i,1)} & \dots & x_{(i,j)} & \dots & x_{(i,m)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{(N,1)} & \dots & x_{(N,j)} & \dots & x_{(N,m)} \end{bmatrix}_{N \times M} \quad (6)$$

Here, the population is the full herd of candidate solutions. Zebra i denotes the i -th member of the herd, and its list of entries stores the proposed values for m decision variables (the j -th entry corresponds to variable j). The herd contains N zebras. Because each zebra encodes a complete solution, the objective (fitness) value can be computed for each zebra from its proposed variable set. The resulting N objective values are collected into a vector, as summarized in equation (2)¹⁹:

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N \times 1} = \begin{bmatrix} F_1(X_1) \\ \vdots \\ F_i(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{N \times 1} \quad (7)$$

Zebras travel to obtain their favorite foods, engage in social interactions, and congregate to sleep at night. They spent 60–80% of their time feeding¹⁸. Each zebra modifies its position as a pioneer zebra (PZ) leads the herd toward its destination. The following equation mathematically defines the foraging phase¹⁹:

$$x_{(ij)}^{new,P_1} = x_{(ij)} + r \left(PZ_j - (\text{round}(1 + \text{rand})) x_{(ij)} \right) \quad (8)$$

$$X_i = \begin{cases} X_i^{new,P_1}, & F_i^{new,P_1} < F_i \\ X_i, & \text{else} \end{cases} \quad (9)$$

Here, the pioneer's position influences the random variations that zebras use to explore new positions. A new position is adopted if it increases the function value of the objective; otherwise, the old location is retained. The function $\text{rand}(\bullet)$ generates random numbers uniformly distributed within the interval $[0, 1]$. The notation X_i^{new,P_1} denotes the updated position of the i -th zebra during the foraging phase, and $x_{(ij)}^{new,P_1}$ represents its coordinates in the j -th dimension. The parameter F_i^{new,P_1} corresponds to the objective function value evaluated at the new position.

Zebras are most vulnerable to predation by lions when grazing, although other carnivorous species, such as leopards, hyenas, cheetahs, and wild dogs, can also pose threats. Furthermore, crocodiles present a major threat when zebras are located close to water sources. Zebras use various defense strategies to combat these threats. When a lion attacks, it can decide to run away, whereas it reacts more violently to hyenas and wild dogs. The zebrafish defense pattern is determined using a specific technique¹⁸. To evade lions, zebras often engage in zigzag and irregular lateral movements that are mathematically described by the following equations¹⁹:

$$x_{(ij)}^{new,P_2} = \begin{cases} S_1 : x_{(ij)} + R(2r - 1) \left(1 - \frac{t}{T} \right) x_{(ij)}, & P_s \leq 0.5 \\ S_2 : x_{(ij)} + r \left(AZ_j - (\text{round}(1 + \text{rand})) x_{(ij)} \right), & \text{else} \end{cases} \quad (10)$$

$$X_i = \begin{cases} X_i^{new,P_2}, & F_i^{new,P_2} < F_i \\ X_i, & \text{else} \end{cases} \quad (11)$$

Additionally, in response to an attack, both predators and other zebras may react dynamically to an attack. To scare or confound the predator, other zebras in the group may move in synchronization toward the targeted member, forming protective barriers. This behavior is then used to update the position of the zebra.

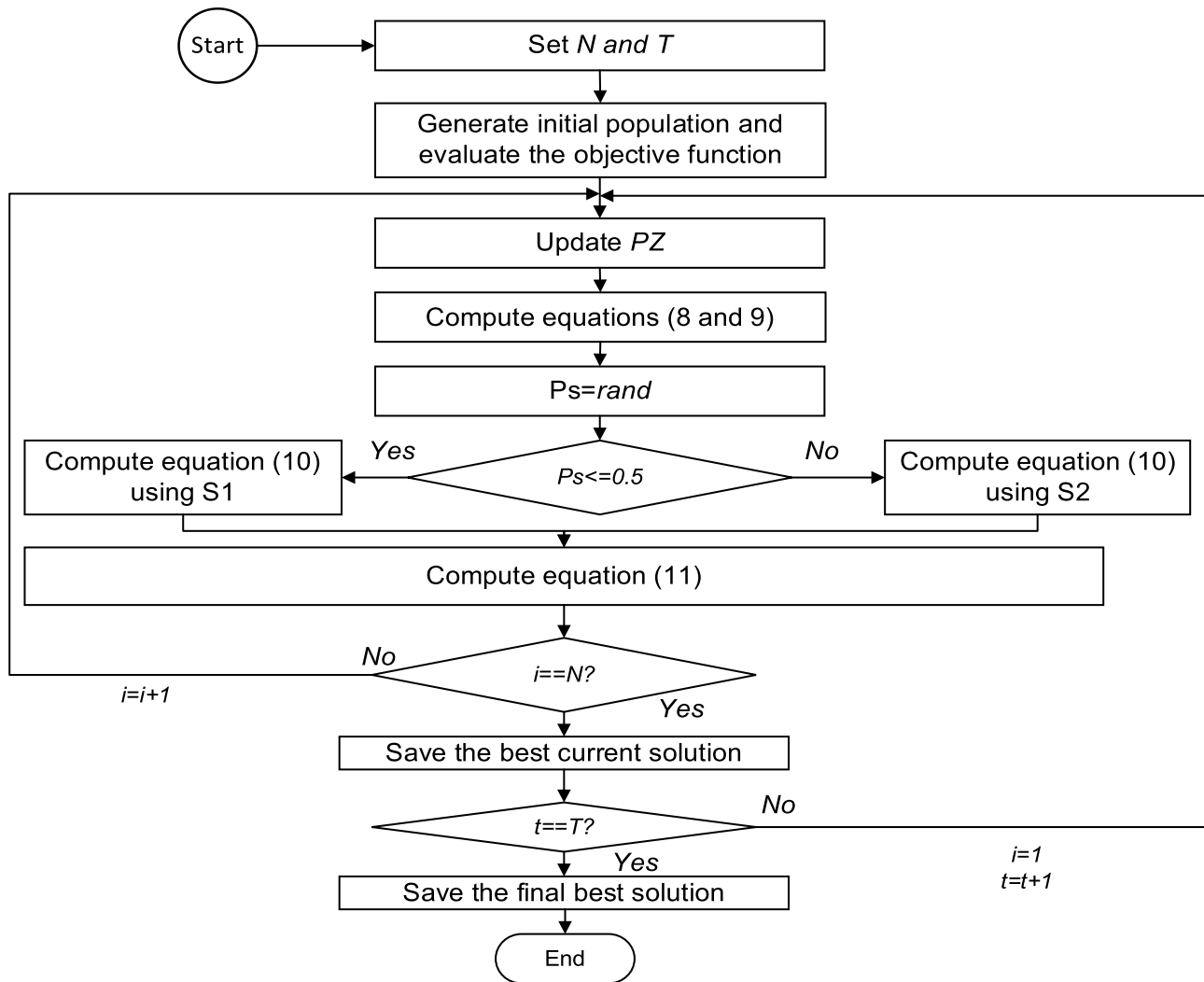


Figure 3. ZOA flowchart

This procedure is a part of a larger optimization mechanism that occurs in two stages during each iteration to continuously move closer to the best solution¹⁸. The flowchart of the ZOA is shown in Figure 3.

5. SIMULATION RESULTS AND DISCUSSION

The following system transfer function for an unstable launch-vehicle attitude is used¹:

$$G(s) = \frac{0.0520(279.74s + 1)}{s(0.7490s - 1)(0.7448s + 1)} e^{-0.05s} \quad (12)$$

The parameters of the proposed controller (k_p , k_i , k_d , λ , μ , and k_f) are computed using ZOA using 30 search agents and 50 iterations. The lower and upper bounds of the parameters are assumed to be $[0.1 \ 0.1 \ 0.1 \ 0.1 \ 0.1 \ 0.1]$ and $[3 \ 3 \ 3 \ 2 \ 2 \ 2]$, respectively. Smaller upper bounds are selected to accelerate the optimization algorithm. Here, integral of squared error (ISE), $ISE = \int_0^\infty e(t)^2 dt$ (See Figure 4), is used as objective function. Figure 5 shows the convergence curve of the ZOA. The final values of the proposed controller are: $k_p = 2.3498$, $k_i = 2.9909$, $k_d = 0.5858$, $\lambda = 0.7981$, $\mu = 1.1513$, and $k_f = 1.1513$. All the

parameters of the proposed algorithm are summarized in Table 1. The performance of the proposed controller is compared with that of the classical PID controller reported in¹. The classical PID parameters are $k_p = 0.8644$, $k_i = 0.2$, and $k_d = 0.2811$ ¹.

Figure 6 compares the tracking and rejection of disturbances between the classical PID controller¹ and the proposed FOPID controller. The plant output is shown in panel (a), and the equivalent control effort is shown in panel (b). At $t = 0$ s, a unit-step reference is applied, and at $t = 15$ s, a disturbance with amplitude 1 is introduced into the closed-loop system. The PID shows a larger overshoot, while the suggested controller generates a nearly monotonic climb with virtually no overshoot. This improvement is accomplished with significantly lower peak actuation values: at $t = 0$ s (zoom-in in (b)), the PID reaches >20 V, whereas the FOPID peaks at just ~ 1 V. In addition to lowering the risk of windup and instabilities from actuator saturation, the decrease in initial control demand maintains an appropriate margin to hardware limits. Compared to PID¹, the FOPID rejects the disturbances with a shorter recovery time after the disturbance at $t = 15$ s (right-hand zoom-in in Figure 6 (b)). The performance is improved without significant control spikes or increased actuation demand. The approach is suitable for

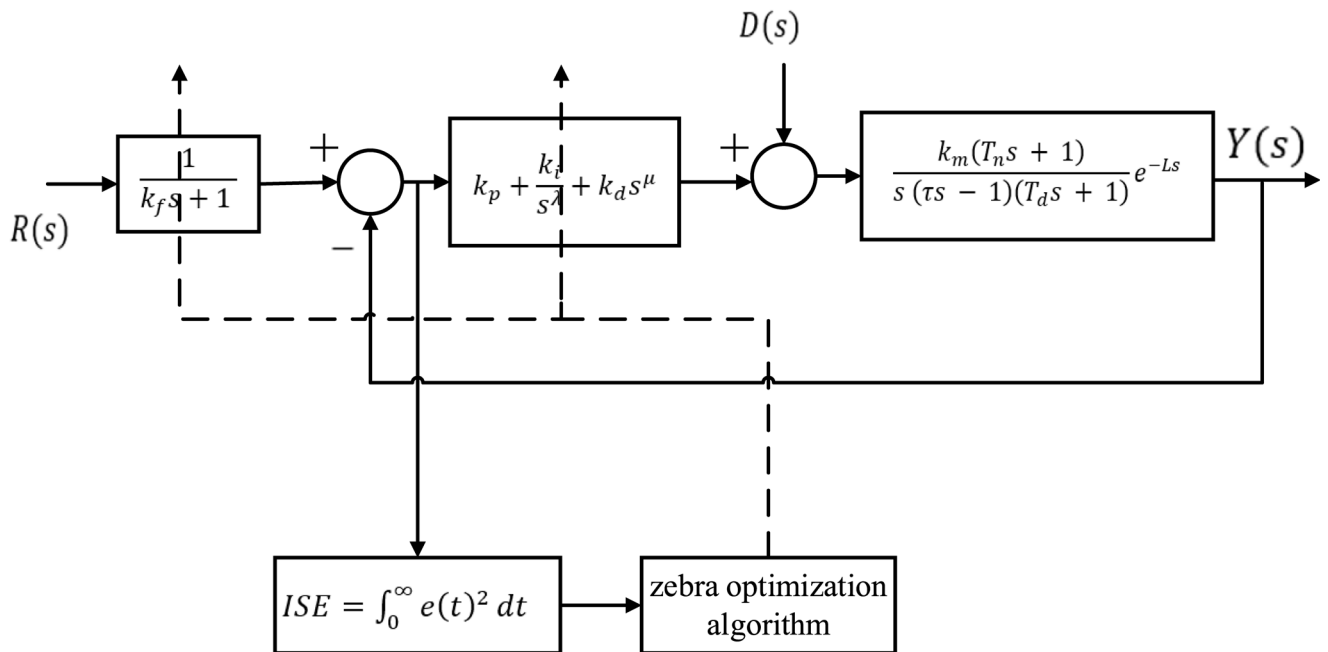


Figure 4. Zebra-optimized FOPID controller pulse setpoint filter

unstable launch-vehicle dynamics with strong actuator limits and atmospheric disturbances through its tight disturbance rejection and minimal control effort.

The proposed method findings are in line with other research showing that FOPID outperform standard PID in terms of robustness and disturbance rejection. However, the suggested solution uses an offline optimization framework to produce better transient response and a large reduction in peak control effort (about 95%) and consequently reduced overshoot, in contrast to various existing approaches such that of PID reported in¹, as clear in Figure 6. This demonstrates how well fractional-order control

and metaheuristic tuning in addition to the setpoint filter work together for unstable launch-vehicle applications.

It is also important to emphasize that the considered plant in equation (3) is characterized by the presence of an integrator and an unstable pole, which are common system models in launch-vehicle attitude dynamics and other unstable aerospace systems. Therefore, the proposed control strategy is not limited to the specific case studied here, but can be extended to systems exhibiting similar dynamic characteristics. This general structure enhances the applicability of the proposed method to a broader class of unstable control problems.

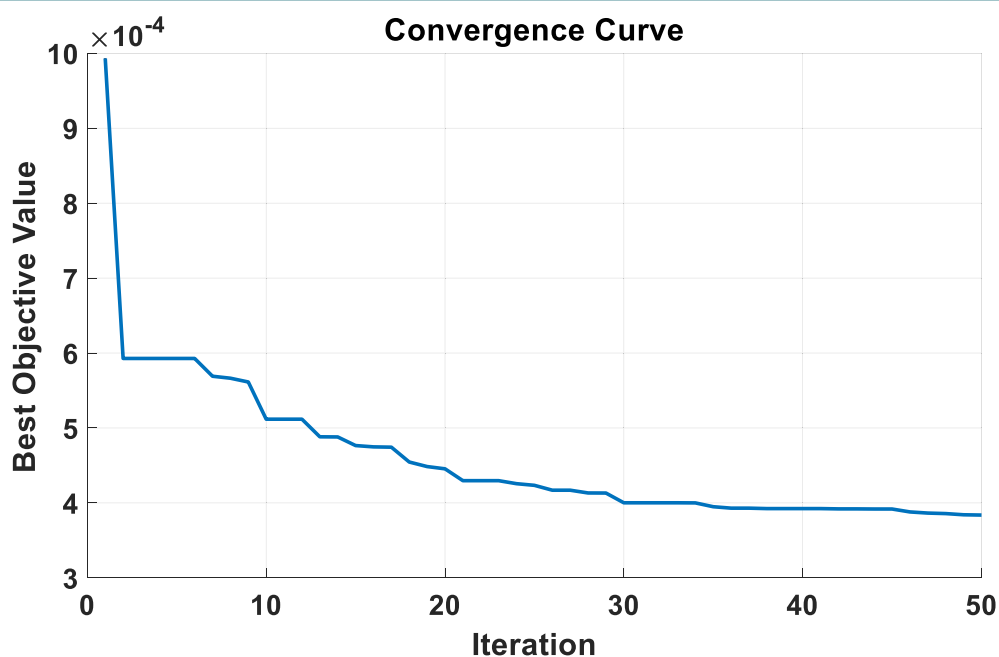


Figure 5. Convergence curve of ZOA

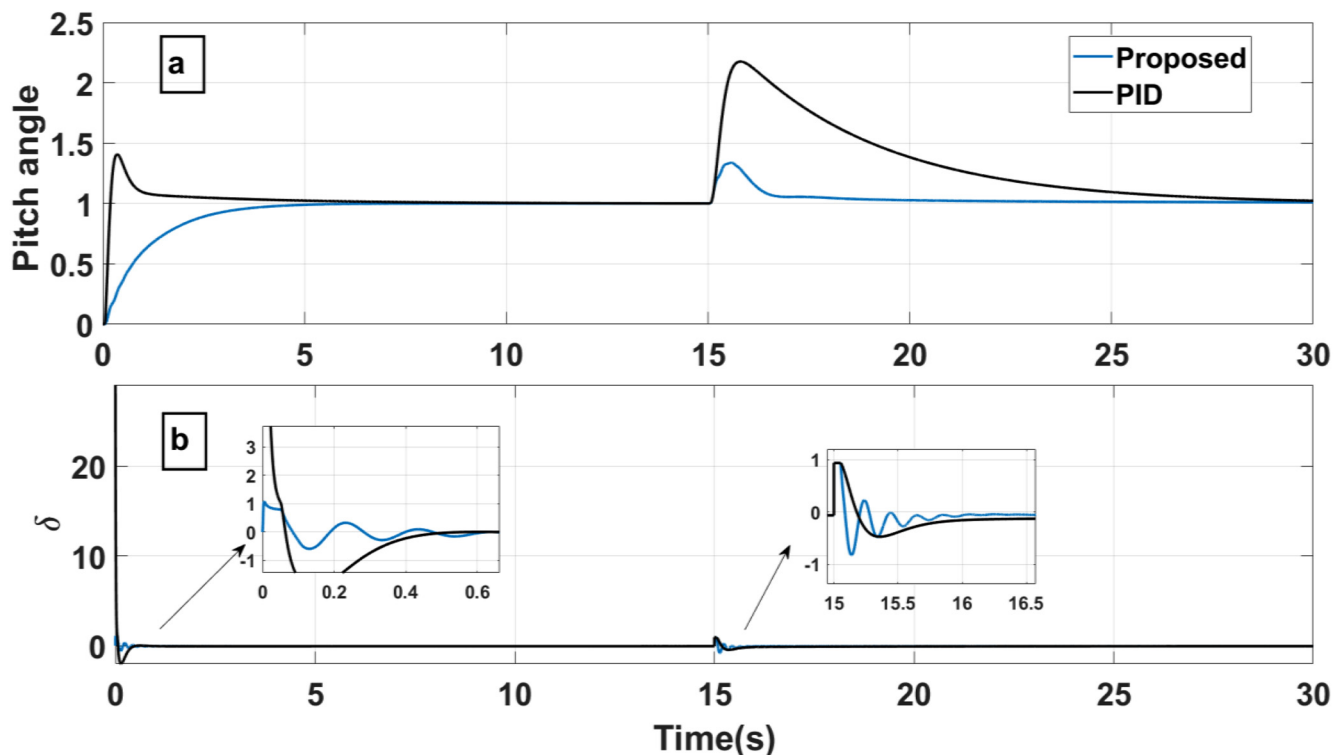


Figure 6. (a) Attitude; (b) control signals

6. CONCLUSION AND FUTURE OUTLOOK

In this study, we investigated the feasibility of using an FOPID pulse setpoint filter controller optimized with the ZOA for attitude control of unstable spacecraft. Achieving attitude control of these systems is challenging because of the existence of unstable and integrated poles. However, the proposed approach successfully solved this difficult control problem. Compared with previously reported PID in the literature strategy for ULV, the proposed method provides enhanced transient response and improved disturbance rejection while maintaining implementation simplicity. However, several limitations should be acknowledged. First, the controller parameters were optimized offline under predefined operating conditions. Therefore, performance under varying system parameters or extreme disturbances was not investigated. Second, the validation was limited to simulation without full-scale hardware-in-the-loop spacecraft validation. Additionally, the stochastic nature of ZOA may introduce slight variability in the obtained gains. Future studies will focus on hardware-in-the-loop implementation and investigation of adaptive or online tuning mechanisms to further enhance robustness under varying mission scenarios.

Table 1. Controller and ZOA parameters

Controller parameters		Optimization algorithm parameters	
k_p	2.3498	Search agents	30
k_i	2.9909	Iterations	50
k_d	0.5858	Optimized variables	6
λ	0.7981	Lower bounds	[0.1 0.1 0.1 0.1 0.1 0.1]
μ and k_f	1.1513	Upper bounds	[3 3 3 2 2 2]

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Yazan Zwawi – Department of Aerospace Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; s202180810@kfupm.edu.sa

Corresponding Author

Fadi Alyoussef – Research Mentor, Post-Doctoral Fellow, Interdisciplinary Research Center for Aviation and Space Exploration, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; fadi.youssef@kfupm.edu.sa

ACKNOWLEDGEMENT

The authors acknowledge the support provided by the Department of Aerospace Engineering and the Interdisciplinary Research Center for Aviation and Space Exploration at King Fahd University of Petroleum and Minerals. This research was funded by the Undergraduate Research Office (URO), KFUPM, Saudi Arabia.

REFERENCES

- (1) Sun, M., Zhang, L., Wang, Z. & Chen, Z. PID pitch attitude control for unstable flight vehicle in the presence of actuator delay: Tuning and analysis. *J. Franklin Inst.* **351**, 5523–5547 (2014).
- (2) Sun, M., Yang, R., Wang, Z. & Chen, Z. Stability margin-based PD attitude control tuning for unstable flight vehicle. *Int. J. Syst. Sci.* **44**, 240–251 (2013).

- (3) Alyoussef, F. & Kaya, I. A New Dynamic Sliding Mode Controller with Disturbance Observer for Controlling Integrating Processes with Time Delay. *Int. J. Control* **97**, 1136–1156 (2023).
- (4) Alyoussef, F. & Kaya, I. Improved adaptive dynamic non-singular terminal sliding mode controller with fractional disturbance observer. *Inf. Sci. (Ny)*. **641**, 119110 (2023).
- (5) Alyoussef, F. & Kaya, I. Simple PI-PD tuning rules based on the centroid of the stability region for controlling unstable and integrating processes. *ISA Trans.* **134**, 238–255 (2023).
- (6) Alyoussef, F., Abdallah, A. & Farahat, A. Robust optimal fractional-order proportional-integral and proportional-derivative controller design for integrating systems with time delays: Real-time application to quadrotors. *ISA Trans.* **167**, 1978–1995 (2025).
- (7) Wu, Q., Sun, M., Chen, Z., Yang, Z. & Wang, Z. Tuning of active disturbance rejection attitude controller for statically unstable launch vehicle. *J. Spacecr. Rockets* **54**, 1383–1389 (2017).
- (8) Sun, M., Jiao, G., Yang, R. & Chen, Z. ADRC for statically unstable longitudinal flight dynamics. *Proc. 30th Chinese Control Conf. CCC 2011* 6274–6280 (2011).
- (9) Bashishtha, T. K., Singh, V. P., Yadav, U. K. & Sahu, U. K. Fractional-order PID controllers and applications: A comprehensive survey. *Annu. Rev. Control* **60**, 101013 (2025).
- (10) Milanesi, M., Visioli, A. & Chen, Y. Q. From PID to PIDD2 α : Performance improvement with a fractional double derivative action. *Annu. Rev. Control* **60**, 101006 (2025).
- (11) Shah, P. & Agashe, S. Review of fractional PID controller. *Mechatronics* **38**, 29–41 (2016).
- (12) Elbaset, A. A., Ismail, M. M. & Alaboudy, A. H. K. Particle Swarm Optimization for layout design of utility interconnected wind parks. *2018 IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. ISGT 2018* 1–6 (2018) doi:10.1109/ISGT.2018.8403367.
- (13) Xu, W., Ismail, M. M., Liu, Y. & Islam, M. R. Parameter optimization of adaptive flux-weakening strategy for permanent-magnet synchronous motor drives based on particle swarm algorithm. *IEEE Trans. Power Electron.* **34**, 12128–12140 (2019).
- (14) Aldhaif Allah, M. M. & Mohamed, M. M. Adaptive Velocity PSO-Based Parameter Optimization for a Permanent Magnet DC Motor Drive in Light Electric Vehicles. *J. Undergrad. Res. Int.* **1**, 1–12 (2025).
- (15) Das, D., Chakraborty, S. & Raja, G. L. Enhanced dual-DOF PI-PD control of integrating-type chemical processes. *Int. J. Chem. React. Eng.* **21**, 907–920 (2021).
- (16) Alyoussef, F. & Kaya, I. Tuning proportional-integral controllers based on new analytical methods for finding centroid of stability locus for stable/unstable first-order plus dead-time processes. *Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng.* **236**, 818–831 (2022).
- (17) Alyoussef, F. & Kaya, I. Proportional–integral and proportional–derivative controller design based on analytically computed centroid point for controlling integrating processes. *Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng.* **237**, 1045–1065 (2023).
- (18) Pazhanimuthu, C., Saravanan, G., Suresh, K. P. & Kumar, R. S. Performance analysis of voltage profile improvement in AVR system using zebra optimization algorithms based on PID controller. *e-Prime - Adv. Electr. Eng. Electron. Energy* **6**, 100380 (2023).
- (19) Trojovska, E., Dehghani, M. & Trojovsky, P. Zebra Optimization Algorithm: A New Bio-Inspired Optimization Algorithm for Solving Optimization Algorithm. *IEEE Access* **10**, 49445–49473 (2022).

Search-and-Track Missions with Limited Communication in Unmanned Aerial Vehicle Swarms: A Centralized Leader-Follower Study

Jinan Ali Shajiri and Mohammed Basheer Mohiuddin*

Cite <https://doi.org/10.64589/juri/218962>

Submitted: November 27, 2025 Revised: January 05, 2026 Accepted: January 28, 2026

ABSTRACT

Unmanned Aerial Vehicle (UAV) swarms exhibit strong potential for complex missions. However, there has been limited research on their behavior during search-and-track missions, particularly under sensing limitations. This study explores a leader-follower UAV swarm that performs a search-and-track mission in which sensing capabilities are concentrated on the leader UAV. In this scenario, the leader detects and tracks the target, while the follower drones operate with limited sensing and receive updates from the leader. Once the target is detected, the swarm transitions into a coordinated ring formation surrounding the target, with the leader guiding the overall ring formation. This study aimed to demonstrate how limited follower sensing affects the formation quality, response to target motion, and overall tracking performance. A simulation framework was developed to evaluate swarm behavior under sensing limitations. The results showed that effective tracking and encirclement can be achieved even when advanced sensing is concentrated on the leader, thereby demonstrating the potential of building scalable and cost-efficient UAV swarm systems.

Keywords: unmanned aerial vehicle swarm, leader-follower control, limited sensing, search-and-track, ring formation, multi-agent systems

1. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) are essential tools for surveillance, reconnaissance, search-and-rescue missions, and environmental monitoring^{1,2}. UAV swarms offer clear advantages over single-UAV systems owing to their ability to coordinate and distribute tasks, as well as cover large areas efficiently³. Inspired by biological swarms, these systems exhibit strong potential for complex missions that require adaptability and cooperation among multiple agents³.

Although many studies have assumed that UAVs within a swarm possess similar sensing and processing capabilities³, real-world constraints often render this assumption unrealistic. Factors such as the weight, power consumption, cost, and sensor limitations create significant differences among agents^{3,4}. Consequently, swarm architectures that rely on heterogeneous sensing, where one UAV carries advanced sensing equipment while others operate with limited capabilities, are gaining attention³. However, research on the behavior of such swarms during search-and-track missions remains limited⁵.

Methodologies for search-and-track operations where only the leader UAV has full sensing capabilities, while the follower UAVs possess limited sensing and take orders from the leader UAV, remain limited^{3,6}. Our study addresses this research gap. The existing literature has explored leader-follower models and coop-

erative search strategies^{7,8}; however, few studies have analyzed how sensing asymmetry affects the performance and coordinated encirclement of a detected target⁸⁻¹⁰.

This study proposes a simulation framework to evaluate the behavior of a leader-follower UAV swarm in which the leader possesses full sensing and processing capabilities, while the follower UAVs operate with limited local sensing and execute commands issued by the leader. The framework simulates the transition of a swarm into a formation surrounding a detected target, with the leader coordinating the movement. The aim is to assess the impact of sensing limitations on swarm performance during tasks such as searching, tracking, and encircling. This proof-of-concept highlights the feasibility of deploying UAV swarms with limited sensing capabilities, offering valuable insights into the design of scalable and cost-effective UAV swarm systems for real-world applications where sensing resources are often unevenly distributed.

The remainder of this paper is organized as follows. Section 2 presents the proposed methodology, including the overall framework, the system model, and the developed algorithm. Section 3 reports the results obtained from the study. Section 4 discusses the findings and their implications. Finally, Section 5 concludes the paper by summarizing the key contributions and outlining potential directions for future work.

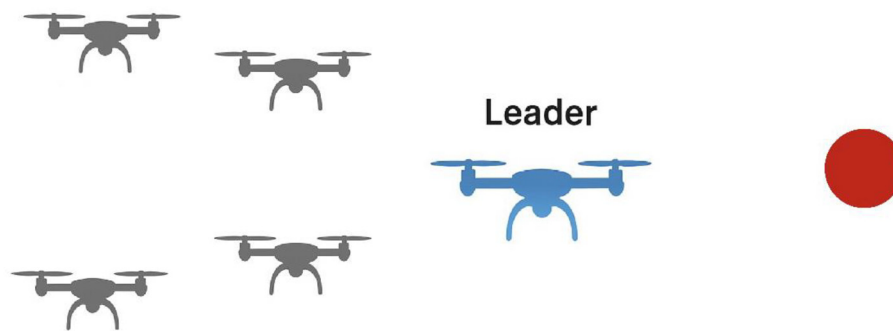


Figure 1. Leader UAV performing the search phase with follower UAVs moving in straight lines before transitioning to a ring formation

2. METHODOLOGY

This study evaluated the behavior of a leader–follower UAV swarm in search-and-track missions. The following subsections describe the simulation setup, swarm model, and the formation and coordination strategies.

2.1. Framework. Swarm behavior was evaluated using a simulation framework developed in Simulink software. The simulation environment models a three-dimensional search area in which a leader–follower UAV swarm performs a search-and-track mission. The number of UAVs, their initial positions, and the target trajectory are configurable parameters. This framework allows the monitoring of UAV positions and velocities at each step, enabling a detailed analysis of swarm coordination and formation dynamics.

2.2. Leader–Follower Model. In the initial phase of the mission, leader UAV performs a search operation using a right-left sinusoidal motion to scan the area, as shown in Figure 1. The follower UAVs move in straight lines until the leader detects the target.

Once the target is found, the follower UAVs dynamically adjust their positions based on the information received from the leader, ensuring synchronization as the leader moves through the search area.

Each swarm UAV is controlled using a velocity controller. A proportional derivative (PD) controller was used. The control law is given by:

$$\mathbf{v}_i = \mathbf{K}_p * (\mathbf{p}_{leader} - \mathbf{p}_i) + \mathbf{K}_d * (\mathbf{v}_{leader} - \mathbf{v}_i), \quad (1)$$

where $\mathbf{v}_i$ is the velocity of the follower UAV i ; $\mathbf{p}_{leader}$ and $\mathbf{p}_i$ are the positions of leader and follower UAVs, respectively; $\mathbf{v}_{leader}$ and $\mathbf{v}_i$ are the velocities of the leader and follower UAVs, respectively; and $\mathbf{K}_p$ and $\mathbf{K}_d$ are the proportional and derivative gains, respectively.

Once the target is detected, the swarm transitions into a ring formation around the target to continue the tracking mission, whereas the leader remains closer to the target than the follower UAVs. This dynamic adjustment ensures that the UAVs maintain a coordinated formation.

2.3. Circular Formation Algorithm. Once the target is detected by the leader UAV, the swarm transitions from search mode to formation mode around the target. As shown in Figure 2, the leader UAV remains close to the target, guiding the follower UAVs in a ring formation. The target remains at the center

of the formation. Although the leader is positioned closer to the target, it still participates in ring formation, ensuring coordinated movement and guiding the follower UAVs.

The position of each follower UAV in the circle is calculated as follows:

$$\mathbf{p}_i = \mathbf{p}_{target} + r * \cos(\theta_i) * \mathbf{i} + r * \sin(\theta_i) * \mathbf{j} \quad (2)$$

where (r) is the radius of the circular formation and (θ_i) is the angle of the UAV relative to the leader and other followers. This ensures that the UAVs maintain fixed positions relative to each other around the target.

2.4. Communication. Communication between UAVs is leader-to-follower only. The leader broadcasts its position and the estimated position of the target at regular intervals. Followers update their velocities based on the latest information, whereas no direct communication occurs between them, thereby simulating a limited hardware capability.

$$\mathbf{v}_i = f(\mathbf{p}_{leader}, \mathbf{p}_{target}, \Delta t) \quad (3)$$

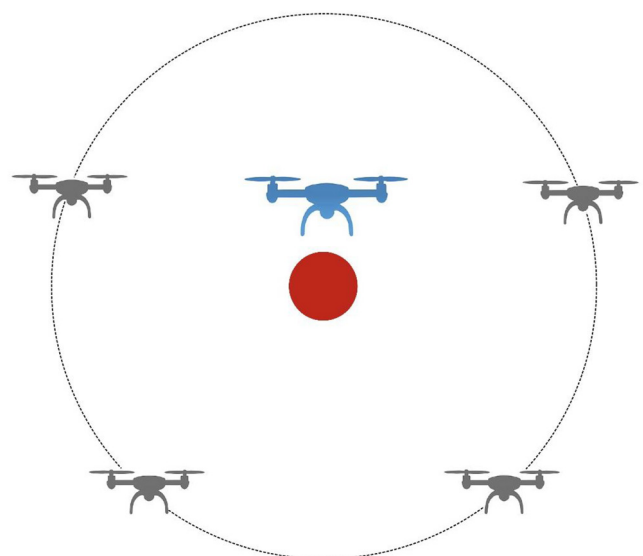


Figure 2. Leader UAV guiding the swarm into a circular formation around the target after detection

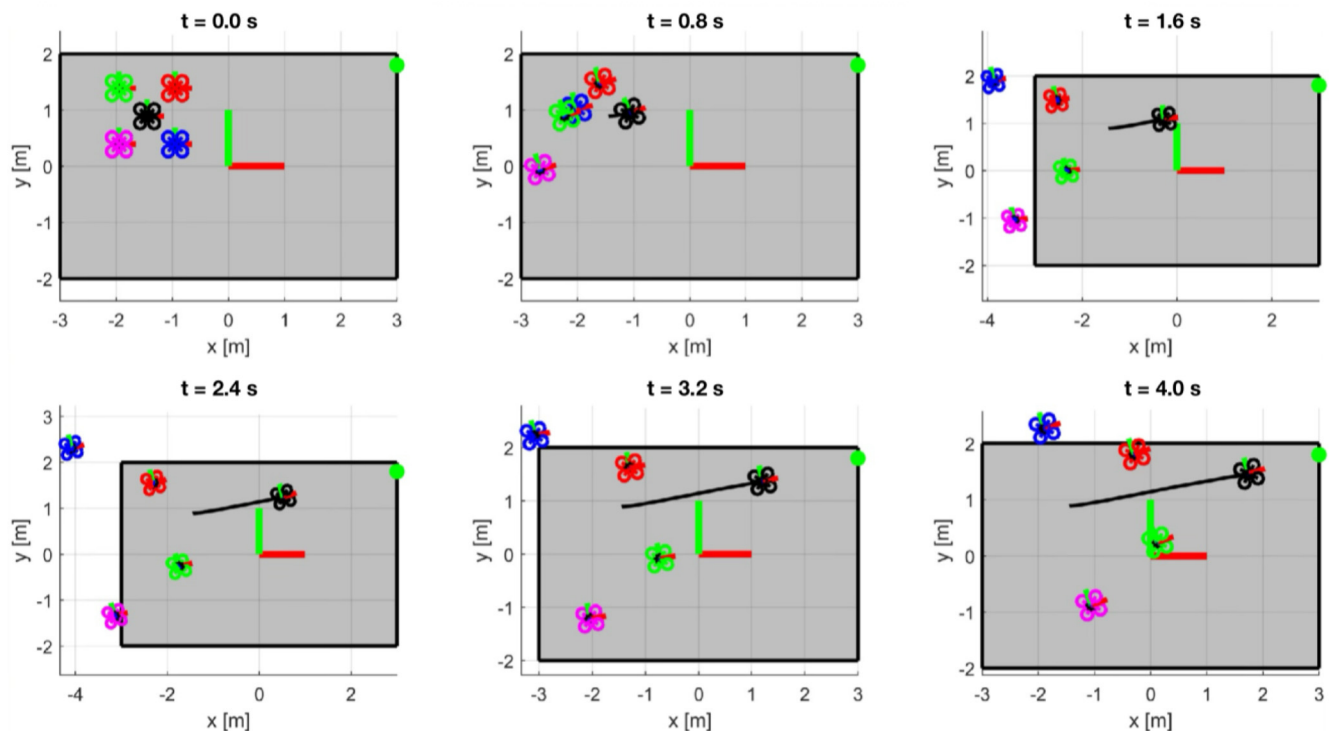


Figure 3. Top views of motion and transition to ring formation of the leader UAV at different times (a) $t = 0.0$ s; (b) $t = 0.8$ s; (c) $t = 1.6$ s; (d) $t = 2.4$ s; (e) $t = 3.2$ s; (f) $t = 4.0$ s

3. RESULTS

This section presents the key findings of the study, highlighting the performance of the leader-follower UAV swarm during the search-and-track mission with limited sensing capabilities in follower UAVs. The results were analyzed in terms of the tracking, formation stability, and response to the target. The data collected from the simulation provide insight into how the sensing capabilities, geometric formation parameters, and follower synchronization of the leader contribute to the overall performance of the swarm.

3.1. Swarm Evolution in a Three-dimensional Arena.

As shown in Figure 3, at $t = 0.0$ s, the UAV swarm begins from the starting position, and the follower UAVs continue to follow the path of the leader. Over time, follower UAVs dynamically adjust their positions based on the movements of the leader, ensuring synchronization. Following the search pattern, the leader UAV gradually narrows to the target, which is marked by a green dot. This movement allows the swarm to explore the search area efficiently.

In the initial phase of the mission, the UAV swarm behavior was analyzed in a three-dimensional environment. As shown in Figure 4, the leader UAV (black) in search mode moves in a right-left sinusoidal motion, whereas the follower UAVs remain in straight lines until the leader detects the target (green dot). It can be observed that at $t = 6.0$ s, the UAVs start transitioning into a ring formation around the target.

Figure 4 shows the evolution of the UAV swarm over time, highlighting the adjustments made by each UAV.

At $t = 3.95$ s, the camera of the leader UAV detects the target for the first time.

3.2. Ring Formation. Once the leader UAV detects the target, the swarm transitions into a circular formation around the target. As shown in Figure 5, the follower UAVs (D2, D3, D4, and D5) adjust their positions to maintain a fixed distance between each other and the target. The final positions of the UAVs in the circular formation show the leader UAV (D1, marked in blue) remaining near the target, and the follower UAVs (D2, D3, D4, and D5) forming a stable perimeter around the target.

The ring formation ensures that the swarm can continuously track the target, maintaining synchronization between the UAVs while preserving the desired formation. The results show that the UAVs successfully maintain a coordinated path and that the leader remains at an optimal distance from the target.

The distance between the leader and the target, as well as the convergence of the followers, was measured over time.

Figure 6 illustrates the convergence of the leader and followers to the target over time. The leader UAV, shown in blue, initially maintains a distance from the target, but the distance is quickly reduced once it detects the target. The follower UAVs follow the movements of the leader and gradually reduce their distance to the target.

The tracking performance measured by the distance to the target indicates that the leader UAV plays a key role in guiding the swarm to track and approach the target effectively.

The ability of the swarm to converge to a circular formation can be observed as the leader UAV guides the followers toward the target. As shown in Figure 7, the trajectory of the leader UAV (in blue) initially follows a path toward the target. The follower UAVs dynamically adjust their positions to ensure that the formation converges around the target. The final position of the leader UAV is indicated by blue dots, and the target is marked in red.

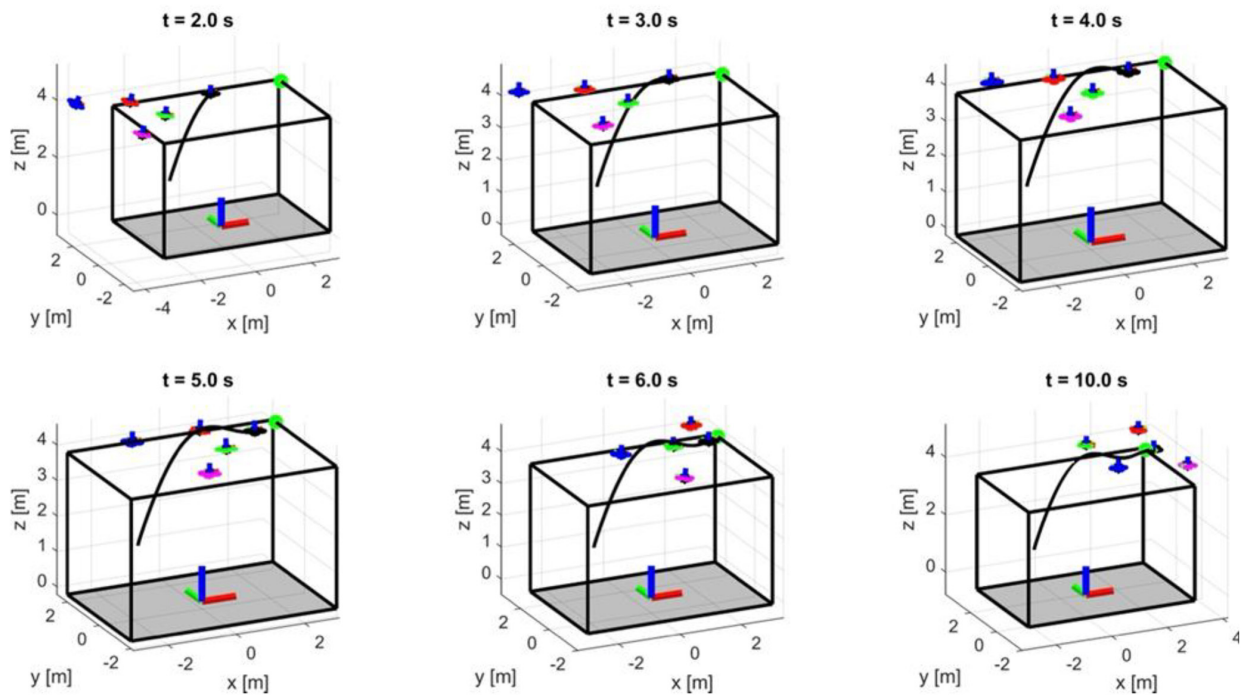


Figure 4. Overview of motion and transition to ring formation of the leader UAV at different times (a) $t = 2.0$ s; (b) $t = 3.0$ s; (c) $t = 4.0$ s; (d) $t = 5.0$ s; (e) $t = 6.0$ s; (f) $t = 10.0$ s

The evolution of the swarm can be observed as the leader moves toward the target and the followers adjust their positions accordingly. A smooth transition to a circular formation is demonstrated as the UAVs synchronize their movements while maintaining the correct formation relative to each other. The results show that the proximity of the leader to the target plays a crucial role in enhancing the coordination and tracking efficiency of the swarm.

4. DISCUSSION

This study provides valuable insights into the behavior of leader-follower UAV swarms during a search-and-track mission, particularly when follower UAVs have limited sensing capabilities. A key

point is the importance of the role of the leader UAV in ensuring effective target tracking and maintaining formation stability. Despite the sensing limitations of the followers, the swarm can adapt and stay synchronized, which highlights the potential of leader-follower systems for missions where communication and sensing resources are constrained.

4.1. Role of the Leader UAV. The results clearly demonstrate that the sensing capabilities of the leader UAV play a critical role in the success of a tracking mission. With the leader being able to detect and track the target efficiently, the follower UAVs can adjust their positions in response to the movements of the leader when the target is detected. This shows that even with limited sensing, follower UAVs can function effectively when they

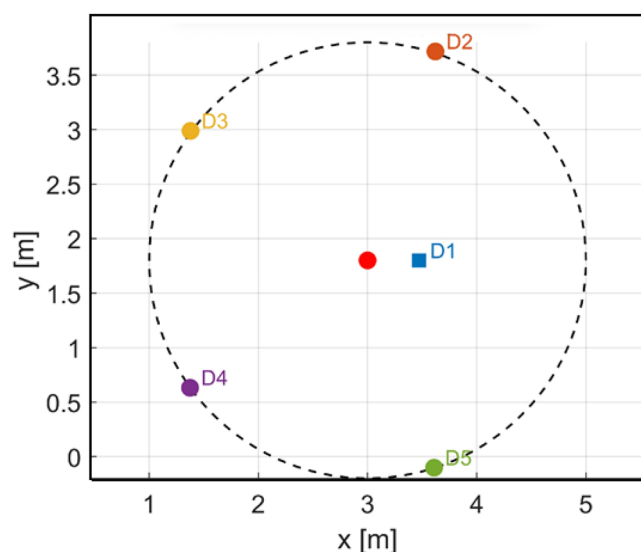


Figure 5. Final ring formation around the target

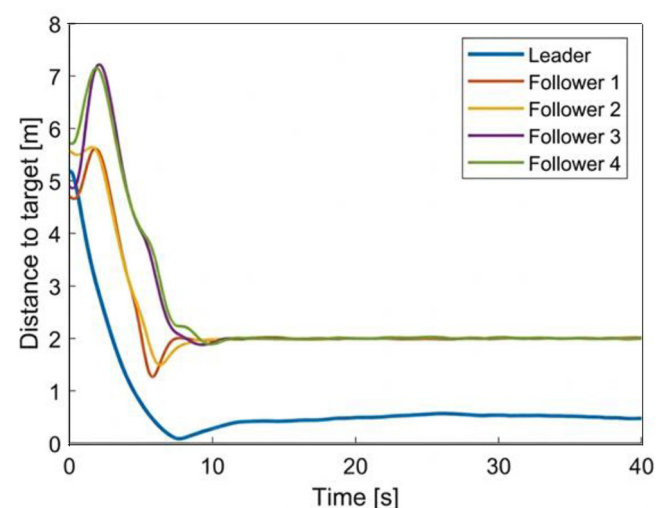


Figure 6. Convergence to the target and tracking performance

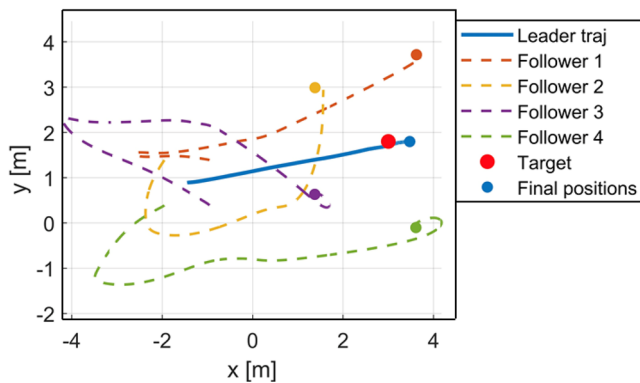


Figure 7. Swarm path and convergence to ring formation

receive accurate information from the leader. This is particularly useful for reducing the need for constant communication among all the UAVs in a swarm, as demonstrated in this study. These findings align with previous research on leader–follower models, which suggests that such systems are effective even with asymmetry in sensing and communication capabilities^{9,10}.

4.2. Proximity of the Leader to the Target. A key decision in this study was to position the leader UAV closer to the target than the follower UAVs. The proximity of the leader allows it to perform critical tasks, such as handling payloads or making precise adjustments to interact with the target, which is crucial for mission success.

The results showed that this asymmetric configuration, with the leader near the target, significantly improved the coordination and efficiency of the swarm. This allowed followers to maintain their formation without requiring extensive resources. This setup is particularly effective for missions such as surveillance and environmental monitoring, where the proximity of the leader to the target can enhance the overall performance of the swarm.

4.3. Limitations. This study does not account for three-dimensional flight dynamics, aerodynamic disturbances, or environmental uncertainties. The communication model assumes idealized periodic updates, which may not fully represent practical wireless network behavior. Moreover, the leader is assumed to possess accurate and reliable sensing. The follower dynamics are also modeled using simplified kinematic control laws without considering actuator saturation or hardware constraints. Finally, the single-leader architecture introduces a dependency on one agent, making the swarm vulnerable to performance degradation in the event of leader failure.

5. CONCLUSIONS

This study investigated the behavior of a leader–follower UAV swarm performing a search-and-track mission with limited sensing capabilities in follower UAVs. This research aimed to assess performance parameters, such as coordination, tracking, formation stability, and response to target motion.

Geometric parameters, such as UAV spacing and formation size, play a critical role in maintaining the formation and effective tracking of the target. The results showed reasonable agreement

with the existing literature on UAV swarm behavior, with tracking accuracy and formation stability aligned with reported values, offering new insights into the behavior of leader–follower UAV swarms with limited sensing. The proximity of the leader UAV to the target also opens new possibilities for advanced applications, such as payload execution and precise target interaction, where the leader can perform more complex tasks. This proximity significantly enhances the operational flexibility of a UAV swarm, allowing for more diverse mission profiles in dynamic environments. The findings support the potential use of leader–follower UAV swarm systems with limited sensing in various applications, such as surveillance and environmental monitoring, provided that further validation under real-world service conditions and long-term assessments are conducted.

The results of this study contribute to a growing body of research focusing on improving the efficiency and applicability of UAV swarms in real-world scenarios. Future work may include long-term performance studies, real-world testing, and computational modeling to optimize swarm behavior and expand the use of limited sensing UAVs for complex missions.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Jinan Ali Shajiri – Aerospace Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; 0009-0007-9935-0252
Email: s202367390@kfupm.edu.sa

Corresponding Author

Mohammed Basheer Mohiuddin – Research Mentor, Post-doctoral Fellow, Interdisciplinary Research Center for Aviation and Space Exploration, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; 0000-0001-5110-2865
Email: mohammed.mohiuddin@kfupm.edu.sa

ACKNOWLEDGEMENTS

The authors acknowledge the support provided by the Department of Aerospace Engineering and the Interdisciplinary Research Center - Aviation & Space Exploration at King Fahd University of Petroleum and Minerals for their support. This research was funded by the Project No. **INAE2500**.

REFERENCES

- (1) Tahir, A., Böling, J., Haghbayan, M.-H., Toivonen, H. T. & Plosila, J. Swarms of Unmanned Aerial Vehicles — A Survey. *J. Ind. Inf. Integr.* **16**, 100106 (2019).
- (2) Mohiuddin, M. B. *et al.* Reinforcement learning for end-to-end UAV slung-load navigation and obstacle avoidance. *Sci. Rep.* **15**, 34621 (2025).
- (3) Alqudsi, Y. & Makaraci, M. UAV swarms: research, challenges, and future directions. *J. Eng. Appl. Sci.* **72**, 12 (2025).

- (4) Li, X. *et al.* AirSwarm: Enabling Cost-Effective Multi-UAV Research with COTS drones. Preprint at <https://doi.org/10.48550/ARXIV.2503.06890> (2025).
- (5) Spyridis, Y., Lagkas, T., Sarigiannidis, P. & Zhang, J. Modelling and simulation of a new cooperative algorithm for UAV swarm coordination in mobile RF target tracking. *Simul. Model. Pract. Theory* **107**, 102232 (2021).
- (6) Javed, S. *et al.* State-of-the-Art and Future Research Challenges in UAV Swarms. *IEEE Internet Things J.* **11**, 19023–19045 (2024).
- (7) Bu, Y., Yan, Y. & Yang, Y. Advancement Challenges in UAV Swarm Formation Control: A Comprehensive Review. *Drones* **8**, 320 (2024).
- (8) Kilic, F., Yang, H. & Hardt, W. Leader-Follower Control and Distributed Communication based UAV Swarm Navigation in GPS-Denied Environment. *Embed. Selforganising Syst.* 3-11 Pages (2023) doi: [10.14464/ESS.V10I8.657](https://doi.org/10.14464/ESS.V10I8.657).
- (9) Han-Lim Choi, Brunet, L. & How, J. P. Consensus-Based Decentralized Auctions for Robust Task Allocation. *IEEE Trans. Robot.* **25**, 912–926 (2009).
- (10) Abro, G. E. M., Abdallah, A. M., Zahid, F. & Ahmed, S. A Comprehensive Review of Next-Gen UAV Swarm Robotics: Optimisation Techniques and Control Strategies for Dynamic Environments. *Intell. Autom. Soft Comput.* **40**, 99–123 (2025).

Embedded Kalman Filter-Based GNSS/INS Fusion for Robust Planetary Rover Navigation

Ahmed Bashammakh and Atif Mahmood*

Cite <https://doi.org/10.64589/juri/219336>

Submitted: November 27, 2025 Revised: February 19, 2026 Accepted: February 25, 2026

ABSTRACT

Rovers require advanced guidance and navigation systems to accurately determine their position and orientation while navigating uneven terrain and avoiding hazards. Therefore, inertial navigation systems (INS) have been integrated with available satellite navigation data to meet these requirements and achieve reliable autonomous navigation on other planets. This study presents a hybrid navigation solution for planetary rovers that targets robust autonomous localization under intermittent or degraded satellite coverage consist of global navigation satellite system (GNSS) and inertial navigation system (INS). The proposed approach combines high-rate inertial measurements with periodic GNSS position fixes using a Kalman-filter-based sensor fusion architecture. By combining the continuous motion propagation of the INS with the long-term stability of GNSS updates, the method bounds the inertial drift while preserving the smooth state estimates during GNSS outages. The predicted update loop was implemented using a low-power Arduino Mega 2560 microcontroller. A hardware prototype integrating an inertial measurement unit and a global positioning system receiver was developed to validate the framework. Experimental results demonstrated that the fused GNSS/INS solution closely tracked the ground truth, reducing position drift by approximately 75% relative to INS-only dead reckoning and improving the overall trajectory accuracy compared with single-sensor approaches. This compact, embedded implementation highlights the feasibility of low-cost and reliable navigation systems for future planetary rover missions.

Keywords: planetary rover, inertial navigation system, global navigation satellite system, navigation, Kalman filter

1. INTRODUCTION

Planetary rover missions, which operate in harsh and uncertain environments and perform complex scientific tasks that rely on precise onboard navigation, are essential for interplanetary exploration, as demonstrated by the Sojourner (1997) and Mars Exploration Rovers^{1,2}. As rovers require advanced navigation systems to accurately determine their position and orientation, researchers have integrated inertial navigation systems (INS) with available satellite navigation data to enable reliable navigation on other planets. The recently developed global navigation satellite system (GNSS)/INS integration methods leverage experience from terrestrial applications and lunar exploration efforts^{3,4}.

The long communication delay with the Earth is a fundamental challenge in Mars exploration. The latency of two-way radio signals between Mars and Earth (>10 min)⁵. Thus, the rover must utilize onboard sensors to safely navigate and make independent decisions while traversing long distances.

The operational range of a rover that depends solely on a GNSS-like satellite navigation system is limited by the availability of reliable satellite signals. Deep-space missions may lack dedicated GNSS coverage and planetary-orbiting satellites provide only occasional position fixes. Nevertheless, GNSS may still

support lunar navigation with errors⁶. However, an INS can continuously propagate the estimated position, velocity, and attitude of the rover by integrating data from onboard accelerometers and gyroscopes⁷. Planetary rovers typically augment the inertial and wheel odometries with vision-based localization to mitigate slippage⁸.

Based on these insights, the hybrid GNSS/INS navigation architecture leverages the complementary advantages of each technology^{9,10}. In this study, a complete GNSS/INS fusion was developed, including a Kalman filter-based navigation algorithm and embedded integration of the GPS receiver and an inertial measurement unit (IMU). Experimental validation and dataset analysis demonstrated a compact, low-cost hybrid navigation solution with improved accuracy compared with standalone inertial dead reckoning.

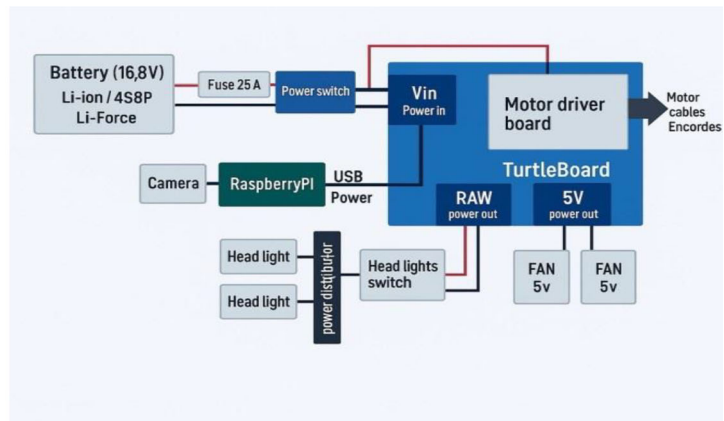
2. METHODOLOGY

This section describes the experimental and algorithmic procedures used to design, implement, and validate the proposed hybrid GNSS/INS navigation system for planetary rovers.

2.1. Rover System Architecture. Figure 1 (a) shows the rover prototype and illustrates its physical implementation. This



(a) Rover prototype



(b) Schematic of the rover system

Figure 1. Prototype of the rover and its circuitry

integrated design emphasizes system modularity and scalability, which allows subsystems to be added or removed with minimal reconfiguration, while ensuring consistent and reliable power delivery¹¹. Figure 1 (b) illustrates the centralized power and control architecture of the rover, featuring a central distribution board at its core to interconnect all the major subsystems via a unified wiring schematic. This is employed as a centralized hub for interfacing all the subsystems to ensure robust operation under varying load conditions¹².

2.2. Kalman Filter-Based Sensor Fusion Framework.

Sensor fusion techniques are essential for ensuring robust navigation in GNSS-degraded or-denied environments. Although GNSS provides absolute position measurements with high accuracy, it lacks the capability to estimate orientations and may experience signal loss owing to planetary terrain or interference. By contrast, IMU continuously estimates motion through accelerometer and gyroscope data; however, its output drifts over time because of sensor bias and noise over time. Integrating both the sensors using a Kalman filter combines the strengths of each sensor and compensates for their individual limitations. Because of its easy implementation and small computational load, as well as the low speed of the rover, the Kalman filter is widely used in real-time embedded navigation sensor-fusion applications¹³.

In this study, a model of the rover motion along a single axis (x-direction) was applied to estimate its position and velocity over time. The IMU of the rover provides acceleration a_x , while the GNSS provides discrete position updates. The system state vector is defined as

$$X_k = [x_k \ v_k]^T \quad (1)$$

where x_k and v_k denote the position and velocity, respectively, at the time step k . The motion model assumes constant acceleration over each time interval Δt , resulting in the following state transition model:

$$A = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \quad (2)$$

$$B = [0.5 \Delta t^2, \Delta t]^T \quad (3)$$

$$X_{k|k-1} = A \cdot X_{k-1} + Bu_k \quad (4)$$

This propagates the previous state forward using IMU acceleration, enabling the prediction equations to estimate the next state and its uncertainty as follows.

$$\hat{x}_{k|k-1} = A \cdot \hat{x}_{k-1} + B \cdot a_k \quad (5)$$

$$P_{k|k-1} = A \cdot P_{k-1} \cdot A^T + Q \quad (6)$$

where $\hat{x}$, $P_{k|k-1}$, and Q denote the predicted state, predicted error covariance, and process noise covariance matrix (tuned empirically), respectively.

The GNSS provides noisy measurements of the position for the updated state, modeled as

$$z_k = H \cdot x_k + w_k \quad (7)$$

$$H = [1, 0] \quad (8)$$

The Kalman gain K_k is computed as:

$$K_k = P_{h|k-1} \cdot H^T \cdot (H \cdot P_{h|k-1} \cdot H^T + R)^{-1} \quad (9)$$

The state and covariance are then updated as

$$\hat{x}_k = \hat{x}_{k|k-1} + K_k \cdot (z_k - H \cdot \hat{x}_{k|k-1}) \quad (10)$$

$$P_k = (I - K_k \cdot H) \cdot P_{k|k-1} \quad (11)$$

where R denotes the measurement noise covariance (from the GNSS uncertainty); z_k is the measured GPS position; and K_k denotes the Kalman gain determining the filter trust in the GPS vs. prediction.

2.3. Embedded Sensor Fusion Implementation. The proposed hybrid navigation framework was validated using a hardware-in-the-loop prototype constructed using an embedded microcontroller platform. This implementation enabled real-time acquisition, processing, and fusion of sensor measurements under practical operating conditions, thereby providing a representative testbed for evaluating the feasibility of the navigation architecture. Table 1 summarizes the primary components used in this implementation and compares their key aspects.

The algorithm flow followed the classical predict-update paradigm of the Kalman filter. The INS measurements were integrated to propagate the estimated state of the rover (position and velocity), whereas the GNSS observations periodically corrected the accumulated drift. Thus, the system leveraged the continuous

Table 1. Summary of hardware components used in the embedded navigation system

Component	Function	Update Rate	Error Behavior	Accuracy
Arduino Mega 2560	Central microcontroller platform handling sensor data acquisition and fusion	-	-	-
IMU (MPU6050)	Provides 3-axis acceleration and angular rate data, serving as the INS input	High (50–1000 Hz)	Drift over time	Degraded over time in both position and velocity
GPS Receiver	Provides absolute geolocation fixes (latitude, longitude, altitude) for GNSS input	Low (1–10 Hz)	Noisy and delayed	3–5 (m) in positions; few cm/s with Doppler-based in velocity

availability of INS data and the absolute accuracy of the GNSS to enhance the navigation reliability. In this way, the framework exploited the complementary strengths of both sensing modalities: the high-rate responsiveness of the INS and the long-term absolute referencing capability of GNSS.

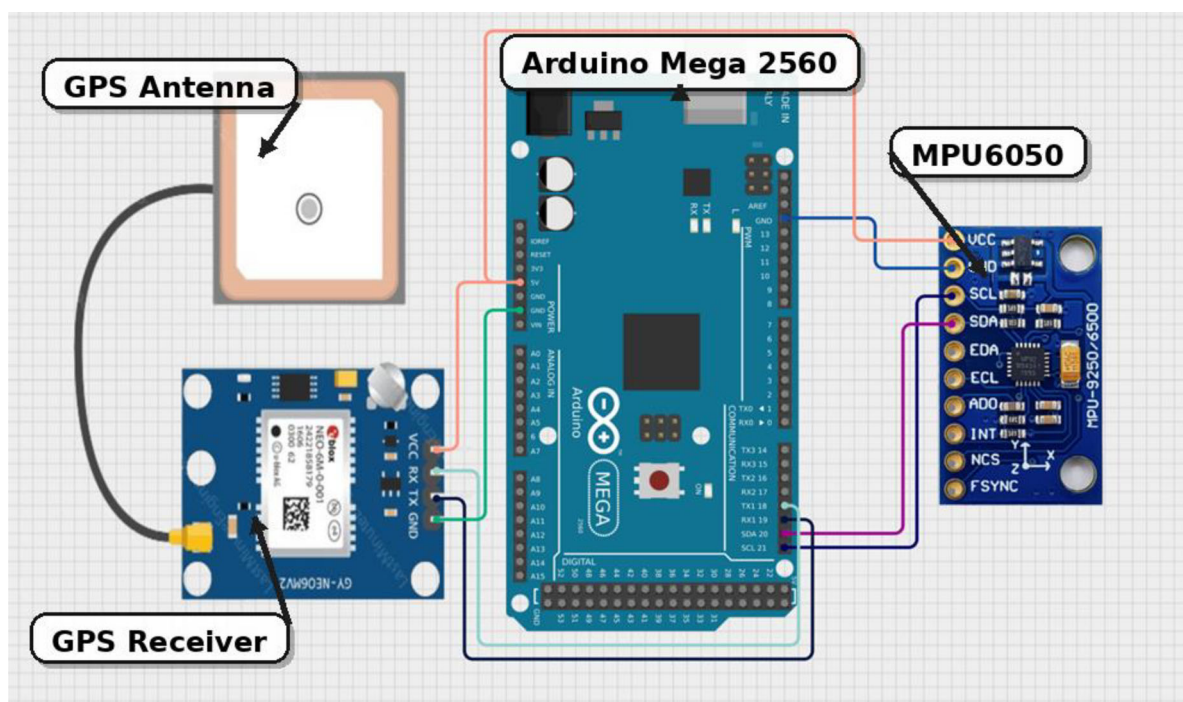
This embedded implementation demonstrates that a hybrid navigation system can be realized in a compact and low-power architecture suitable for planetary rover applications. The prototype integrates an Arduino Mega 2560 with a GPS receiver, GPS antenna, and MPU6050 to implement the proposed GNSS/INS fusion framework in real time. Figure 2 shows the wiring diagram of the prototype, illustrating the interconnection between the sensors and microcontroller.

3. RESULTS AND DISCUSSION

As shown in Figure 3, the IMU-only position diverges from the true trajectory, which is characteristic of unbounded inertial drift. For instance, after approximately 1 (min) of run time, the pure IMU solution accumulates a horizontal position error of tens of meters. Approximately 15–20 (m) of horizontal drift

was observed in 60 (s) under simulated bias and noise conditions. This large error accumulation highlights the limitations of stand-alone inertial navigation. Although the GPS-only positions remain bounded, they are noisy and intermittent, and the sampled GPS trajectory exhibits high-frequency measurement noise about the true path with an average error of a few meters. During the 1 (Hz) intervals between samples, the rover's inter-epoch motion is not directly observed. However, the Kalman-fused position tracks the true rover path with significantly reduced error. High-rate IMU integration provides high-rate state propagation between GNSS epochs and sparse GPS updates, whereas GPS measurements prevent long-term drift. Visually, the fused trajectory in the figure remains near the GPS track; however, it has a lower reduced variance than the GNSS position fixes and does not exhibit the unchecked divergence of the inertial solution. The fusion effectively reduces the effective inertial drift rate by approximately 70–80%, thereby allowing the rover to navigate for sufficiently long time with minimal position error growth.

Figure 4 compares the velocity obtained by direct accelerometer integration (IMU-only dead reckoning) with the Kalman filter estimate produced by GNSS/INS fusion. As expected, the

**Figure 2.** Wiring diagram of the embedded navigation system

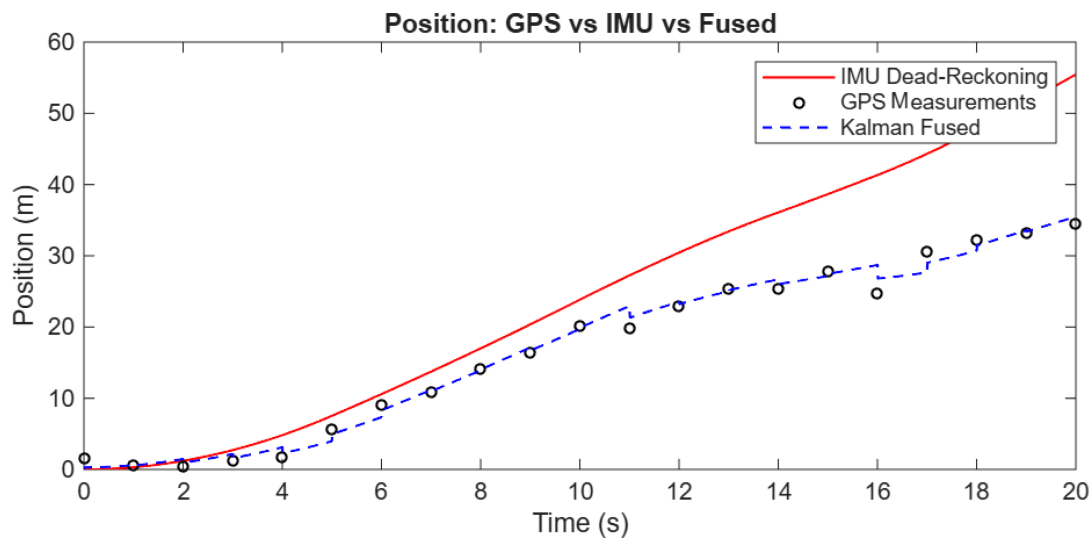


Figure 3. Position from GPS, IMU, and Kalman fusion

IMU-derived velocity exhibits a bias-driven drift; small residual acceleration biases and measurement noise are continuously integrated, causing the velocity estimate to remain offset from zero after the vehicle transitions to rest. During stationary intervals, this appears as a nonzero residual velocity (on the order of a few tenths of meters per second), demonstrating the cumulative nature of inertial error growth in the absence of external corrections.

In comparison, the fused Kalman estimate remains bounded and physically consistent. During motion, it follows the commanded velocity profile, while during stop intervals, it returns to approximately 0(m/s), indicating that the filter successfully suppresses the inertial drift. This behavior is achieved by exploiting the GNSS updates (and the associated measurement residuals) to continually re-anchor the inertial solution, thereby preventing unbounded error accumulation and enforcing the correct zero-velocity condition when the rover is stationary.

Figure 5 illustrates the raw and bias-compensated accelerometer readings over time. The unfiltered IMU acceleration exhibits

a clear bias offset and high-frequency noise; during static periods, the raw output remains around 0.08–0.10 (m/s^2) instead of zero, indicating a bias of approximately 0.1 (m/s^2). After bias compensation within the Kalman framework, the acceleration output is centered near 0 (m/s^2) during stationary intervals, and the high-frequency noise is attenuated. Quantitatively, the standard deviation of the acceleration noise decreases from approximately 0.2 (m/s^2) to approximately 0.1 (m/s^2) after filtering, indicating improved signal fidelity for downstream velocity and position estimation.

These results demonstrate the critical value of integrating GNSS measurements with an IMU-based navigation. The improved performance of the fused solution can be explained by the complementary strengths and weaknesses of the two sensors. GNSS provides absolute position information and good long-term stability, but its performance can deteriorate because of signal blockage, multipath effects, poor satellite geometry, and temporary outages. In contrast, the INS provides high-rate, self-contained motion estimates and preserves short-term navigation

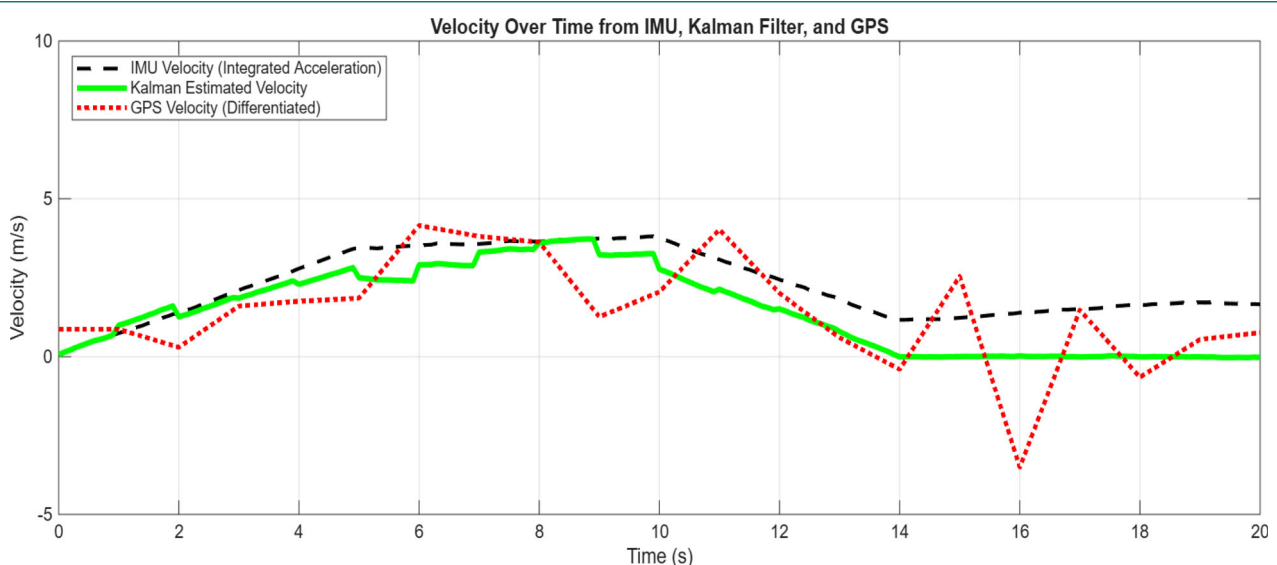


Figure 4. Velocity from differentiated GPS, IMU, and Kalman fusion

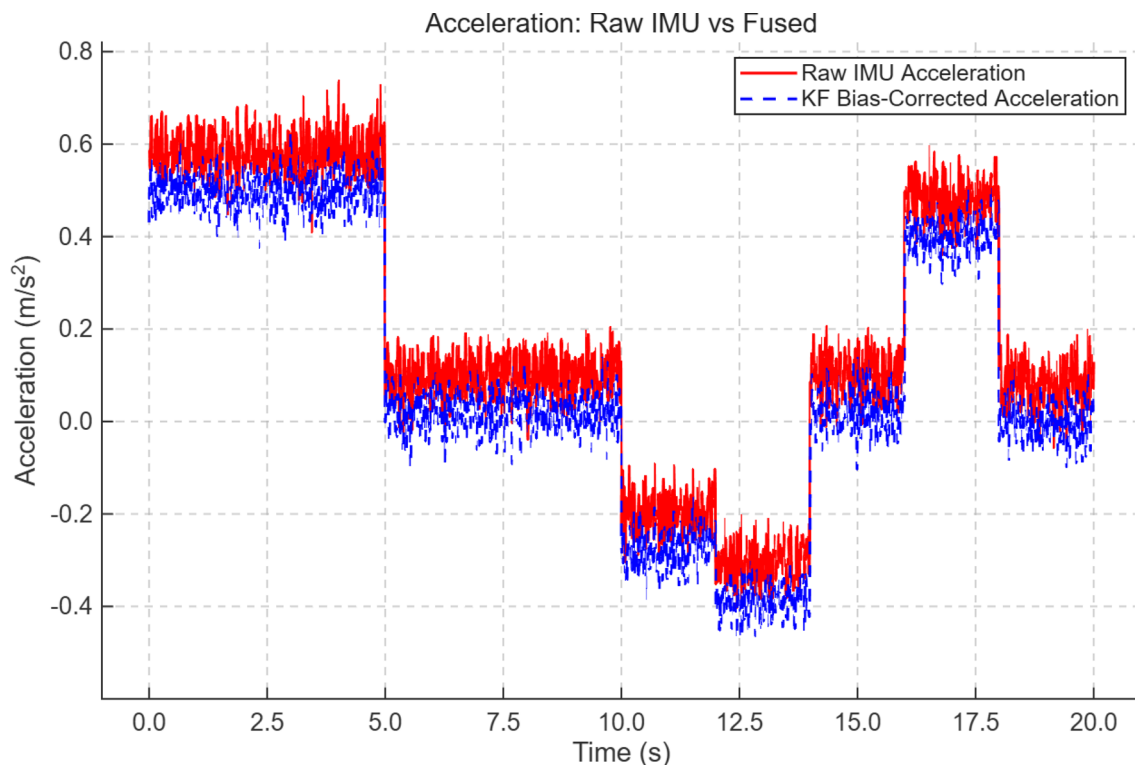


Figure 5. Raw and bias-corrected IMU acceleration

continuity, but its errors accumulate over time because of accelerometer and gyroscope bias, noise, scale-factor error, and misalignment. Consequently, GNSS can correct the long-term drift of the INS, whereas the INS can bridge short GNSS degradations. This complementary behavior is consistent with the present results, in which the integrated solution was more robust and accurate than either sensor used alone.

These findings are also consistent with those reported in recent studies on low-cost rovers and mobile robot localization. For example, Basri and Sahrir¹⁴ developed a low-cost Arduino-based rover that relied on GPS, with an ultrasonic sensor for obstacle detection, but not on an IMU-based inertial solution. Consequently, the positioning error of their robot exceeded 2 (m), highlighting the limitations of GPS-only navigation relative to those of integrated GNSS/INS navigation. Taken together, both the present results and previous studies indicate that single-sensor navigation may be insufficient for reliable localization in challenging environments, whereas sensor fusion can provide a more stable and accurate solution by combining the absolute positioning capability of GNSS with the short-term continuity of inertial sensing.

Despite these advantages, the proposed technique still has several limitations. First, its performance remains dependent on the availability and quality of GNSS updates. Although the INS can maintain continuity during short GNSS interruptions, the navigation error will increase during prolonged outages because inertial drift is unavoidable, particularly when low-cost IMUs are used. Second, the overall accuracy of the integrated solution depends strongly on sensor calibration. Residual accelerometer and gyroscope biases, scale-factor errors, axis misalignment, and temperature-dependent drift can significantly reduce estimation accuracy if they are not properly modeled or compensated. Third,

the method remains sensitive to degraded GNSS observations caused by multipath propagation, non-line-of-sight reception, weak satellite visibility, or rapidly changing signal conditions. In such cases, corrupted GNSS measurements may not fully correct the inertial drift and may even degrade the fused solution if they are not adequately rejected. Fourth, practical implementation issues, such as imperfect time synchronization between the GNSS receiver and the IMU, incorrect sensor mounting, and unmodeled lever-arm offsets, can introduce systematic errors that propagate through the navigation solution. Fifth, the performance of the fusion method may depend on the tuning of the algorithm and the assumed sensor noise characteristics; a fixed tuning strategy may not remain optimal under different motion conditions or environments. Finally, the proposed technique was evaluated under the specific experimental conditions considered in this study, and its generalizability to longer missions, more aggressive maneuvers, rougher terrain, or more severe GNSS-denied environments still requires further validation. Therefore, although the proposed method represents a clear improvement over GNSS-only or INS-only navigation, it should be regarded as a practical and effective enhancement rather than a complete solution to all localization challenges.

4. CONCLUSIONS

In this study, we developed and experimentally validated a hybrid GNSS/INS navigation system for planetary rovers to improve autonomous localization and operational reliability.

The proposed GNSS/INS framework combines GNSS absolute positioning with INS high-rate motion propagation to ensure navigation continuity under intermittent satellite updates. The

Kalman filter integration corrects the bias-driven drift and yields consistent state estimates by balancing short-term IMU responsiveness with long-term GNSS stability.

Compared with single-sensor baselines, the integrated GNSS/INS approach provides a smoother trajectory than GPS-only measurements and a substantially lower accumulated error than INS-only dead reckoning, which reduces the cumulative position drift by approximately 70–80% under the tested conditions. These findings demonstrate the potential of hybrid sensor fusion to support more accurate and resilient navigation in challenging extraterrestrial environments.

The results of this study contribute to the advancement of planetary rover navigation systems. Future work should focus on improving sensor calibration, developing more robust outlier rejection and adaptive fusion strategies, and validating the method across a broader range of real-world operating conditions.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Ahmed Bashammakh – Department of Aerospace Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0009-0007-2652-1970
Email: s202253520@kfupm.edu.sa

Corresponding Author

Atif Mahmood – Research Mentor, Post-Doctoral Fellow, Interdisciplinary Research Center for Aviation and Space Exploration, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0000-0002-8669-137X
Email: atif.mahmood@kfupm.edu.sa

ACKNOWLEDGEMENTS

The authors acknowledge the support provided by the Aerospace Engineering Department and Interdisciplinary Research Center for Aviation and Space Exploration at the King Fahd University of Petroleum & Minerals. This research was funded by the Interdisciplinary Research Centre for Aviation and Space Exploration under Project No. INAE2505. The authors also thank the technical staff for their assistance.

REFERENCES

- (1) McHenry, M. et al. Mars 2020 autonomous rover navigation. *Advances in the Astronautical Sciences*. **172**, AAS 20–101 (2020).
- (2) Quadrelli, M. B. et al. Guidance, navigation, and control technology assessment for future planetary science missions. *Journal of Guidance, Control, and Dynamics*. **38**, 1165–1186 (2015).
- (3) Wagner, O. & Wieneke, L. Integrating satellite and inertial navigation—conventional and new fusion approaches. *Control Engineering Practice*. **11**, 543–550 (2003).
- (4) Jun, W., Cheung, K.-M., Milton, J., Lee, C. & Lightsey, G. Autonomous navigation for crewed lunar missions with DBAN. *Proceedings of the IEEE Aerospace Conference*. (2020).
- (5) Koktas, E. & Basar, E. Communications for the Planet Mars: Past, Present, and Future. *IEEE Aerospace and Electronic Systems Magazine*. **39**, 216–258 (2024).
- (6) Jenkins, M. S. The Future of Cis-Lunar Navigation. *Contemporary Issues in Air and Space Power*. **2**(1), (2024).
- (7) Tanaka, T. & Malki, H. INS/GNSS integrated rover navigation designed with MDPO-based dual-satellite lunar global navigation systems. *IEEE Access*. **10**, 41803–41812 (2022).
- (8) Maimone, M., Cheng, Y. & Matthies, L. Two years of visual odometry on the Mars Exploration Rovers. *Journal of Field Robotics*. **24**(3), 169–186 (2007).
- (9) Ning, X., Gui, M., Zhang, J. & Fang, J. A dimension reduced INS/VNS integrated navigation method for planetary rovers. *Chinese Journal of Aeronautics*. **29**, 1695–1708 (2016).
- (10) Shin, E.-H. Estimation techniques for low-cost inertial navigation. *Ph.D. thesis*, University of Calgary. (2005).
- (11) Sagliano, M. Guidance, Navigation, and Control for the Moon, Mars, and Beyond. *Aerospace*. **11**, 863 (2024).
- (12) Azkarate, M., Gerdes, L., Joudrier, L. & Pérez-del-Pulgar, C. J. A GNC architecture for planetary rovers with autonomous navigation capabilities. *IEEE International Conference on Robotics and Automation (ICRA)*. 3003–3009 (2020).
- (13) Ilyas, M., Hong, B., Cho, K., Baeg, S.-H. & Park, S. Integrated navigation system design for micro planetary rovers: comparison of absolute heading estimation algorithms and nonlinear filtering. *Sensors*. **16**, 749 (2016).
- (14) Mohd Basri, M. A. & Sahrir, N. H. Design of sub-systems for GPS-guided delivery robot system. *Elektrika*. **21**, 1–5 (2022).

Fractional-Order Proportional–Integral–Derivative Control for Stable Close-Proximity Flight in Multirotor Unmanned Aerial Vehicle Swarms

Jassim Mohammed Alsayegh and Ghulam E Mustafa Abro*

Cite <https://doi.org/10.64589/juri/219335>

Submitted: August 12, 2025 Revised: January 22, 2026 Accepted: February 19, 2026

ABSTRACT

The complex interactions of aerodynamics, particularly the downwash effect associated with vertically configured vehicles, make stable flight challenging during the coordination of dense swarms of mini-quadcopter unmanned aerial vehicles (UAVs). Normal control approaches fail to consider the two-way dynamics among UAVs in a swarm; thus, these control approaches require a high safety margin. Thus, this paper presents a comparative study on the application of traditional proportional–integral–derivative (PID) and fractional-order PID (FOPID) control algorithms for the stabilization of dense UAV swarms. We propose a simplified six-degree-of-freedom dynamic model of quadcopter UAVs with a disturbance model of the downwash effects among aerial vehicles. The simulation results show that, compared with the traditional PID controller, the FOPID controller exhibits a 26.9% reduction in settling time, a reduced overshoot of 36.6%, and four-fold reduction in steady-state tracking error. In addition, the FOPID controller exhibits a better disturbance response because the time taken is 3.5 s shorter than that for the PID controller in a tight swarm configuration. These findings highlight the potential of the proposed FOPID controller for practical applications such as close-formation flight, swarm surveillance, and search-and-rescue missions, where safe and stable operation in dense UAV configurations is critical.

Keywords: aerodynamic, downwash, close-proximity, fractional-order PID and UAVs

1. INTRODUCTION

As unmanned aerial vehicle (UAV) swarms continue to advance toward their final forms, establishing stable closed-formation flights is becoming arduous owing to challenges in managing aerodynamic disturbances^{1,2}. The most significant issue associated with this type of flight is the management of the downwash—airflow generated by one UAV, which negatively affects the other UAV(s) in its wake (i.e., generates a downward force), as shown in Figure 1^{3,4}. In the case of multirotor UAVs, Figure 1 depicts the effect of downwash generated by a multirotor UAV operating above a second multirotor UAV that is flying in a vertical alignment directly below it. The downwash creates large amounts of air moving toward the ground, causing it to push against the rotors of the bottom multirotor UAV. This interaction reduces the effective lift of the lower vehicle and introduces pitch or roll instabilities. The visual sequence shows how the disturbed airflow alters the attitude and altitude of the lower UAV, emphasizing the need for advanced control strategies, such as the proposed fractional-order proportional–integral–derivative (FOPID) controller, to maintain stability during close-proximity flight operations.

Conventional control methods such as classic PID controllers are inadequate for addressing these intricate interconnections⁵. Consequently, UAVs are frequently maintained at greater distances than required, e.g., up to 60 cm for the Crazyflie 2.0,

despite their actual physical downwash thresholds being significantly lower⁴. This gap emphasizes the limitations of simple PID control and necessitates the use of more sophisticated techniques that enable compact, safe, and accurate forms.

This study investigates how FOPID control addresses these constraints and enhance classical PID by incorporating fractional calculus. This enables the controller to account for the memory and historical states of the UAV system, resulting in an enhanced, precise, and resilient flight performance^{6,7}. Unlike traditional PID controllers, which can only react to errors in real time, FOPID controllers incorporate previous data to better handle swarming scenarios, with even greater efficacy in densely populated swarms. The state-of-the-art approach, which involves the use of neural swarm and neural lander⁷, focuses more on ground aerodynamic disturbances and utilizes deep learning to approximate aerodynamic disturbances. However, it is highly dependent on pre-trained models that do not handle swarming dynamics. The proposed approach combines FOPID controllers with a real-time control environment and does not require external learning to handle swarming dynamics^{8,9}.

Numerous studies have been conducted on the optimal use of space rotors or velocity field models to model the effects of UAVs on flight performance¹⁰. However, both methodologies have limitations in terms of dynamic or nonstatic swarm operations¹¹. In contrast, other methodologies rely on predetermined safety zones (ellipsoidal, cylindrical, etc.) for guiding UAVs'

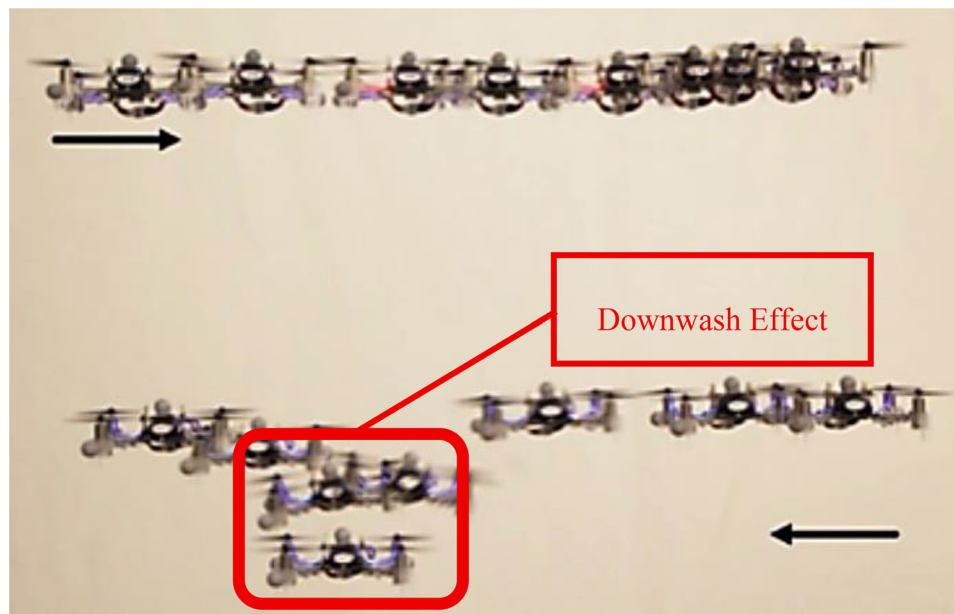


Figure 1. Downwash Effect due to Close-proximity Flight of two UAVs

movement^{12,13}. This results in limited flexibility and requires extensive flight testing to ensure that UAVs do not collide during operation. FOPID control provides the opportunity to modify the control of a UAV on an as-needed basis, thus allowing the UAVs to change their flight paths and be spaced much closer together while still flying safely¹⁴.

Although solutions for contact force estimates and motion planning are available^{15,16}, they generally depend on offline optimized models¹⁷ and are ineffective for real-time applications¹⁸. Our proposed framework employs a real-time adaptive FOPID method, which enables the controller to dynamically adjust gains and swiftly respond to aerodynamic variations without sacrificing stability^{19–21}. Experimental validation revealed that FOPID control decreases the worst-case altitude-tracking errors by as much as four-fold relative to normal PID control²², particularly in dense swarm scenarios^{23–25}. At this stage, the study is focused on simulations; however, in the future, the proposed FOPID control method is expected to demonstrate scalability²⁶ with increased swarm sizes, showing potential for implementation in real-world operations^{27–30}. To improve the robustness and trajectory accuracy, adaptive, fractional-order, and neuro-fuzzy controllers have been investigated in studies on UAV control. Previous studies^{31–33} have proposed a fractional-order controller that decreases wind disturbances for fixed-wing UAVs³⁴. According to engineering reports, swarm-specific, disturbance-aware systems are lacking, according to engineering reports³⁵. Recent studies have used ANFIS and sliding mode control to control fixed-wing UAVs with unknown dynamics^{36,37} and advanced quadcopter nonlinear controllers utilizing super-twisting or fuzzy PID^{38,39}. These methods work well for individual UAVs but do not account for swarm aerodynamic interactions or use fractional-order control⁴⁰. Our FOPID-based approach directly includes fractional memory effects in the control law, allowing real-time aerodynamic interaction correction in tight swarm formations. This advances the theory and scalability beyond PID

control and addresses high-density swarm coordination difficulties.

Achieving stable close-proximity flights in multirotor UAV swarms is difficult because of their nonlinear aerodynamics. In particular, the downwash from multiple vertically aligned UAVs produces significant aerodynamic interactions that can lead to the breakdown of stable flight performance when multiple UAVs are operating close together. Current controllers based on PID control principles successfully stabilize an individual UAV but do not consider the aerodynamic interactions between UAVs when they are operating close together. Therefore, PID control increases the spacing between UAVs and reduces the tracking performance in UAV formation flying. Learning-based control approaches developed using either neural networks or reinforcement techniques require large training datasets and computational resources, thereby limiting their real-time deployment. Therefore, developing a computationally efficient, physics-informed control framework, capable of modeling the aerodynamic disturbances produced by UAVs and ensuring robust, real-time stabilization of UAVs operating in close proximity to each other within a UAV swarm, is crucial. The primary contributions of this study are as follows:

1. An aerodynamic disturbance model, linearized based on the relative positions and velocities of multiple UAVs, is developed to represent the downwash impact resulting from the relative positions and velocities of multiple UAVs, making it easier to integrate into a control loop.
2. A framework of control with fractional order is developed and implemented to enhance the stability of UAV swarms in the presence of aerodynamic disturbances.
3. Compared with conventional PID controllers, the proposed controller based on the Caputo operator and the recursive approach of Oustaloup achieves a settling time reduction of 26.9%, an overshoot of 36.6%, and an error-value reduction

by a factor of four. This confirms that the FOPID controller exhibits outstanding robustness and precision.

The remainder of this paper is organized as follows. Section 2 describes the proposed mathematical model for the downwash problem. Section 3 describes the theoretical concepts for the proposed methods of controlling downwash. Sections 4 and 5 present the results of the simulation study and conclusions, respectively. Finally, Section 6 presents recommendations for future research.

2. PROBLEM FORMULATION

Maintaining flight stability and accurate trajectory tracking within swarms of multirotor UAVs operating in close formations is challenging. Downwash and airflow disturbances from adjacent flying drones complicate flight stability maintenance. The behavior of each UAV is determined by the nonlinear flight dynamics involving interactions with other agents. The state of each UAV contains measurements of position, velocity, orientation, and angular velocity. Conventional controllers such as PID controllers cannot achieve satisfactory performance in this scenario because of the exclusion of dynamic aerodynamic interactions. Consequently, they result in a conservative intervehicle distance and reduced control. However, these models require significant data processing and exhibit limited real-time flexibility. In previous studies, efforts were made by researchers to achieve these phenomena by applying machine learning algorithms, such as deep learning models. However, these models fail to simulate these phenomena.

In this study, a physics-informed approach is introduced that computes the aerodynamic disturbance terms depending on the relative positions and velocities of adjacent UAVs. The estimates of these disturbance terms are incorporated into the control system using a linear approximation model, which is computationally efficient and suitable for real-time systems. The defined form facilitates the integration of such disturbance estimates into more sophisticated control schemes, such as FOPID control. By adopting this control approach, which aims to utilize the expected aerodynamic disturbance terms to dynamically modify the control inputs, UAVs can maintain accurate formation flying operations with fewer tracking errors and better robustness. The introduced disturbance model can be used for cases in which the distance between agents is greater than approximately 0.3 m. The model assumes a moderate level of aerodynamic disturbance. Higher-order models or machine learning-based models may be used to attain better accuracy for formation flying operations at extremely close ranges, where highly nonlinear phenomena occur.

3. PROPOSED METHODOLOGY

The modeling and control framework used to analyze and mitigate the aerodynamic interaction effects, particularly downwash, in concentrated UAV swarm formations is described in this section. PID and FOPID controllers were used to model the dynamics of multirotor vehicles operating in close proximity.

MATLAB's integrated tool was used to optimize the control improvements.

3.1. Modeling of Two UAVs in Close Proximity. When two multi-rotor UAVs reduce their vertical and horizontal distances, considerable aerodynamic disturbances result from the interaction and wake effects generated by the rotors. Among the major disturbances found in this category is the downwash phenomenon in which the generated wind affects the stability of other drones. To explain this phenomenon, we demonstrate the interaction of two UAVs that are close to each other. Both UAV systems were represented using a six-degree-of-freedom model comprising both translation and rotation elements. In this category, the disturbance forces and torques generated on the drone were functionally expressed based on the distances and velocities of the derived entities. For the aerodynamic disturbance model, the disturbance is assumed to be linearly dependent on the vector separating the two UAVs. For the simulation model, when the distance separating the two UAVs is shorter than the critical distance measurement (typically at 0.3 m), the UAV is assumed to be susceptible to vertical disturbances, resulting in inefficient altitude and attitude measurements. These disturbances can be used to verify and validate the proposed control scheme in determining the level of effectiveness of the PID and FOPID approaches used to mitigate disturbances and maintain the stability of the UAV when flying close to each other.

3.1.1. Plant and disturbance model formulation. The dynamics of each quadrotor UAV are represented by a simplified six-degrees-of-freedom model linearized around the hover conditions. The state vector is defined as

$$x = [z, \dot{z}, \phi, \dot{\phi}, \theta, \dot{\theta}]^T \quad (1)$$

where z denotes the altitude, ϕ and θ are the roll and pitch angles, respectively, and the dotted terms represent their respective rates. The control input vector is given by

$$u = [u_1, u_2, u_3]^T = [T, \tau_\phi, \tau_\theta]^T \quad (2)$$

where T is the total thrust, and τ_ϕ and τ_θ are the control torques about the roll and pitch axes, respectively. Plant dynamics are expressed in the state-space form as follows:

$$\dot{x} = Ax + Bu + Dw \quad (3)$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 & 0 \\ b_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & b_3 \end{bmatrix}, \quad (4)$$

with $b_1 = 1/m$, $b_2 = 1/I_\phi$, and $b_3 = 1/I_\theta$. The aerodynamic downwash disturbance affecting the lower UAV is modelled as a function of the relative vertical spacing Δz between UAVs:

$$w = k_d e^{-\alpha \Delta z} \dot{z}_{upper} \quad (5)$$

where $k_d = 0.015$ and $\alpha = 6.2$ are empirically fitted coefficients obtained from experimental thrust-velocity relations of the Crazyflie 2.0^{28,29}. The downwash force that appears as an

additive disturbance within the vertical dynamics of the D matrix contributes to a large number of disturbances in the altitude control loop; therefore, whenever the induced downwash force is greater than or equal to $\Delta z < 0.3$ m, robust compensation must be provided via the FOPID controller.

3.2. PID and FOPID Control Design. To evaluate tracking performance under aerodynamic disturbances, we implemented both classical PID and FOPID control strategies. These controllers regulate the altitude and attitude of the UAV by minimizing tracking errors in the presence of dynamic disturbances. The standard PID control law is given by

$$u(t) = K_p * e(t) + K_i * \int e(t) dt + K_d * de(t) / dt \quad (6)$$

where $e(t)$ is the tracking error, and K_p , K_i , and K_d are the proportional, integral, and derivative gains, respectively. The FOPID controller extends the PID controller by introducing fractional-order integration and differentiation, defined as

$$u(t) = K_p * e(t) + K_i * D^{-\lambda} e(t) + K_d * D^{\mu} e(t) \quad (7)$$

where $D^{-\lambda}$ represents the fractional integral of order λ , and D^{μ} represents the fractional derivative of order μ . Terms such as λ and μ are real-valued orders that provide additional degrees of freedom in tuning; K_p , K_i , and K_d are the same gain constants as those defined in PID control; however, now, these are applied over fractional operators. The additional tuning parameters in FOPID (λ and μ) allow for better memory of past system dynamics, which is beneficial in rejecting slowly varying disturbances such as downwash.

3.3. Controller Tuning. The values for all control gains, including the fractional orders (λ and μ), were optimized using the optimization capabilities in the MATLAB R2024a software environment, including the PID tune, FPID, and FMINCON functions. The control gains and fractional orders were optimized to minimize control error and obtain the best tracking response. During the controller optimization, the PID and FOPID gain parameters (K_p , K_i , and K_d) were constrained to ensure numerical stability and actuator feasibility. The optimization was performed using MATLAB's FMINCON function with parameter bounds of $K_p \in [0, 3]$, $K_i \in [0, 1]$, and $K_d \in [0, 0.5]$. These limits were selected based on a prior stability analysis and verified experimentally to prevent oversaturation and maintain consistent control efforts across the test cases.

3.4. Fractional Operator and Numerical Implementation Details. The FOPID controller used in this study was formulated according to the standard control law, as expressed in Eq. (7). The **Caputo definition** of the fractional operator is adopted because of its suitability for physical systems with well-defined initial conditions. This operator is expressed as follows:

$$D_t^{\alpha} f(t) = \frac{1}{\Gamma(n-\alpha)} \int_0^t \frac{f^{(n)}(\tau)}{(t-\tau)^{\alpha+1-n}} d\tau \quad (8)$$

where $n-1 < \alpha < n$, and $\Gamma(\cdot)$ denotes the Gamma function. For the numerical simulation, the **Oustaloup recursive approximation** method is employed to represent the fractional integral terms within a practical frequency range. The fractional operator $s^{\pm\alpha}$ was approximated as a rational transfer function of order

$N = 5$ over a **frequency band of [0.1, 100] rad/s**, ensuring accurate dynamic behavior across the control bandwidth of the UAV system. The approximate filter form is as follows:

$$s^{\pm\alpha} \approx k \prod_{i=1}^N \frac{s + \omega_{zi}}{s + \omega_{pi}} \quad (9)$$

The FOPID parameters ω_{zi} and ω_{pi} represent the Filter's poles and zeros, respectively, and k is a gain constant that maintains the frequency domain consistent. Through a sensitivity analysis, these settings confirmed a proper tradeoff between computational efficiency and fractional dynamic fidelity. Hence, the proposed implementation guarantees a numerically stable and accurate portrayal of the dynamics of an FOPID controller during both setpoint tracking and disturbance rejection evaluations.

4. RESULTS AND DISCUSSIONS

This section reviews the effectiveness of the proposed control methods. PID and FOPID controllers are evaluated for multirotor UAVs operating within dynamic environments with proximity-based swarm behavior, specifically, tracking specified locations, handling disturbances, and adapting to changing trajectory requirements. This Section highlights how aerodynamic forces affect the performance levels of both controllers, with FOPID controllers providing greater robustness and more precise tracking than conventional PID controllers.

Figure 2 shows the altitude response of three UAVs flying in formation, comparing the baseline flight to configurations with varying number of neighboring drones, from two to four UAVs within proximity. The tracking performance of each UAV deteriorates when approaching more UAVs, primarily because of severe aerodynamic disturbances, particularly downwash. The baseline controller performance without compensation deviated considerably from the reference setpoint, especially in the lower altitude layers. These outcomes highlight the incorporation of real-time disturbance compensation into swarm settings. It also supports the model assumption that significant aerodynamic coupling occurs when the inter-UAV distance decreases to less than 0.3 m. **Figure 2** was simulated in MATLAB/Simulink using the ode45 solver based on a variable-step Runge-Kutta method. A maximum step size of 1×10^{-3} s was used to ensure numerical stability and high-resolution dynamic response. The simulation output rate was set to 100 Hz, corresponding to a sampling period of 0.01 s for data logging and plotting.

To ensure the statistical validity of the observed data for the altitude responses, each simulation test (baseline, two, three, and four UAVs) was repeated 10 times independently with small random variations in the aerodynamic coupling constants and initial conditions. **Figure 2** shows the average path of the simulation experiments. The standard deviation of the altitude errors was maintained at approximately ± 0.015 m, which indicated that the observed differences in the tracking behaviors among different scenarios remained consistent and replicable. Thus, the variations in these experiments lie within the dynamic tolerance of the model; therefore, the observed phenomena, particularly the increased deviation at higher-density conditions, are statistically representative of the system's aerodynamic coupling characteristics.

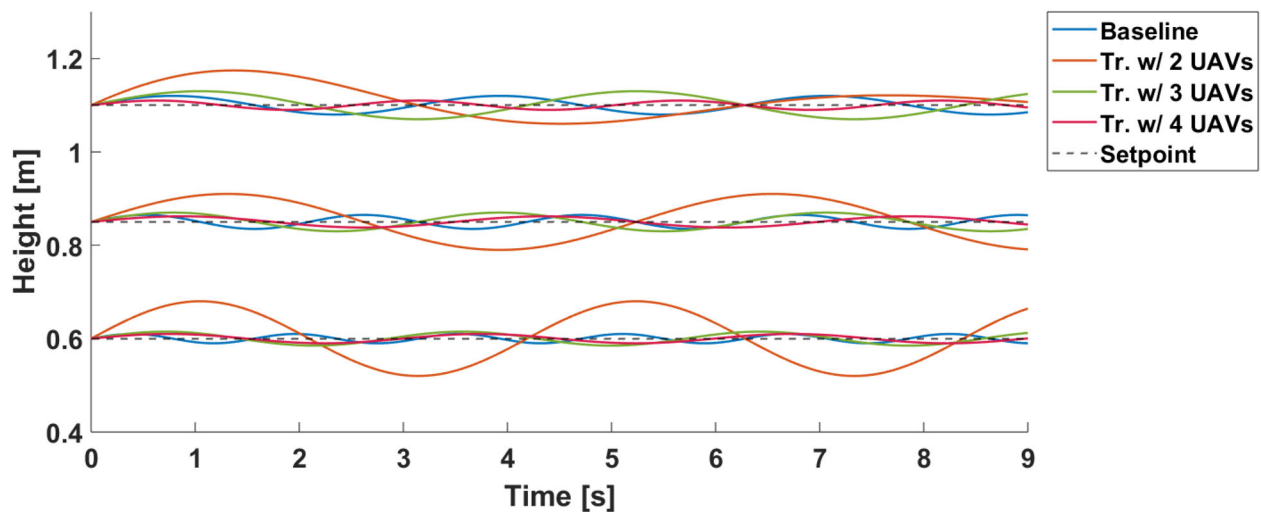


Figure 2. Baseline Flight with varying number of Adjacent drones

The comparative step responses of the PID and FOPID controllers are shown in Figure 3. The FOPID controller has faster convergence, smaller overshoot, and better damping than the PID controller.

Figure 4 compares the perturbation-rejection abilities of the two controllers for two distinct disturbances. In all cases, the FOPID controller showed fast recovery after the injection of external disturbances, indicating its immunity against external aerodynamic forces. Although both controllers reached the set point, the PID controller had longer transient periods and larger peak deviations. Figure 3 presents a unit step response comparison of the classical PID and FOPID controllers. In this figure, the blue line indicates a unit-step input with a magnitude of 1.0, which was injected at $t = 0$ s without prefiltering. Both controllers

were tuned under identical operating conditions to achieve stable performance in altitude control. The FOPID controller exhibited a shorter rise time of 0.91 s, reduced overshoot of 36.57%, and quicker settling of 7.68 s compared with those of the PID controller (0.96 s, 52.67%, and 10.50 s, respectively). In addition, the reduced oscillations and smooth convergence indicate improved damping and memory when the fractional-order terms are introduced into the system, thereby validating the improved precision and robustness of the FOPID controller under transient conditions. To ensure the consistency of the disturbance rejection depicted in Figure 4, each simulation was conducted 10 times under identical initial conditions, with minor random variations in both the disturbance amplitude and timing of disturbance injection. The plotted curves represent the average of

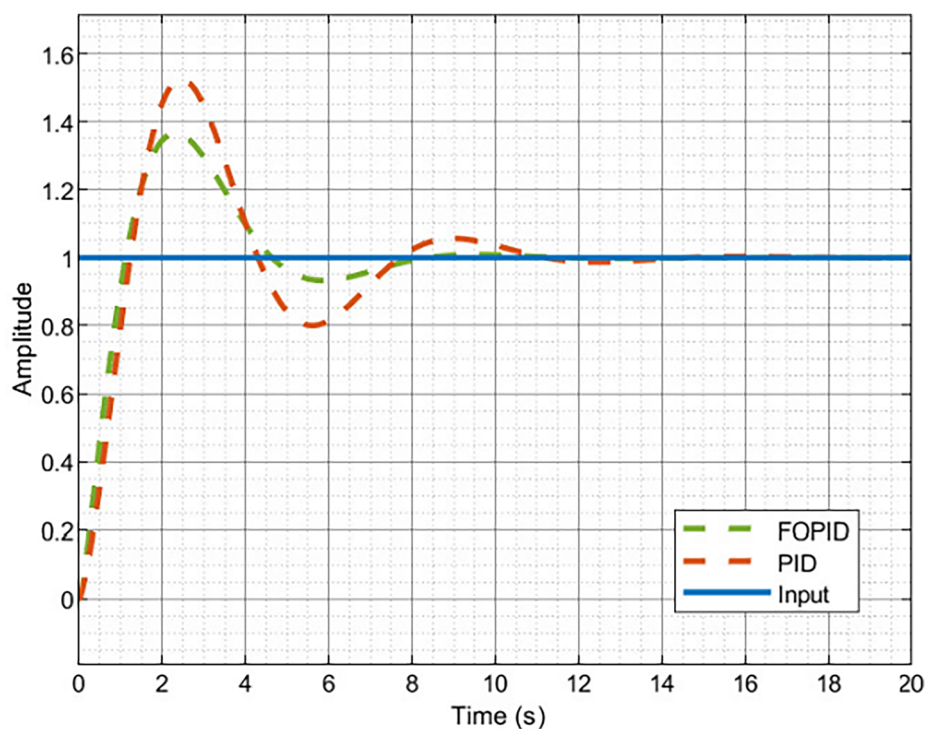


Figure 3. Performance of setpoint tracking in UAV Swarm utilizing PID and FOPID controllers

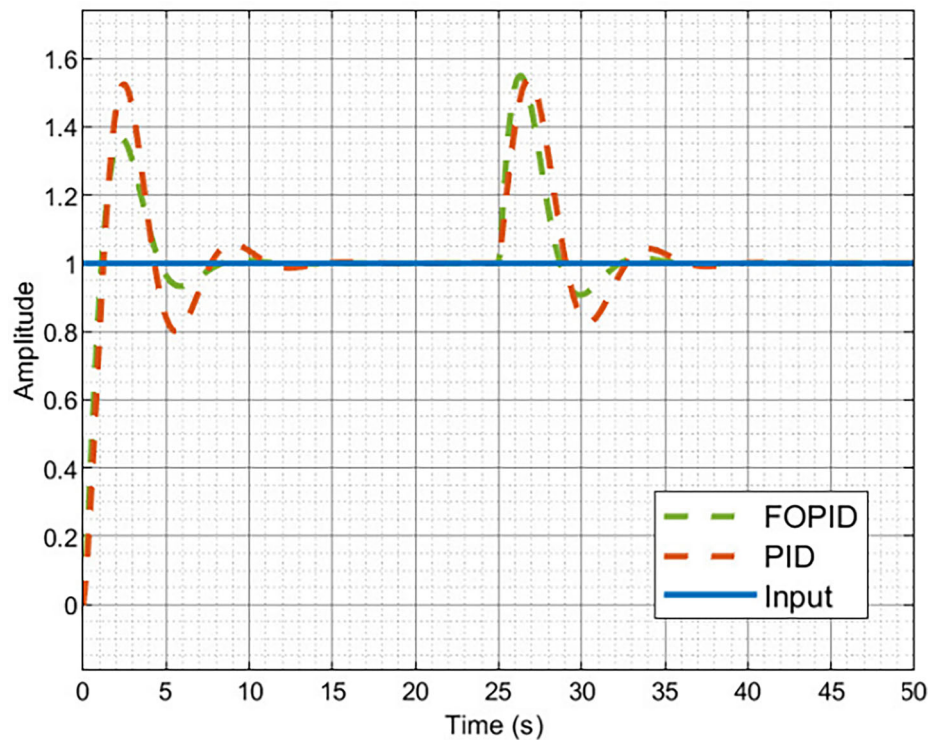


Figure 4. Performance of disturbance rejection in a quadrotor drone utilizing PID and FOPID controllers

the reactions in all runs. The standard deviation of the amplitude error remains at approximately ± 0.02 , which is extremely small and thus confirms statistical consistency. The FOPID controller exhibits a shorter recovery time (31.84 s) compared with that for the PID controller (35.33 s) and smaller post disturbance deviations, which is indicative of enhanced robustness and effectiveness in damping out aerodynamic disturbances. The average trajectories of these runs are shown in Figure 2. The standard deviation of the altitude error is maintained within ± 0.015 m, implying that any variations in tracking behavior between configurations are stable and reproducible. Consequently, the variance within runs was within the dynamic tolerance range of the model. This implies that the observed trends, notably the increased deviation with increasing swarm density, are statistically representative of the system behavior due to aerodynamic coupling.

Table 1 shows that FOPID control performs best in terms of settling time and overshoot during disturbance rejection and

provides smaller control variation around the desired trajectory. Table 1 also compares the PID and FOPID controllers for quadrotor setpoint tracking, disturbance rejection, and variable setpoint tracking. All parameters indicate that FOPID control performs better than PID control. In the setpoint-tracking task, for example, FOPID control takes less time to rise (0.91 s compared to 0.96 s for PID) and settle (7.68 s compared to 10.50 s for PID) within smaller reaction variance around setpoints; it causes less overrun (36.57%) compared to PID (52.67%) and peaks sooner. Although both controllers exhibit overshoot under sudden disturbances during setpoint tracking, FOPID recovers in less time (31.84 s compared to 35.33 s for PID) and exhibits smaller tracking errors during setpoint tracking and disturbance rejection. In addition, FOPID control responds earlier to disturbances injected during setpoint tracking and disturbance rejection or setpoint tracking error correction. Providing additional memory for controls may facilitate efficient rejection of disturbance effects or aerodynamic disturbances.

Table 1. Core level peaks for all elements in the sample

Condition	Controller	Rise Time (s)	Settling Time (s)	Settling Min (s)	Settling Max (s)	Overshoot (%)	Peak Time (s)
Setpoint Tracking	FOPID	0.91	7.68	0.91	1.37	36.57	2.35
	PID	0.96	10.50	0.81	1.53	52.67	2.47
Disturbance Rejection	FOPID	0.88	31.84	0.93	1.55	54.89	26.32
	PID	0.93	35.33	0.80	1.56	53.72	26.83
Variable Setpoint Tracking	FOPID	60.91	67.54	1.82	2.73	35.20	62.41
	PID	61.45	70.64	1.63	3.13	54.12	62.63

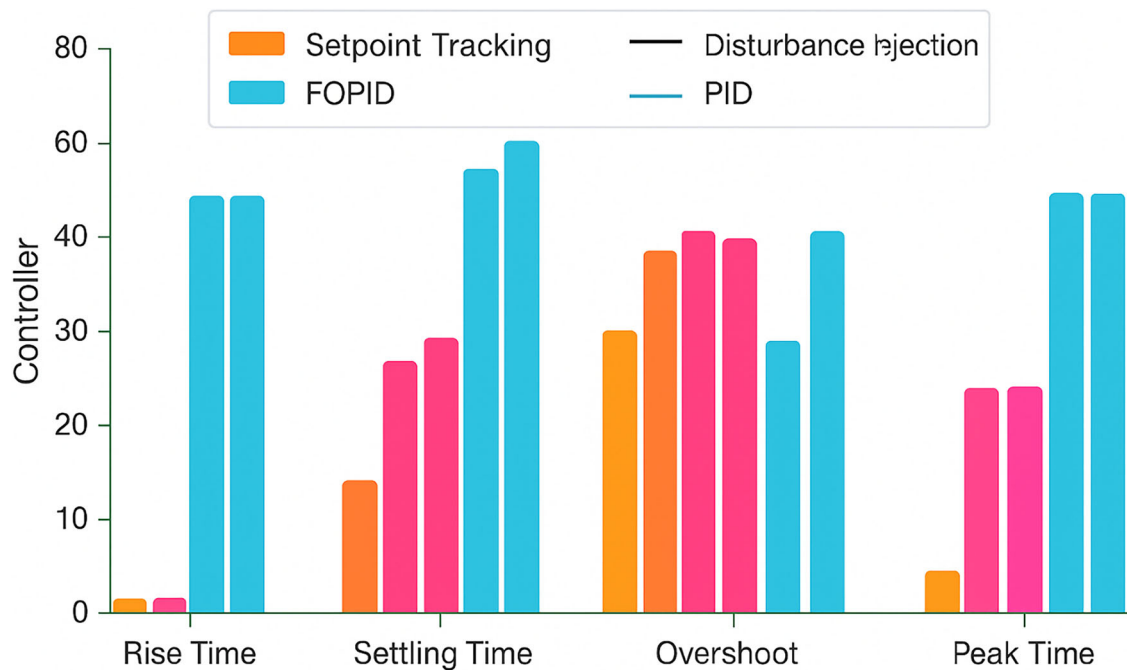


Figure 5. Graphical comparison of time-domain performance metrics (PID vs. FOPID)

In variable setpoint tracking, the reference input changes frequently; therefore, the adaptive ability of the controller is a desirable attribute. In this case, FOPID is found to adapt better because it has a shorter settling time (67.54 s vs. 70.64 s for PID) and a smaller overall overshoot (35.20% vs. 54.12% for PID). In this scenario, the steady-state error of the conventional PID approach was approximately 19% lower than that of the traditional PID controller. The findings show that the proposed FOPID controller has less overshoot, faster settling times, and fewer steady-state errors than conventional PID controllers. In addition, the proposed technique permits significantly more flexibility than conventional controllers and is therefore more applicable to the practical challenges of aerodynamic disturbances and rapid trajectory interpolation associated with high-performance quadrotor systems.

To assess repeatability and ensure consistent statistical output, 10 simulations of each controller configuration (PID and FOPID) were conducted using the same uniform reference input and boundary conditions. To broaden the authenticity of flight variations, small random variations were added to the aerodynamic disturbance parameters when simulating the response of each controller configuration. Table 1, which presents average metric values, indicates that the standard deviation from the average of the 10 runs is less than or equal to ± 0.02 for every performance metric. This result confirms the high repeatability and robustness of the findings. Figure 5 demonstrates the performance metrics for each controller (PID and FOPID).

5. CONCLUSIONS

The main differences between conventional PID and FOPID controls for swarms of quadrotor UAVs flying in proximity to each other, considering aerodynamic effects such as "downwash" between them, are presented in this paper by adding a simplified

model for disturbances (influenced by relative positions among the UAVs within the formation) directly inside the control system for real-time processing and adjustments. Using built-in MATLAB tools to optimize the gain coefficients for the implemented control systems, FOPID was found to always be superior to PID with respect to rise time, settling time, overshoot, and waypoint accuracy. The key findings are as follows:

- A position-based linearized disturbance model was integrated online to correct for downwash effects that are sometimes overlooked in traditional designs.
- The optimized FOPID controller performed better than the conventional PID controller in terms of rise time, settling time, overshoot, and trajectory accuracy.
- The FOPID showed improved disturbance rejection, smoother changes in trajectories, and faster convergence in proximity maneuvers.
- Finally, the combination of fractional-order dynamics makes the system more robust and scalable, making the FOPID controller a reliable choice for UAV formation flight control in practical applications such as search-and-rescue missions, environmental monitoring, surveillance, and delivery operations.

6. FUTURE RECOMMENDATION AND DIRECTIONS

Although the FOPID controller performed better when the aerodynamic coupling was moderate, a few areas can still be explored to further extend this research. Future studies will be focused on scalability experiments involving larger swarm sizes (such as swarms consisting of three, five, and nine UAVs) to investigate the distribution of control efforts, limits to the distance between UAVs, and stability analysis. The effectiveness and applicability of the FOPID method have already been validated for cases involving close-range UAV formations and pairs. However, a comprehensive analysis involving its scalability, including

an analysis of its performance and control effort, for higher-density swarm formations will be performed in future studies. In future studies, the disturbance model needs to be further developed to capture the stronger nonlinearities attributed to aerodynamic couplings achieved at closely approaching ranges (less than 0.3 m). This needs to be achieved by implementing learnable paradigms such as neural networks and adaptive observer techniques. Real-time implementation and validation involving a multi-quadrotor setup are essential for determining the computational and communication overheads involved during coordinated operations.

Moreover, future extensions should examine adaptive FOPID or hybrid FOPID-linear quadratic regulator (LQR) controllers to provide increased robustness against system uncertainties and disturbances such as wind or sensor noise. Expanding the idea of UAV swarms composed of different UAVs operating remotely in a GPS environment will add to their practical applicability in multiagent aerial robotics.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Jassim Mohammed Alsayegh – Department of Aerospace Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0009-0001-4531-608X
Email: s202161710@kfupm.edu.sa

Corresponding Author

Ghulam E Mustafa Abro – Research Mentor, Postdoctoral Fellow, Interdisciplinary Research Centre for Aviation and Space Exploration, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0000-0003-1874-1889
Email: ghulam.abro@kfupm.edu.sa

ACKNOWLEDGEMENTS

The authors acknowledge the support provided by the Department of Aerospace Engineering and the Interdisciplinary Research Center for Aviation and Space Exploration at King Fahd University of Petroleum & Minerals, Saudi Arabia. We would also like to thank the external mentorship provided by the University Technology PETRONAS, Malaysia.

REFERENCES

- (1) S.-J. Chung, A. A. Paranjape, P. M. Dames, S. Shen, and V. Kumar, "A survey on aerial swarm robotics," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 837–855, 2018.
- (2) D. Morgan, G. P. Subramanian, S.-J. Chung, and F. Y. Hadaegh, "Swarm assignment and trajectory optimization using variable-swarm, distributed auction assignment and sequential convex programming," *International Journal of Robotics Research*, vol. 35, no. 10, pp. 1261–1285, 2016.
- (3) X. Kan, J. Thomas, H. Teng, H. G. Tanner, V. Kumar, and K. Karydis, "Analysis of ground effect for small-scale UAVs in forward flight," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3860–3867, Oct. 2019.
- (4) W. Honig, J. A. Preiss, T. K. S. Kumar, G. S. Sukhatme, and N. Ayanian, "Trajectory planning for quadrotor swarms," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 856–869, 2018.
- (5) J. van den Berg, S. J. Guy, M. C. Lin, and D. Manocha, "Reciprocal n-body collision avoidance," in *International Symposium on Robotics Research*, vol. 70, pp. 3–19, 2009.
- (6) P. L. Bartlett, D. J. Foster, and M. Telgarsky, "Spectrally-normalized margin bounds for neural networks," in *NeurIPS*, pp. 6240–6249, 2017.
- (7) G. Shi et al., "Neural Lander: Stable drone landing control using learned dynamics," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2019, 9784–9790.
- (8) M. Zaheer et al., "Deep sets," in *Neural Inf. Process. Syst. (NeurIPS)*, 2017, 3391–3401.
- (9) H. M. Le et al., "Smooth imitation learning for online sequence prediction," in *Proc. Int. Conf. Mach. Learn. (ICML)*, vol. 48, 2016, 680–688.
- (10) A. J. Taylor et al., "Episodic learning with control Lyapunov functions for uncertain robotic systems," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2019, 6878–6884.
- (11) R. Cheng et al., "Control regularization for reduced variance reinforcement learning," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, 1141–1150.
- (12) C. D. McKinnon and A. P. Schoellig, "Learn fast, forget slow: Safe predictive learning control for systems with unknown and changing dynamics performing repetitive tasks," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 2180–2187, 2019.
- (13) M. Saveriano et al., "Data-efficient control policy search using residual dynamics learning," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2017, 4709–4715.
- (14) T. Johannink et al., "Residual reinforcement learning for robot control," in *ICRA*, pp. 6023–6029, 2019.
- (15) I. Cheeseman and W. Bennett, "The effect of ground on a helicopter rotor in forward flight," *ARC Reports and Memoranda*, 1955.
- (16) D. Yeo et al., "An empirical model of rotorcraft UAV downwash for disturbance localization and avoidance," in *AIAA Atmos. Flight Mech. Conf.*, 2015.
- (17) D. Shukla and N. Komerath, "Multirotor drone aerodynamic interaction investigation," *Drones*, vol. 2, no. 4, 2018.
- (18) K. P. Jain et al., "Modeling of aerodynamic disturbances for proximity flight of multirotors," in *Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, 2019, 1261–1269.
- (19) D. Morgan et al., "Model predictive control of swarms of spacecraft using sequential convex programming," *Journal of Guidance, Control, and Dynamics*, vol. 37, no. 6, pp. 1725–1740, 2014.
- (20) M. Debord et al., "Trajectory planning for heterogeneous robot teams," in *IROS*, pp. 7924–7931, 2018.
- (21) D. Mellinger et al., "Mixed-integer quadratic program trajectory generation for heterogeneous quadrotor teams," in *ICRA*, pp. 477–483, 2012.
- (22) A. Liu et al., "Robust regression for safe exploration in control," *CoRR*, vol. abs/1906.05819, 2019.
- (23) S. Bandyopadhyay et al., "Nonlinear attitude control of spacecraft with a large, captured object," *Journal of Guidance, Control, and Dynamics*, vol. 39, no. 4, pp. 754–769, 2016.
- (24) T. Lee et al., "Geometric tracking control of a quadrotor UAV on SE(3)," in *CDC*, pp. 5420–5425, 2010.
- (25) S.-J. Chung et al., "Phase synchronization control of complex networks of Lagrangian systems on adaptive digraphs," *Automatica*, vol. 49, no. 5, pp. 1148–1161, 2013.
- (26) H. Khalil, *Nonlinear Systems*, 3rd ed., Pearson Education, Prentice Hall, 2002.
- (27) J. A. Preiss et al., "Crazyswarm: A large nano-quadcopter swarm," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2017, 3299–3304.

- (28) Bitcraze, "Crazyflie 2.0 thrust investigation," 2015. Available: <https://wiki.bitcraze.io/misc:investigations:thrust>
- (29) J. Forster, "System identification of the Crazyflie 2.0 nano quadcopter," Master's thesis, ETH Zurich, 2015.
- (30) M. Faessler et al., "Thrust mixing, saturation, and body-rate control for accurate aggressive quadrotor flight," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 476–482, 2017.
- (31) O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," in *ICRA*, pp. 500–505, 1985.
- (32) A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *NeurIPS*, pp. 8024–8035, 2019.
- (33) A. Molchanov et al., "Sim-to-(multi)-real: Transfer of low-level robust control policies to multiple quadrotors," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2019, pp. 59–66.
- (34) M. Metekia, E. Wondem, W. A. Asfaw, C. M. Abdissa, and L. N. Lemma, "Control of a fixed wing unmanned aerial vehicle using a robust fractional order controller," *Scientific Reports*, vol. 15, no. 1, p. 19954, 2025.
- (35) E. A. Gedefaw, N. B. Abera, and C. M. Abdissa, "A review of modeling and control techniques for unmanned aerial vehicles," *Engineering Reports*, vol. 7, no. 6, p. e70215, 2025.
- (36) F. T. Yareshe, N. W. Madebo, C. M. Abdissa, and L. N. Lemma, "Trajectory tracking of fixed-wing UAV using ANFIS-based sliding mode controller," *IEEE Access*, 2025.
- (37) T. K. Mohammed, W. A. Asfaw, C. M. Abdissa, and L. N. Lemma, "Constrained robust adaptive control design for fixed wing UAV under parameter uncertainties and external disturbances," *Engineering Research Express*, vol. 7, no. 2, p. 025254, 2025.
- (38) N. B. Abera, C. M. Abdissa, and L. N. Lemma, "An improved nonsingular adaptive super twisting sliding mode controller for quadcopter," *PLOS ONE*, vol. 19, no. 10, p. e0309098, 2024.
- (39) E. A. Gedefaw, C. M. Abdissa, and L. N. Lemma, "An improved trajectory tracking control of quadcopter using a novel sliding mode control with fuzzy PID surface," *PLOS ONE*, vol. 19, no. 11, p. e0308997, 2024.
- (40) N. W. Madebo, C. M. Abdissa, and L. N. Lemma, "Enhanced trajectory control of quadrotor UAV using fuzzy PID based recurrent neural network controller," *IEEE Access*, 2024.

Design and Comparative Analysis of a Ka-band Microstrip Patch Antenna Array for Deep-Space Exploration: Rovers Communications

Hussah Abdulrahman Aldawish and Bazilah Baharom*

Cite <https://doi.org/10.64589/juri/218956>

Submitted: September 14, 2025 Revised: February 15, 2026 Accepted: February 21, 2026

ABSTRACT

Under the harsh Martian environment, maintaining reliable communication during Martian exploration missions is challenging. Limited power availability and mass constraints necessitate innovative solutions for communication systems. This study presents the design and simulation of a medium-gain Ka-band microstrip patch antenna optimized for rover-to-orbiter communication links. Two configurations, a single patch and 1×2 series array, were developed, analyzed, and optimized using the CST Studio Suite antenna software. The single-element antenna achieved a bandwidth of 1.8 GHz (25.38–27.18 GHz) with a return loss of –16.95 dB at 26.5 GHz. The 1×2 array achieved a wider bandwidth of 3.8 GHz (24.7–28.5 GHz) with a return loss of –23.24 dB, centered at 26.5 GHz, implemented on the Rogers RT/Duroid 5880 lossy material, whose dielectric constant ϵ_r is 2.2. The proposed antenna design was optimized for lightweight implementation while achieving reliable medium-gain performance for rover-to-orbiter links, satisfying the Ka-band requirement and providing a wide field of view. The proposed design leverages microstrip patch antenna technology to improve the signal directionality and adaptability for planetary research while maintaining lightweight structures. A preliminary simulation analysis indicated promising performance characteristics, demonstrating the feasibility of this antenna for deep-space communication. Future extensions include expanding the array size to meet the Earth-return communication requirements.

Keywords: microstrip patch antenna, array antenna, Ka-band, satellite relay orbiter, Mars rovers, rover-to-Earth communication, deep-space exploration, CST studio suite

1. INTRODUCTION

With the advancement of planetary missions and the increasing need for high-data-rate telemetry, reliable communication between rovers and orbiters has become a critical component for mission success. Traditional communication systems have relied on the S-band and X-band; however, with the growing demand for bandwidth and efficiency, the Ka-band (26.5–40 GHz) is emerging as a viable alternative because of its ability to support higher data rates, better spectral efficiency, and reduced antenna size for a given gain requirement¹. The concept of planetary rovers with onboard low-, medium-, and high-gain antennas is illustrated in Figure 1. Rovers normally send most data via UHF to a Mars orbiter, which can handle up to ~2 Mb/s and then forward it to Earth; the UHF band and modest-gain omni or low-gain antennas are robust against terrain or attitude changes. For Direct-to-Earth, rovers use steerable X-band high-gain antenna, narrow beams, and higher gain for reliable long-range links to the DSN, and some missions and orbiters demonstrate or plan the Ka-band for even higher downlink efficiency². The two primary communication pathways illustrated in Figure 2 for Mars rovers are Direct-to-Earth links for immediate transmission and relay

communications via the Mars Reconnaissance Orbiter, which enhance the reliability and efficiency of the data transfer.

The Mars Global Surveyor Ka-Band Link Experiment (KaBLE-II) validated the effectiveness of Ka-band communication in space, demonstrating a 5–6 dB link improvement over the X-band under similar conditions¹. This improvement is significant for applications such as planetary rovers, where constraints on mass, volume, and power require highly efficient subsystems. However, designing a compact, high-gain antenna that can reliably operate in the Ka-band on rover platforms poses notable challenges owing to increased signal attenuation and fabrication sensitivity at higher frequencies.

Microstrip patch antennas (MPAs) are suitable for addressing these challenges owing to their low profiles, ease of fabrication, conformability to curved surfaces, and lightweight structures. Thus, MPAs are ideal for integration on space-constrained platforms such as Mars rovers. Additionally, patch antennas can be configured into arrays to achieve a higher gain and improved directivity without significantly increasing the physical footprint^{3,4}.

This study presents the simulated design of a high-gain microstrip antenna operating at 26.5 GHz, optimized for Mars

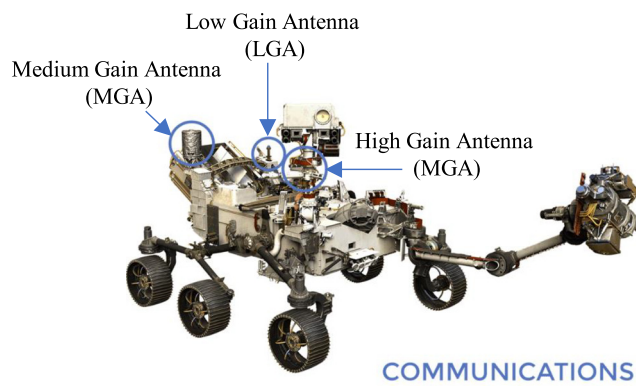


Figure 1. Planetary rovers with onboard low, medium, and high-gain antenna Orbiter (credit: NASA²)

rover-orbiter communication. Both a single patch and 1×2 patch array were evaluated in terms of return loss, gain, and bandwidth to assess their viability for robust high-frequency communication links in extraterrestrial environments.

2. LITERATURE REVIEW

Several researchers have advanced the development of Ka-band antennas for terrestrial and spatial communication. Chrysler et al.³ developed a circularly polarized 2×2 subarray at 26 GHz that achieved a gain of over 12 dBi while maintaining a compact feed network, showing potential for application in mobile and aerospace systems. Similarly, Patil et al.⁵ compared integrated and conventional feeding networks for 2×2 patch arrays and concluded that integrated feeding provided better impedance matching and higher efficiency.

A compact single-element Ka-band antenna with an enhanced bandwidth and return loss was proposed by Md. Rana et al.⁶, targeting wireless links at 26.5 GHz. Their antenna design exhibited improved performance when using inset-fed patches and optimized ground planes. Similarly, Johari et al.⁷ emphasized the importance of ground plane modifications and feed positioning in achieving high gains for small antenna arrays used in 5G and satellite systems.

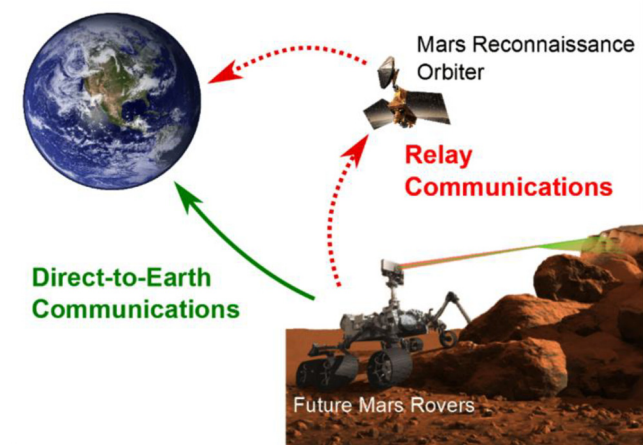


Figure 2. Direct-to-Earth and relay communications via Mars Reconnaissance Orbiter (credit: NASA²)

For future 5G and deep-space applications, Arivarasu et al.⁴ designed a multiband patch antenna with slotting techniques to enhance the bandwidth and minimize the return loss. These modifications have proven to be effective for tuning the resonant frequency and bandwidth without increasing the antenna size, which is critical for aerospace applications. Complementary work by Nabil et al.⁸ focused on the geometry for gain enhancement.

In addition to single- and small-array configurations, phased arrays have been explored for adaptive beamforming in deep-space missions. For instance, Morabito et al.¹, detailed the success of a Mars Global Surveyor experiment using a 1.5 m parabolic antenna with a Ka-band link, highlighting the benefits of frequency scaling for increased antenna gain. At the structural level, phase-array systems for spacecraft face strict mechanical and thermal constraints, as analyzed by Wang et al.⁹, where reliability in launch and space environments is paramount.

In terms of system intelligence and adaptability, Dakulagi et al.¹⁰ have outlined various smart antenna techniques, including beamforming and direction-of-arrival estimation methods. They identified robust algorithms, such as LMS and MUSIC, as critical for interference suppression and accurate signal tracking in dynamic communication environments. These adaptive features are increasingly relevant for autonomous systems such as rovers, which must maintain communication links despite changes in position and orientation.

Gao et al.¹¹ presented a comprehensive review of antenna integration into small satellites, detailing challenges such as multipath mitigation and structural loading, which are also relevant for compact rover platforms. Finally, Gatti et al.¹² discussed the phased-array antenna design strategies specifically tailored for deep-space communications, providing foundational design considerations for Ka-band platforms.

Collectively, the literature underscores the growing relevance of compact high-gain microstrip antenna systems in Ka-band communications, particularly for applications such as Martian surface missions. The combination of lightweight design, array scalability, and smart signal processing positions MPAs as a key technology for next-generation planetary communication systems.

Unlike prior studies that primarily focused on large-phase arrays or complex feeding networks, this study emphasizes a compact and low-complexity Ka-band microstrip patch array suitable for planetary rover platforms. The contribution of this study lies in the direct comparative analysis of a single patch and a simple 1×2 series-fed array under identical constraints, demonstrating clear performance gains while maintaining minimal structural and fabrication complexities. This makes the proposed design particularly relevant for lightweight and power-constrained deep-space rover applications.

3. ANTENNA DESIGN CONFIGURATION AND PROPOSED MATERIAL

This section describes the implementation of the simulation procedures, with an emphasis on the antenna design and material selection. The antenna configurations for both the single patch and the 1×2 patch array, modeled in the CST Studio Suite,

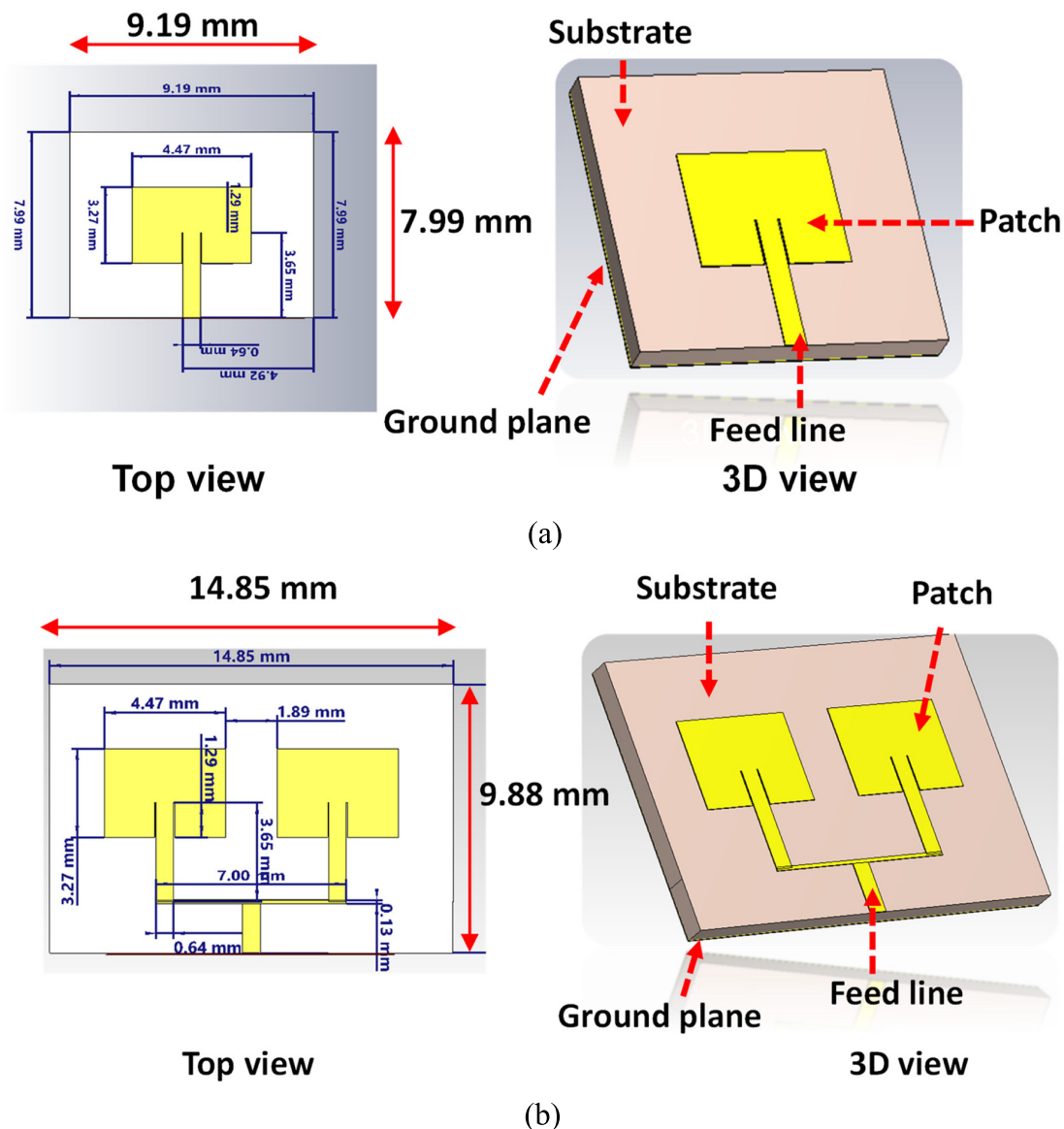


Figure 3. Configuration of antenna design for single and 1×2 array patch from CST Studio Suite (a) Antenna Design for Single Patch and (b) Antenna Design for 1×2 Patch Array

are illustrated in Figure 3. As shown in Figure 3 (a), the single-patch antenna consists of a rectangular patch element fed by a microstrip line, with a substrate layer and ground plane forming the basic structure. The dimensions of $9.19 \text{ mm} \times 7.99 \text{ mm}$ are optimized for resonance at 26.5 GHz. As demonstrated in Figure 3 (b), the 1×2 patch array integrated two identical rectangular patches connected by a microstrip feed network on the same substrate. The overall size of $14.85 \text{ mm} \times 9.88 \text{ mm}$ reflects the larger aperture, enabling higher gain and directivity compared to the single patch.

The selected substrate for both designs is Rogers RT Duroid 5880, known for its low dielectric constant ($\epsilon_r = 2.2$) and loss tangent ($\tan \delta = 0.0009$), which minimizes signal loss at millimeter-wave frequencies^{13–15}. A substrate thickness t of 0.787 mm was used to achieve a balance between compactness and bandwidth.

For the single-patch design, a rectangular microstrip patch antenna was designed using standard transmission-line model

equations to determine the patch width W and length L based on the target resonance at 26.5 GHz. The antenna was fed using a microstrip line with an inset feed to improve impedance matching⁶.

In the 1×2 array configuration, two identical rectangular patches were arranged along the x axis and connected via a microstrip series feed network. The series-fed design was selected because of its simplicity, compactness, and ease of fabrication—features well-suited for planetary rovers, where system integration is a major concern.

4. SIMULATION RESULTS AND DISCUSSION

The antenna design models were discretized using a tetrahedral high-frequency mesh in the CST Studio Suite with the frequency-domain solver. An automatic surface-based meshing scheme was employed, enabling adaptive refinement in electrically critical regions such as the patch edges, feed line, and substrate–air

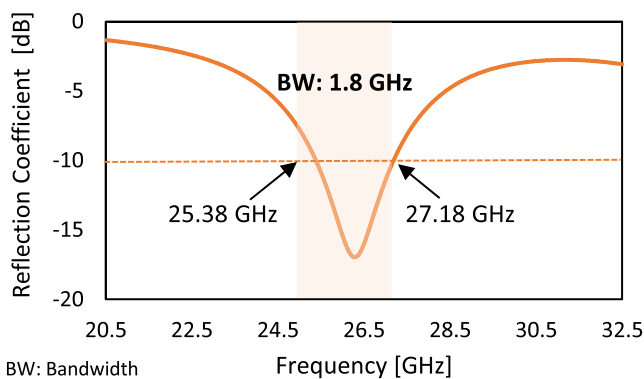


Figure 4. Single-patch antenna reflection coefficient (S_{11}) versus frequency

interfaces. The final mesh consisted of approximately 84,076 tetrahedra with a minimum edge length of 0.001 mm and a maximum edge length of 4.21 mm.

Mesh quality was monitored to ensure numerical accuracy, resulting in an average mesh quality of 0.71. The mesh convergence was verified by confirming stable S-parameters and gain responses across successive adaptive refinement steps. This adaptive meshing strategy ensures reliable computation of the S-parameters, radiation characteristics, and gain at Ka-band frequencies.

4.1. Single-Patch Antenna Performance. The single-patch antenna (Figure 4) was designed to resonate at a center frequency of 26.5 GHz. The CST Studio Suite simulation results and S_{11} plot revealed that the antenna achieved a -10 dB bandwidth from 25.38 GHz to 27.18 GHz, yielding a total operational bandwidth of 1.8 GHz with a return loss of -16.95 dB at 26.5 GHz. This wide bandwidth ensures stable performance for Ka-band rover communication, with some tolerance for frequency drift or mismatch.

The voltage standing wave ratio (VSWR) performance of the single-patch antenna is illustrated in Figure 5. At the resonant frequency of 26.5 GHz, the antenna exhibits a VSWR of approximately 1.38, which indicates good impedance matching and efficient power transfer between the feed line and the radiating patch. The VSWR remains below the standard threshold of 2 within the -10 dB bandwidth (25.38–27.18 GHz), confirming stable operation and minimal reflection losses across the

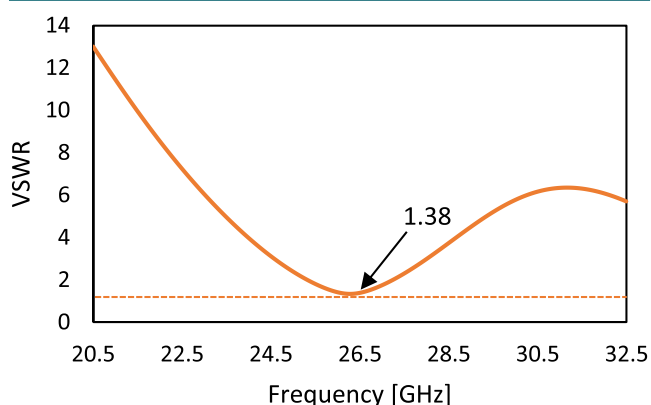


Figure 5. VSWR of the single-patch antenna

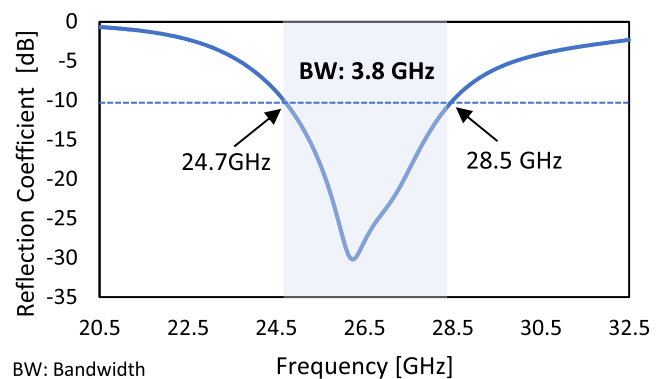


Figure 6. 1×2 array patch antenna reflection coefficient (S_{11}) versus frequency

intended Ka-band frequency range. These characteristics make the single-patch design suitable for scenarios requiring moderate gain with broad angular coverage.

4.2. 1×2 Patch Array Antenna Performance. The 1×2 series microstrip array antenna was simulated under the same design constraints, incorporating two identical radiating elements with a series feed configuration. According to the simulation results shown in Figure 6, this configuration achieved a broad -10 dB bandwidth from 24.7 GHz to 28.5 GHz, corresponding to a total bandwidth of 3.8 GHz. The return loss reached -23.24 dB at the center frequency of 26.5 GHz, representing a significant enhancement compared to the single patch antenna.

The VSWR of the 1×2 array patch antenna across the simulated frequency range is shown in Figure 7. At the resonant frequency of 26.5 GHz, the antenna achieves a VSWR of approximately 1.08, which is very close to the ideal value of 1, indicating excellent impedance matching and minimal reflection. Furthermore, the VSWR remains well below the acceptable threshold of 2 across the -10 dB bandwidth (24.7–28.5 GHz), confirming efficient power transfer and stable performance throughout the operating band. This result demonstrates the superior matching characteristics of the array compared with the single-patch design.

4.3. Performance Comparison. The comparative performances of the single-patch and 1×2 patch array antennas are

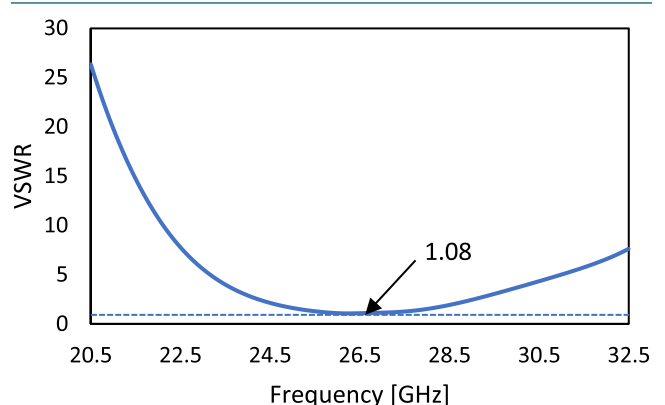


Figure 7. VSWR of the 1×2 array patch antenna

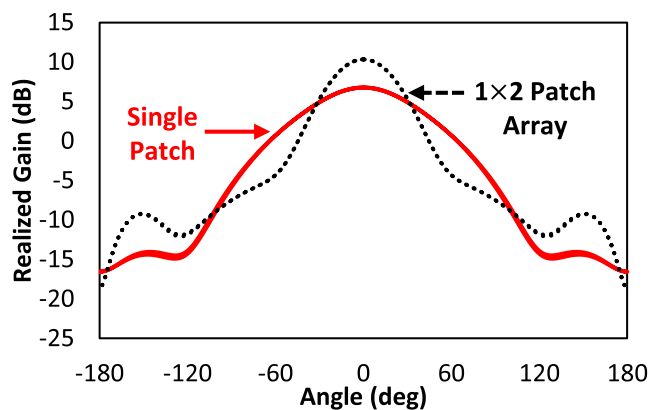


Figure 8. Radiation pattern comparison of single-patch antenna and 1×2 patch array antenna

summarized in Table 1. The array demonstrated superior performance with a wider bandwidth 3.8 GHz, higher gain 10.7 dB, and narrower half-power beam width (46.1°) than a single patch.

According to the classical array theory, combining multiple radiating elements in a phase, results in an increase in gain owing to constructive interference. While larger arrays generally provide higher gain, they also increase the structural complexity and feeding network losses. The proposed 1×2 array achieves a gain of 10.7 dB, demonstrating that efficient performance can be

Table 1. Comparative performance of single patch and 1×2 patch array antenna

Antenna Characteristics / Type	Single Patch	1×2 Patch Array
Frequency range (GHz)	25.38–27.18	24.7–28.5
Bandwidth (GHz)	1.8	3.8
Gain (dB)	7.7	10.7
HPBW (deg)	86.1°	46.1°
Side-lobe-level (dB)	-16.0	-19.5

Additionally, the array achieved a lower side-lobe level, indicating improved radiation efficiency and directivity.

obtained with a minimal number of elements, which is advantageous for compact rover platforms¹⁶. The validation analysis shows that the theoretical gain increase from a single patch to a 1×2 array follows classical array theory, where combining two identical in-phase radiating elements ideally increases the gain by approximately 3 dB. This theoretical expectation aligns well with the simulation results, where the single patch and 1×2 array achieve gains of 7.7 and 10.7 dB, respectively.

Figure 8 compares the radiation patterns of the single patch and 1×2 patch array at 26.5 GHz. The single patch exhibits a peak realized gain of 7.7 dB, whereas the 1×2 array achieves a higher gain of 10.7 dB. This confirms that the array configuration successfully enhances the antenna performance by approximately 3 dB owing to the constructive combination of the radiated fields

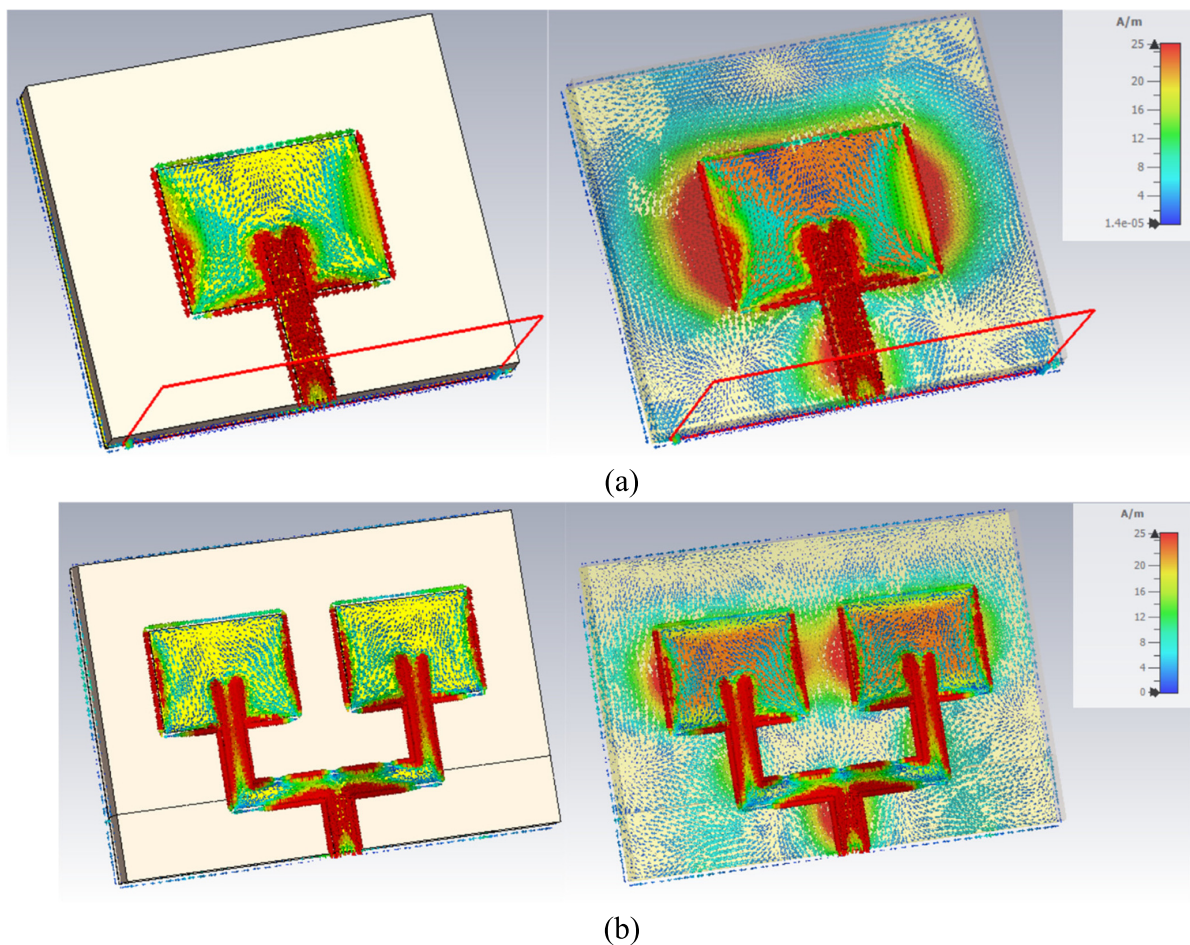


Figure 9. Surface-current distribution at 26.5 GHz (a) Single-patch antenna: (left) non-transparent view and (right) transparent view. (b) 1×2 patch array antenna: (left) non-transparent view and (right) transparent view

from both elements. The proposed antenna operates with linear polarization and is commonly employed in rover communication systems owing to its simplicity and ease of integration. In the single-patch configuration, the radiation exhibits broader angular coverage, resembling semi-omnidirectional behavior. When extended to a 1×2 array, the polarization remains linear but becomes more directionally focused due to constructive field combination, resulting in improved gain and beam shaping. Additionally, the array provides improved directivity while maintaining broadside radiation characteristics.

Figure 9 shows the current distribution, where the left-hand side presents a non-transparent view, and the right-hand side shows a transparent view for both the single patch and 1×2 array patch antenna at 26.5 GHz. Figure 9 (a) shows the surface-current distribution of the single-microstrip patch antenna operating at 26.5 GHz. The current was predominantly concentrated along the feed line and central portion of the patch, with reduced intensity observed toward the radiating edges. This distribution validates the excitation of the maximum current at the patch center and minimum at the open-circuit edges. The transparent view further illustrates the fringing fields around the patch periphery, which are responsible for emitting radiation into free space.

Figure 9 (b) shows the surface-current distribution of the proposed 1×2 microstrip patch array at 26.5 GHz. The current is strongly concentrated along the feed network and symmetrically distributed across both patches, indicating effective excitation of the array elements. The highest current density appears near the feed junction and patch centers, whereas the edges exhibit weaker currents, which is consistent with the radiating boundaries.

This distribution validates the proper operation of the series-fed configuration, which ensures in-phase excitation of both patches, achieving a higher gain and enhanced directivity compared to those achieved using the single-element design.

5. CONCLUSIONS

In this study, we designed and simulated a medium-gain Ka-band microstrip patch antenna, which was optimized for rover-to-orbiter communication links. The simulation results revealed that both the single-microstrip patch antenna and 1×2 series array exhibited high performance at Ka-band frequencies and were thus suitable candidates for Mars rover-to-orbiter communication systems. However, distinct differences in bandwidth, gain, and radiation patterns highlighted the tradeoffs between compactness and performance.

In rover-to-orbiter communication, antenna performance characteristics, such as gain, bandwidth, and beamwidth, directly influence the link reliability and data throughput. A higher gain improves the link margin and supports higher data rates, whereas a wider bandwidth ensures robustness against frequency drift and fabrication tolerances. The narrower beamwidth of the 1×2 array enhances directional energy transmission toward orbiters with predictable trajectories, thereby improving communication reliability compared with a single patch antenna.

The single-patch antenna provides a moderately wide bandwidth of 1.8 GHz (25.38–27.18 GHz) with a peak gain of 7.7 dB, suitable for short-range, omnidirectional, or semi-directional links. Its broader 3 dB beamwidth of 86.1° ensures wide angular

coverage, which is ideal for applications in which alignment with an orbiter cannot always be precisely controlled. Consequently, a single element is advantageous in low-complexity and low-weight environments, especially when the gain demands are modest.

In contrast, the 1×2 array delivers significantly enhanced performance in terms of gain and bandwidth. The array achieves a broader bandwidth of 3.8 GHz (24.7–28.5 GHz) and a main lobe gain of 10.7 dB, which are critical for reliable high-data-rate transmission and extended link distances. The narrower beamwidth of 46.1° indicates focused energy radiation, which is particularly beneficial for directional communication with orbiters flying over the head or in known positions. Additionally, the improved side-lobe-level suppression (19.5 dB) reduced interference and energy loss.

These findings are consistent with those reported in the literature, in which array configurations have been shown to enhance the gain and improve link margin in Ka-band systems^{3,5,8}. The high total efficiencies greater than ~90% in both designs also validate the use of Rogers RT Duroid 5880 as a low-loss substrate for millimeter-wave applications. Ultimately, the 1×2 series-fed patch array provides superior performance and is recommended for primary deployment in Martian surface communication platforms. However, a single patch remains valuable in scenarios requiring reduced hardware complexity and power consumption.

A single-patch antenna serves as a fundamental and widely accepted baseline for evaluating improvements in array performance. The comparative results obtained under identical substrate material, operating frequency, and feed configuration conditions directly highlight the performance enhancement achieved solely through array formation, without introducing additional design complexity.

Future investigations may extend the proposed design to higher-order array configurations for 2×2 or larger arrays to achieve a higher gain and narrower beamwidth. Such extensions would further enhance link robustness for longer rover-to-orbiter distances. In addition, experimental fabrication and measurements were performed to validate the simulated performance.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Hussah Abdulrahman Aldawish – Aerospace Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0009-0003-6048-1475
Email: s202273600@kfupm.edu.sa

Corresponding Author

Bazilah Baharom – Research Mentor, Postdoctoral Fellow, Interdisciplinary Research Centre Aviation and Space Exploration, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0000-0003-3193-6197
Email: bazilah.baharom@kfupm.edu.sa

ACKNOWLEDGEMENTS

The authors acknowledge the support provided by the Interdisciplinary Research Center for Aviation and Space Exploration

at the King Fahd University of Petroleum and Minerals. The author also extends her appreciation to the advisor for providing their guidance throughout the RES 200 Independent Research course, particularly for understanding antenna operations in space applications, conducting a literature review, and supporting the antenna design process.

REFERENCES

- (1) D. Morabito, S. Butman, and S. Shambayati, "The Mars Global Surveyor Ka-Band Link Experiment (MGS/KaBLE-II)," *Telecommunications and Mission Operations Progress Report*, 42-137, Jet Propulsion Laboratory, NASA, Pasadena, CA, USA, May (1999).
- (2) NASA, "Mars 2020 Mission: Perseverance Rover," 2021. [Online]. Available: https://www.nasa.gov/wpcontent/uploads/2021/01/mars_2020_mission_perseverance_rover.pdf. (2021).
- (3) R. N. Simons, F. A. Miranda, A. Chrysler, and C. Furse, "A Ka-Band (26 GHz) Circularly Polarized 2×2 Microstrip Patch Sub-Array with Compact Feed," in *Proc. IEEE Antennas and Propagation Society Int. Symp.*, Cleveland, OH, USA, (2013).
- (4) M. Arivarasu et al., "Microstrip Patch Antenna Development at K Band for Satellite Communication," *International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET)*, 10, 1, 120–125. (2024).
- (5) V. Patil and N.B. Sambre, "Design and Comparison of 2x2 Array Antenna using Integrated and Conventional Feeding Network," in *International Journal of Innovative Research in Technology (IJIRT)*. 9, 10, 280-283. March (2023).
- (6) Md. S. Rana et al., "Design and Analysis of Ka-band Microstrip Patch Antenna for Wireless Applications," *International Journal of Scientific Research in Engineering and Management (IJSREM)*, 5, 8, 1–6. (2021).
- (7) M.N. Johari et al., "28 GHz Microstrip Patch Antennas for Future 5G," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 6, 9, 7781–7787. (2017).
- (8) M. Nabil and M.M.A. Faisal, "Design, Simulation and Analysis of a High Gain Small Size Array Antenna for 5G Wireless Communication," *Wireless Personal Communications*, 118, 1885–1902. (2021).
- (9) C. Wang et al., "Space Phased Array Antenna Developments: A Perspective on Structural Design," *IEEE Aerospace and Electronic Systems Magazine*, 35, 7, 30–45. (2020).
- (10) V. Dakulagi and M. Bakhar, "Advances in Smart Antenna Systems for Wireless Communication," *Wireless Personal Communications*, 113, 655–679. (2020).
- (11) S. Gao, Y. Rahmat-Samii, R. E. Hodges and X. -X. Yang, "Advanced Antennas for Small Satellites," in *Proceedings of the IEEE*, 106, 3, 391-403, March (2018).
- (12) M.S. Gatti, "A Phased Array Antenna for Deep Space Communications," *Jet Propulsion Laboratory, California Institute of Technology*, Pasadena, CA, Technical Report. (2003).
- (13) M. I. Khattak, A. Sohail, U. Khan, et al., "Elliptical Slot Circular Patch Antenna Array with Dual Band Behaviour for Future 5G Mobile Communication Networks," *IEEE Access*, 7, 1–10, 2019.
- (14) J. Mazierska, M. V. Jacob, D. Ledenyov and J. Krupk, "Loss tangent measurements of dielectric substrates from 15K to 300K with two resonators: investigation into accuracy issues," *2005 Asia-Pacific Microwave Conference Proceedings*, Suzhou, China, 4, 2005.
- (15) Alsawaf, H. A., and Ahmad, B. M. "Performance of circular patch microstrip antenna for adaptive modulation and coding applications". *Bulletin of Electrical Engineering and Informatics*, 11(3), 1370-1380. (2002)
- (16) Shah, Md. Farid and Singh, Aheibam Dinamani. "Design and Analysis of Microstrip Patch Antenna Arrays for Millimeter Wave Wireless Communication". *International Journal of Engineering and Advanced Technology (IJEAT)*, 9, 3, 281–286, (2020).

A hydrodynamics quantification of tortuosity and residence time of TPMS based bone scaffold

Btoul S. Alkhalil and Ahmed S. Dalaq*

Cite <https://doi.org/10.64589/juri/218963>

Submitted: November 29, 2025 Revised: January 24, 2026 Accepted: January 28, 2026

ABSTRACT

A critical component of the biocompatibility of a bone scaffold lies in its mechanical, chemical, and physical properties. However, less attention has been paid to the scaffold's transport properties, which determine its interaction with the host biological environment. We propose a hydrodynamic measure of tortuosity in relation to mean particle residence time (MRT) and its effect on bone scaffold performance. Using high-fidelity computational fluid dynamics (CFD), we quantified MRT, volumetric flow rates, and geometric tortuosity to establish their relationship with permeability and hydraulic resistance. Our results demonstrate that increased geometric tortuosity reduces MRT under pressure-driven flow due to localized nozzle-like acceleration. This study provides a measure of tortuosity, both geometrically and from particle residence time, with the overarching objective of matching it to bone in future studies.

Keywords: bone scaffold, computational fluid dynamics, tortuosity, particle residence time

1. INTRODUCTION

Bone defects due to trauma, infection, tumors, congenital disabilities, and degenerative diseases remain major clinical challenges in modern orthopedic surgery¹. Although autografts and allografts are used globally, they are limited by the postsurgical immune response, risk of disease transmission, insufficient availability, and donor-site morbidity, especially in cases with severe bone loss or critical-sized defects²⁻⁴. Therefore, following the foundational work of Vacanti⁵, bone tissue engineering has introduced the concept of combining scaffolds, cells, and bioactive cues to regenerate functional bone, rather than merely replacing missing tissue. Bone tissue engineering is considered a promising alternative for developing biomimetic scaffolds because of its ability to guide successful osseointegration, transmit loads, and enable nutrient diffusion and vascularization⁶. As these outcomes largely depend on how the scaffold interacts with its biological environment, the architecture of the transplanted scaffold governs both the mechanical stability and transport behavior experienced by biomolecules, significantly affecting the success of these bioengineered scaffolds.

Various scaffold architectures have been proposed to optimize the balance among mechanical strength, biological performance, and manufacturability (Figure 1). Generally, scaffold architectures can be categorized into two main types: whole-structure and unit-cell-based designs. Whereas whole-structure design focuses on the entire geometry being designed as a continuous structure (e.g., functionally graded or topologically optimized forms), unit-cell-based design is constructed from repeating motifs⁷. Unit-cell scaffolds can be generated through

parametric approaches, including Voronoi and triply periodic minimal surfaces (TPMS), or nonparametric approaches, such as body-centered cubic, face-centered cubic, polyhedral, or honeycomb lattices.

TPMS scaffolds, such as gyroids, double gyroids, Schwarz diamonds, and primitive structures, are constructed using specific trigonometric equations called level-set equations with mean curvatures close to zero⁸. TPMS have gained significant attention compared to traditional lattice designs because of their high surface-area-to-volume ratio, tunable porosity, and enhanced permeability, each of which contributes to improving cell adhesion, proliferation, and transport of growth factors, oxygen, and nutrients, as well as the removal of waste¹⁰⁻¹². TPMS possess complex interconnected pore networks that replicate the trabecular structure of native bone, explaining their rapid adoption in next-generation bone implants. Several studies have investigated how the mechanical architecture of a TPMS exhibits optimized load-bearing behavior and energy-dissipating capability that match the behavior of cancellous bone. These studies show how the continuity of the TPMS surfaces enables a uniform stress distribution, which minimizes localized stress concentration and lowers the probability of mechanical failure^{13,14}. These advantages explain their rapid adoption as a next-generation bone implant.

Moreover, considering that transport behavior within porous scaffolds is instrumental in determining tissue regeneration success, several studies have analyzed velocity and pressure profiles, quantified Darcy's permeability and wall shear stress, and tested mass transfer performance to demonstrate how

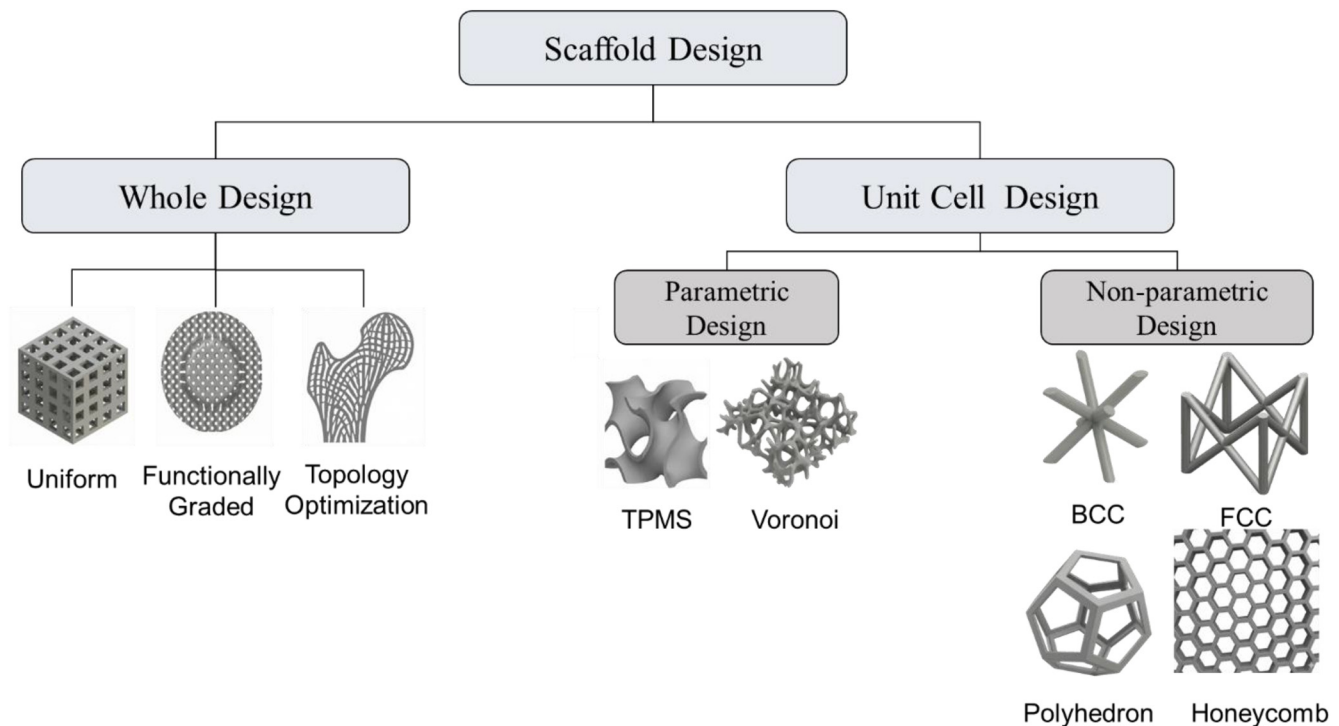


Figure 1. Bone Scaffold Design. Classification of scaffold design approaches, including whole-structure designs and unit-cell-based designs, adopted from ^{8,9}

geometric offsets affect the tortuosity and sinuosity of the porous medium, while improving mass transfer behavior^{13,15–19}. However, a few studies have fully investigated the mean particle residence time (MRT; the time a single particle spends within the pores and paths of the structure before leaving) of TPMS structures. In addition, the current literature fails to address the correlation between tortuosity and MRT within TPMS-based bone scaffolds. In particular, the multifaceted biocompatibility requirements remain incomplete without matching the transport properties of the scaffold, which should receive the same level of attention as its chemical and mechanical characteristics (e.g., avoiding stress shielding).

In this study, we investigated how TPMS scaffold geometry influences MRT and its relationship with tortuosity using a computational fluid dynamics-based approach. By assessing the tortuosity geometrically and through the MRT, this study provides a framework for predicting the tortuosity of scaffolds and general lattice structures via fluid flow.

2. METHODS

2.1. Overall Workflow. The workflow integrates numerical fluid dynamics simulations with morphological analyses to extract key transport metrics relevant to bone tissue engineering.

As illustrated in Figure 2, the workflow commenced with the generation of scaffold geometries using MS-Lattice software 1.0²⁰. Numerical flow simulations were performed using COMSOL Multiphysics 6.3²¹ to compute the hydraulic resistance and mean residence time (MRT). In parallel, morphological analysis was performed using Dragonfly²² to quantify scaffold tortuosity based on geometric path analysis.

2.2. Computational Domain and Scaffold Design.

Gyroid and double gyroid architectures were selected based on *in vivo* studies reporting superior bone ingrowth compared to that of other TPMS designs²³. The fluid domain consisted of a prismatic cylindrical region with length $L = 10$ mm and diameter $D = 10$ mm. Figure 3a shows the non-architected domain used as a reference. We analyzed gyroid structures with two unit-cell sizes, $0.5 D$ and D , and a double gyroid of unit size D . TPMS structures were generated using MS-Lattice²⁴. The selected TPMS geometries are standard and have been extensively studied from a mechanical perspective, although to a lesser extent in terms of their transport properties. All TPMS scaffolds were generated with a porosity of $\phi \sim 40\%$, representing an optimal balance between mechanical strength and permeability, a level known to provide high robustness relative to both lower and higher porosities (i.e., enhanced strength and toughness), and therefore adopted here²⁵. For each TPMS case, the fluid domain corresponded to the complementary space obtained by the Boolean subtraction of the solid TPMS from the cylindrical domain. An empty case without a scaffold geometry was simulated under the same boundary conditions and material properties to provide a baseline for evaluating the influence of the porous scaffold on flow transport.

The model was driven by a pressure difference of $\Delta P = 1$ Pa, with a no-slip condition imposed at the interface between the TPMS walls and fluid domain, as well as on the circumferential sides of the fluid domain to simulate the pipe enclosure (Figure 3b). This pressure difference does not precisely mimic a physiological value, but rather falls within the range of pressure differences typically used; for example, in studies of interstitial fluid flow around osteocytes²⁶ and bioreactor cultures of bone tissue constructs²⁷, which are typically on the ~ 10 mm scale²⁸.

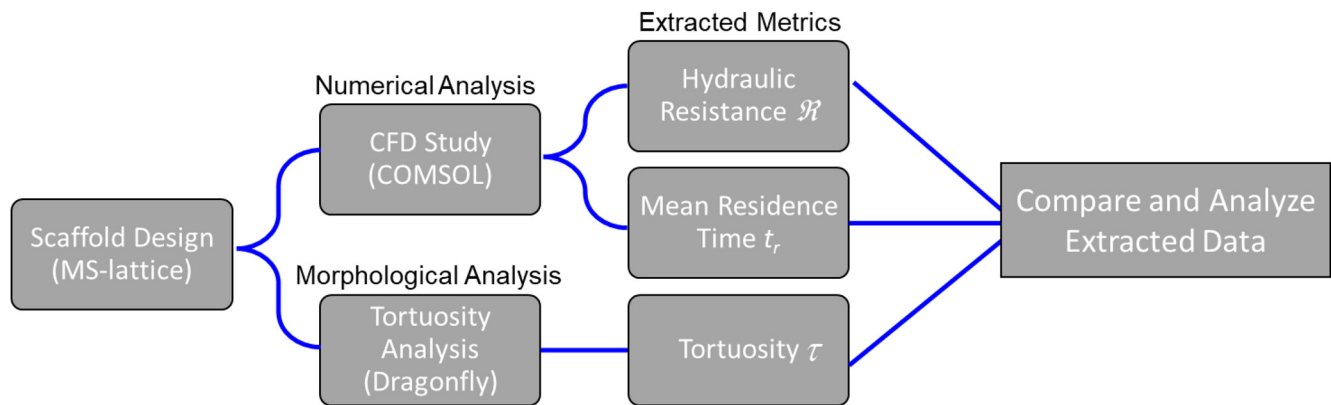


Figure 2. Overview of the computational workflow integrating CFD simulations and morphological analysis to extract hydraulic resistance, mean residence time, and tortuosity of scaffolds

The model assumed laminar flow (Reynolds number <2200) and a Newtonian fluid governed by a constitutive law with viscosity $\mu = 3 \text{ mPa} \cdot \text{s}$ and density $\rho = 1000 \text{ kg/m}^3$. The fluid domain was governed by the well-known Navier–Stokes equations²⁹.

First, the velocity field and fluid kinematics were mesh independent in the mesh sensitivity study. The velocity magnitude fields for all geometries are shown in Figure 4a–4d for the non-architected case, two gyroids, and double gyroid. We begin with the non-architected case (i.e., fluid passing through a pipe). We selected gyroid variants owing to their strong geometric resemblance to natural bone³⁰. The reference case exhibited the expected parabolic velocity profile, where the velocity contours were zero near the wall (blue, approximately $y = D/2$) and increased parabolically toward a maximum at the center. A cross-sectional plot of the velocity along the longitudinal axes x , v_x confirms this parabolic trend (Figure 4b). This case represented Poiseuille flow. The computational fluid dynamics (CFD) results showed an error $<2\%$ relative to the analytical model.

With confidence in the established CFD framework, we introduced a gyroid architecture with a $0.5D$ unit-cell. This structure immediately produces spatial heterogeneity in the velocity field, with a maximum of approximately 0.02 m/s occurring in narrow, nozzle-like regions where the flow accelerates. However, this peak velocity is only about one-tenth that of the non-architected reference case, indicating that the TPMS significantly slows the overall flow. The qualitative distribution of the velocity contours is similar among the TPMS cases, with the highest velocities observed for the gyroid with a unit-cell size of D (Figure 4e). Overall, all three TPMS geometries generate alternating high- and low-velocity regions; the greater the number of tortuous channels and stagnation pockets, the lower the overall velocity magnitude.

The flow within these TPMS geometries is highly complex, and determining the specific channel features that contribute to flow acceleration or deceleration requires further analysis. These scaffolds are expected to exhibit varying particle transport and residence-time behaviors.

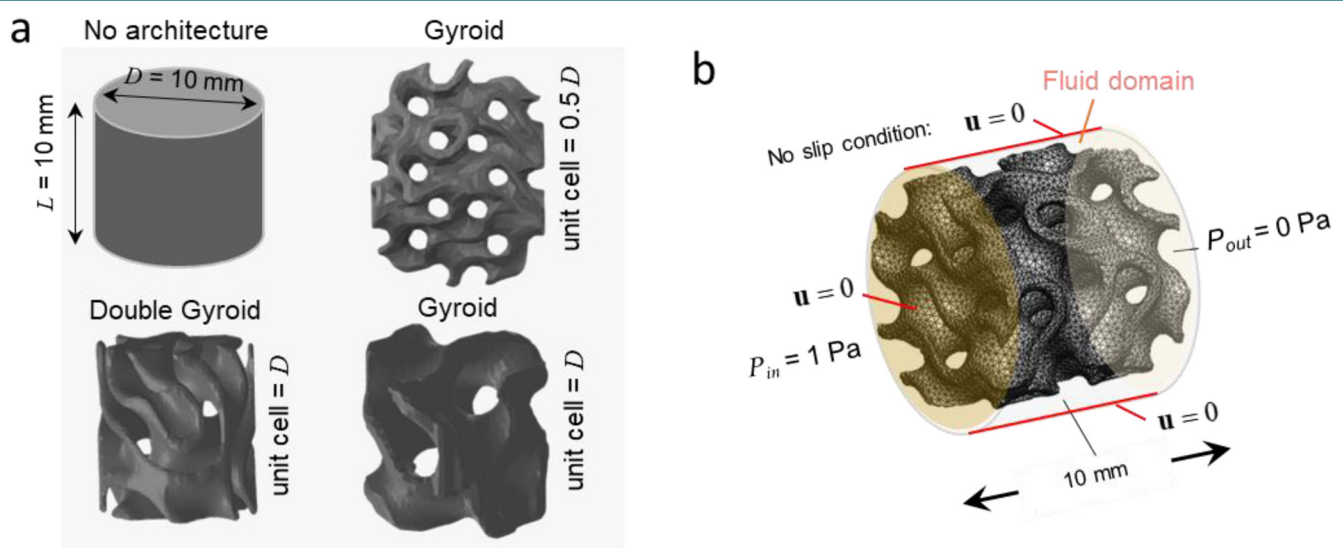


Figure 3. Adopted TPMS structures and CFD model. (a) No architecture (reference), Gyroid with unit-cell size $0.5D$, double Gyroid, and Gyroid with unit-cell size D . The fluid domain is a cylindrical domain of length $L=10 \text{ mm}$ and diameter $D=10 \text{ mm}$. (b) Simulation setup driven by a pressure difference ($\Delta P = 1 \text{ Pa}$), showing the fluid domain (orange) and boundary conditions. The full computational domain includes the surrounding pipe along with the imposed inlet and outlet pressure boundary conditions used in the CFD analysis

2.3. Governing Equations and Material Properties.

The flow within the domain was modeled using the steady incompressible Navier–Stokes Eqs²⁹:

$$\rho (\mathbf{u} \cdot \nabla) \mathbf{u} = \nabla \cdot [-P\mathbf{I} + \mathbf{K}] + \mathbf{F}$$

$$\mathbf{K} = \mu (\nabla \mathbf{u} + (\nabla \mathbf{u})^T)$$
(1)

where ρ is the fluid density, which equals 1000 kg/m^3 , $\mathbf{u}$ is the velocity field vector, P is the static pressure, and $\mathbf{K} = \mu (\nabla \mathbf{u} + (\nabla \mathbf{u})^T)$ is the viscous stress tensor. $\mathbf{F} = 0$ because no body forces were considered. The fluid was assumed to be Newtonian with material viscosity $\mu = 3 \text{ mPa.s}$.

2.4. Boundary Conditions. At the inlet of the pipe, a pressure condition was applied with a suppressed backflow and normal flow enforcement. The system was driven by a pressure difference, where the inlet pressure was 1 Pa, and outlet of the pipe had a fixed pressure of 0 Pa with a suppressed backflow (Figure 3b).

$$P(x = 0) = 1 \text{ Pa and } P(x = L) = 0 \text{ Pa}$$
(2)

All walls of the pipe and scaffold were assigned a no-slip condition, except at the inlet and outlet boundaries (Figure 3b).

$$\mathbf{u} = 0$$
(3)

2.5. Solver Configuration. Equations (1)–(2) were solved using COMSOL Multiphysics. The study was conducted using a stationary (steady-state) solver with a pseudo-time range of 0–1 in increments of 0.1. The simulation results were analyzed to assess the velocity and pressure distributions, with particular attention to the formation of gradients in and around the scaffold region.

2.6. Meshing Strategy & Mesh Sensitivity Analysis. A physics-controlled mesh is employed and iteratively refined by tuning the minimum element size, curvature resolution, growth rate, and boundary-layer thickness. Mesh independence was verified by monitoring the convergence of the velocity field and pressure drop across the scaffold. A mesh was adopted after a stable saturated solution was obtained.

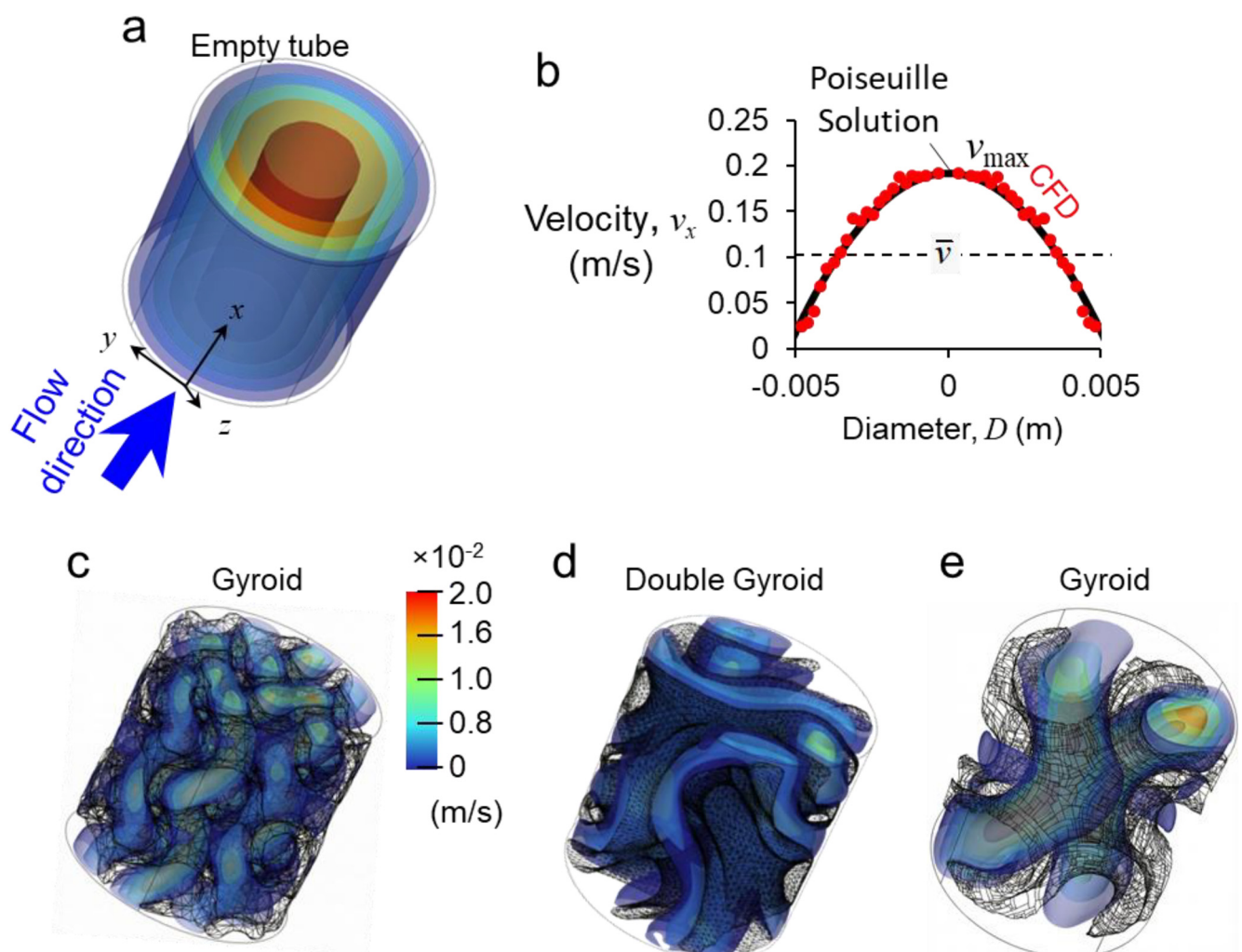


Figure 4. Velocity field contours and verification of the CFD model. (a) No architecture (reference), (b) verification of the CFD axial velocity, v_x (in red), against Poiseuille's solution (black continuous line). $\bar{v}$ is the mean velocity while v_{max} is the maximum velocity located at $y = 0$. (c) Gyroid with unit-cell size $0.5 D$, (d) double Gyroid, and (e) Gyroid with unit-cell size D . The color bar range of v_x is from 0–2.0 m/s. Flow direction is indicated by the blue arrow in panel (a)

To ensure the accuracy and mesh independence of the MRT results, a sensitivity test was conducted on the gyroid scaffold geometry. The MRT was calculated for several mesh domain element sizes until it was observed that the MRT stabilizes after 3.42 million refinement elements (Figure 5). Mesh physics corresponding to this element size were adopted to perform the MRT analysis for other bone scaffold designs.

2.7. Residence Time Computation. The residence time was evaluated using two methods to determine the transport characteristics of the porous scaffolds. The domain-averaged residence time, denoted as t_r , was defined in COMSOL by introducing the custom variable:

$$\begin{aligned} V &= \iiint dx dy dz \\ Q &= v \cdot \mathbf{n} dA \\ t_r &= \frac{V}{Q} \end{aligned} \quad (4)$$

where V is the overall volume across the entire geometry domain obtained using nonlocal coupling integration, and Q is the volumetric flow rate at the outlet obtained by integrating the normal velocity over the outlet boundary (Eq. (3)). The resulting MRT, t_r , is defined as the volume divided by the flow rate (Eq. (3)).

In addition to the average residence time measurements, the flow velocity field and pressure contour plots were analyzed to understand the flow behavior and gradient. A streamline plot based on the velocity field was added under the velocity field graph to visualize the flow trajectories and analyze the transport pathways in relation to the geometrical properties of the scaffold. These analyses provide both quantitative and qualitative insights into scaffold-fluid interactions.

2.8. Tortuosity Analysis. Geometric tortuosity τ is the ratio between the mean length of the geometric flow paths within a porous medium and direct straight-line distance across it³¹:

$$\tau = L / [r_2 - r_1] \quad (5)$$

Where L is the actual flow path, and the straight line is between the points (r_1, r_2) represented by the vector $[r_1 - r_2]$ ¹⁹.

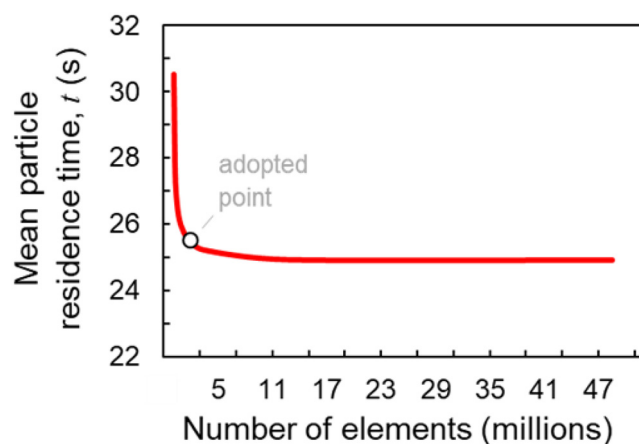


Figure 5. Mesh sensitivity analysis. Variation in mean residence time was observed as the mesh volume increased. The results converge beyond 3.4 million elements, indicating sufficient mesh independence for subsequent simulations

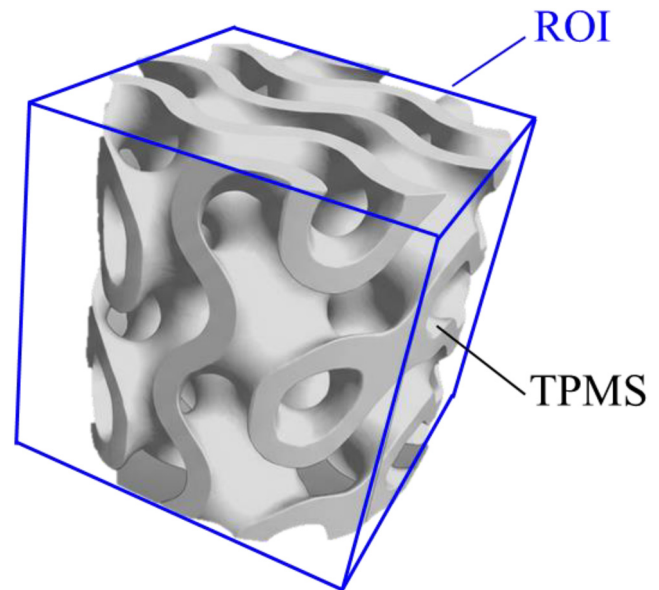


Figure 6. Bounding Box of Region of Interest (ROI) of the scaffold

Effective tortuosity was computed using Dragonfly²², via the skeletonization of the pore space. In this approach, the void region enclosed by the TPMS sheets within a user-defined region of interest (ROI), corresponding to the entire TPMS scaffold shown in Figure 6, was converted into a skeletonized graph representing the pore network. First, STL files were generated for each geometry and imported into Dragonfly, where a bounding box was defined for each structure to establish a structural grid as Region of Interest 1 (ROI1; Figure 6). The interior of ROI1 was filled in three dimensions, and a convex hull was generated to create the second region of interest (ROI2). Subtracting ROI1 from ROI2 yielded a third region of interest representing the pore space (ROI3). Pore Network Modeling was then applied to ROI3 to generate a network representation of the pores. Tortuosity was calculated from pore network measurements, specifically using longest-path analysis. The resulting graph consists of nodes (endpoints) and segments that define the pathways connected through the TPMS structure. For each path, the tortuosity was computed as the ratio of the actual path length along the segments to the straight-line (Euclidean) distance between the path endpoints (Eq. (4)). Histograms of the curvilinear path lengths were obtained, and the mean and median tortuosity values were determined. This procedure was repeated for all three scaffold geometries.

2.9. Permeability and Hydraulic Resistance. Permeability and hydraulic resistance were quantified using Darcy's law based on the volumetric flow rate Q obtained from the CFD simulation. The pressure difference ΔP across the scaffold was measured from the imposed inlet pressure condition and zero-pressure outlet. The effective permeability k was then determined using the following.

$$k = \frac{\mu Q L}{A \Delta P} \quad (6)$$

where μ is the fluid viscosity, L is the scaffold length, and A is the inlet cross-sectional area (Eq. (6)).

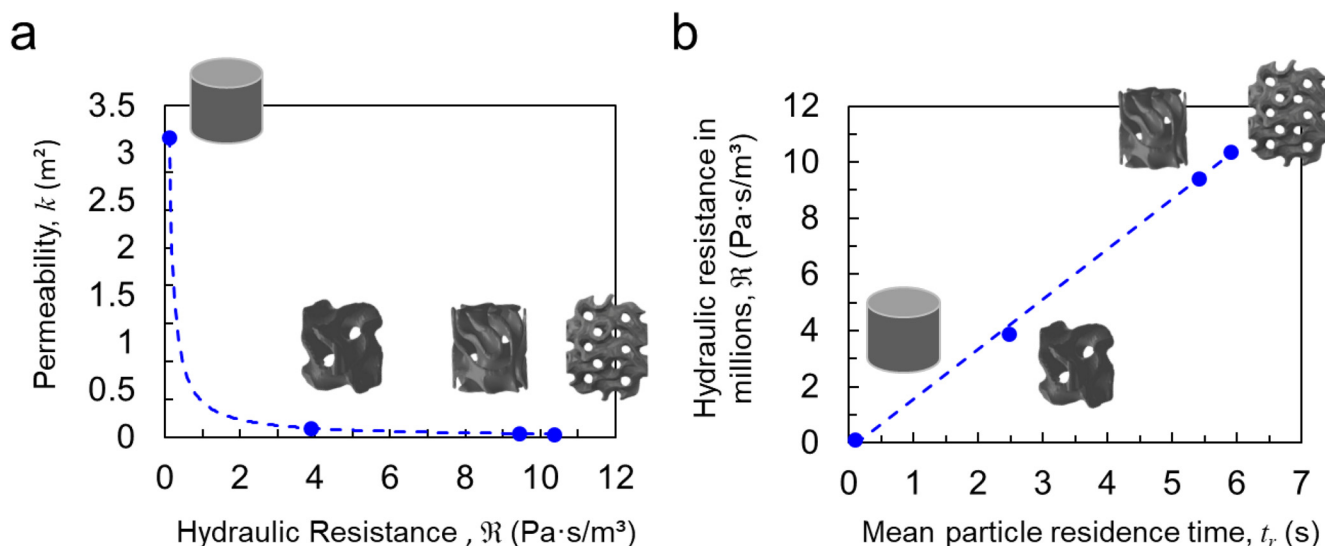


Figure 7. Transport characteristics under varying hydraulic resistance. (a) Permeability–resistance curve showing an inverse dependence of k on $\mathfrak{R}$. (b) Residence-time–resistance curve showing direct proportional growth of mean residence time with $\mathfrak{R}$

Hydraulic resistance $\mathfrak{R}$ was calculated as:

$$Q = \frac{\Delta P}{\mathfrak{R}} \quad (7)$$

which, when substituted into Eq. (6) and solved for $\mathfrak{R}$, becomes:

$$\mathfrak{R} = \frac{\mu L}{A} \frac{1}{k} \quad (8)$$

This characterizes the opposition of the structure to the flow (Eq. (5)). These calculations provide geometry-dependent transport parameters that are directly linked to the applied physics and simulated flow fields.

3. RESULTS AND DISCUSSION

3.1. Hydraulic Resistance and Permeability. As the simulation can capture the overall field flow, we can obtain the permeability, hydraulic resistance, and MRT (Methods). The relationship between permeability, k , and hydraulic resistance, $\mathfrak{R}$, for the examined TPMS scaffolds is shown in Figure 7a. As expected for laminar flow through porous structures, the permeability decreased with increased hydraulic resistance, following a power-law profile. Scaffolds with lower $\mathfrak{R}$ exhibited higher k , indicating reduced opposition to fluid transport. In geometrically constricted architectures, such as the 0.5D gyroid, exhibited significantly lower k and higher $\mathfrak{R}$ compared to that of the D unit-cell size gyroid.

This behavior is driven by geometric restrictions controlling fluid transport through porous networks. Permeability reflects the ability of a scaffold to accommodate flow, which is primarily controlled by the effective cross-sectional openness and continuity of pore pathways. Scaffolds with frequent constrictions reduce the hydraulic diameter, which increases geometric obstruction, leading to reduced permeability. This inverse trend reflects the theoretical relationship between structural complexity and flow as defined by Darcy's law (Eqs. (6) and (5)) and quantifies

how structural complexity influences the transport performance across scaffold designs.

The relationship between hydraulic resistance, $\mathfrak{R}$, and MRT, t_r , is shown in Figure 7b. As shown in the figure, the MRT is directly proportional to the hydraulic resistance. The four studied architectures, including flow in a simple tube, provide four (t_r , $\mathfrak{R}$) combinations that followed a linear trend. Designs exhibiting the highest $\mathfrak{R}$ generate the longest t_r values, whereas low-resistance geometries have shorter particle transit through the pore network.

This linear correlation occurs because higher hydraulic resistance reflects increased obstruction, which reduces the average flow velocity under a constant pressure drop (Eq. 7), thereby extending the time particles remain within the pore network.

From a transport standpoint, under a constant imposed pressure drop (Figure 3b), geometries with higher hydraulic resistances resulted in lower volumetric flow rates (Eq. (7)). This reduction in the flow rate for a given scaffold geometry decreases the average velocity of the flow, thereby increasing the residence time according to Eq. (4).

These observations imply that the MRT can be indirectly tuned through scaffold architecture modifications that alter the hydraulic resistance. Designs optimized for enhancing the prolonged exposure of particles inside scaffolds would therefore benefit from architectures that minimize hydraulic losses. Hence, the MRT acts as a sensitive indicator of the redistribution of hydrodynamic loading within the scaffold, and its direct increase with hydraulic resistance provides a powerful metric for predicting the transport performance across different TPMS designs.

3.2. Tortuosity and Particle Residence Time. Figure 8 illustrates the tortuosity distributions for the three architected scaffolds compared to that of a non-architected pipe ($\tau = 1$). The red dotted vertical line indicates the median tortuosity. The cases are arranged from the lowest to the highest median, and τ . The gyroid with unit-cell size D exhibited the lowest median τ , followed by the 0.5D gyroid and double gyroid. All three TPMS structures exhibited comparable distribution widths. The 0.5D

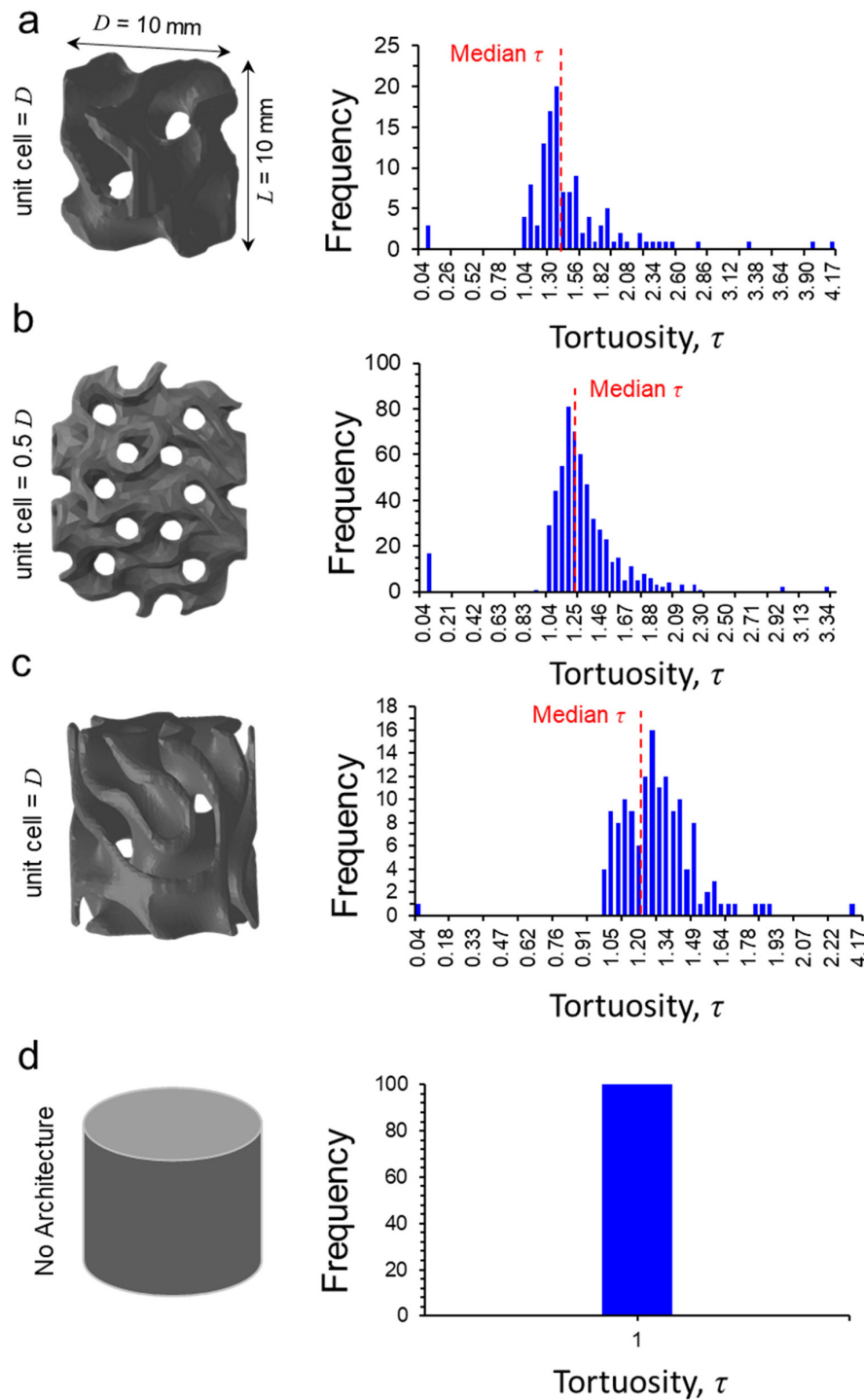


Figure 8. Tortuosity histograms and distributions for (a) a Gyroid with unit-cell size D , (b) a Gyroid with unit-cell size $0.5D$, (c) a double Gyroid with unit-cell size D , and (d) a simple pipe without the TPMS architecture. The histograms show the frequency versus tortuosity of the elements within each TPMS structure, along with their respective median tortuosity, (τ , dimensionless)

gyroid contained more surface elements in the tortuosity computation, resulting in higher raw frequency values. The raw median τ values for the three TPMS cases (Figure 8a–c) are 1.24, 1.26, and 1.34, respectively.

Figure 9 shows the MRT t_r versus the median tortuosity τ for the architected and non-architected cases (Eqs. (3) and (4), and Methods). The flow in a pipe without any internal architecture exhibited the lowest t_r and, as expected, the lowest tortuosity

($\tau = 1$). The TPMS cases display an inverse correlation between t_r and median tortuosity τ : cases with higher median tortuosity τ showed shorter residence times t_r .

Such a trend is typically associated with the conditions under which the inlet velocity (or volumetric flow rate Q) is imposed, such that Q is effectively fixed, and the pressure drop adjusts to the geometry (i.e., to different TPMSs). Under these conditions, narrow passages within the TPMS act as nozzles, locally

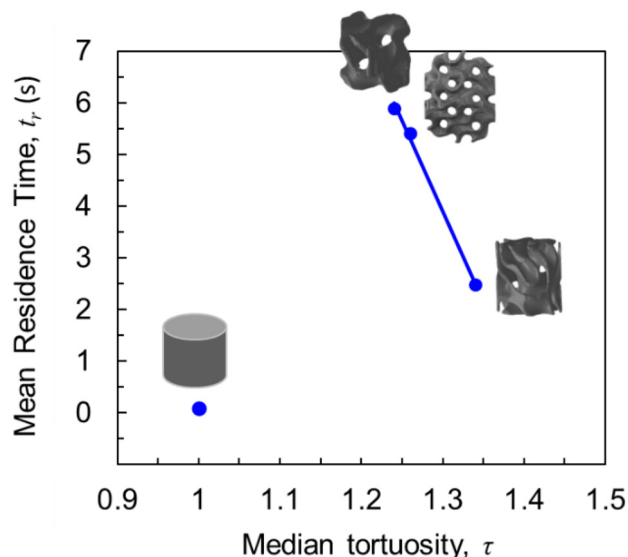


Figure 9. Mean residence time versus median tortuosity τ

accelerating the flow and enhancing advective transport, which reduces t_r as τ increases^{32–34}.

However, in our simulations, a fixed pressure difference was applied across the TPMS. While the Darcy regime theory suggests, one would then expect Q to decrease with increasing tortuosity, and thus t_r to increase with increased τ ³⁵. Observing a decreasing t_r with increased τ suggests that the imposed pressure difference placed the system in a forcing-dominated regime³⁶. In this regime, the resulting Q and local velocities are sufficiently high that the nozzle effects inside the TPMS dominate the flow. As such, this high constant ΔP has a similar effect to a velocity-inlet-pressure-outlet condition, which makes the observed reduction in t_r with increasing tortuosity plausible.

4. CONCLUSIONS

We investigated the influence of TPMS-based scaffold geometry on the transport behavior by quantifying hydraulic resistance, tortuosity, and MRT. Using a validated CFD model applied to gyroid and double gyroid structures, we demonstrated how geometric features dictate internal fluid dynamics. Our results indicate that increased hydraulic resistance correlates with decreased permeability and increased MRT, aligning with Darcy-type behavior under a fixed pressure difference.

Tortuosity analysis revealed distinct architectural differences among the three TPMS cases. Although the gyroid with unit-cell size D exhibited the lowest median tortuosity and double gyroid the highest, the MRT varied inversely with tortuosity. This is physically consistent with the trend, which appears counterintuitive but becomes physically consistent when considering the high forcing level created by the imposed pressure difference, wherein narrow TPMS passages act as nozzles that locally accelerate the flow.

Overall, this work provides a robust methodology to quantify scaffold tortuosity using fluid flow, relating it to residence time and hydraulic resistance. These relationships highlight the importance of integrating transport properties with mechanical and

biological criteria when designing TPMS scaffolds for bone tissue engineering. Future work should explore a broader range of pressure differences to examine the conditions whereby increased tortuosity scales directly with MRT and perform a direct comparison with native trabecular bone.

In summary, the analytical approach presented herein establishes a unified framework for quantifying tortuosity and residence time in architected porous scaffolds. These findings provide foundational tools for designing TPMS structures with tunable transport performance, potentially improving nutrient delivery, antimicrobial efficacy, waste clearance, and functional optimization in bone tissue engineering applications.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Btoul S. Alkhalil – Bioengineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; 0009-0004-6658-101X
Email: s202274200@kfupm.edu.sa

Corresponding Author

Ahmed S. Dalaq – Research Mentor, Assistant Professor Bioengineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; 0000-0002-3070-4215
Email: ahmed.dalaq@kfupm.edu.sa

ACKNOWLEDGMENTS

This work was supported by the Interdisciplinary Research Center of Biosystems and Machines at King Fahd University of Petroleum & Minerals, Saudi Arabia.

DATA AVAILABILITY

Data will be made available upon request

REFERENCES

- (1) Koons, G. L., Diba, M. & Mikos, A. G. Materials design for bone-tissue engineering. *Nat. Rev. Mater.* **5**, 584–603 (2020).
- (2) Baldwin, P. *et al.* Autograft, Allograft, and Bone Graft Substitutes: Clinical Evidence and Indications for Use in the Setting of Orthopaedic Trauma Surgery. *J. Orthop. Trauma* **33**, 203 (2019).
- (3) Annamalai, R. T. *et al.* Injectable osteogenic microtissues containing mesenchymal stromal cells conformally fill and repair critical-size defects. *Biomaterials* **208**, 32–44 (2019).
- (4) Garrity, C., Arzi, B., Haus, B., Lee, C. A. & Vapniarsky, N. A Fresh Glimpse into Cartilage Immune Privilege. *CARTILAGE* **13**, 119–132 (2022).
- (5) Vacanti, C. A. The history of tissue engineering. *J. Cell. Mol. Med.* **10**, 569–576 (2006).
- (6) Chan, B. P. & Leong, K. W. Scaffolding in tissue engineering: general approaches and tissue-specific considerations. *Eur. Spine J.* **17**, 467–479 (2008).
- (7) Ataollahi, S. A review on additive manufacturing of lattice structures in tissue engineering. *Bioprinting* **35**, e00304 (2023).

- (8) Kanwar, S. & Vijayavenkataraman, S. Design of 3D printed scaffolds for bone tissue engineering: A review. *Bioprinting* **24**, e00167 (2021).
- (9) Chen, H. *et al.* Porous Scaffold Design for Additive Manufacturing in Orthopedics: A Review. *Front. Bioeng. Biotechnol.* **8**, (2020).
- (10) Melchels, F. P. W. *et al.* Mathematically defined tissue engineering scaffold architectures prepared by stereolithography. *Biomaterials* **31**, 6909–6916 (2010).
- (11) Yoo, D.-J. Advanced porous scaffold design using multi-void triply periodic minimal surface models with high surface area to volume ratios. *Int. J. Precis. Eng. Manuf.* **15**, 1657–1666 (2014).
- (12) Jung, Y. & Torquato, S. Fluid permeabilities of triply periodic minimal surfaces. *Phys. Rev. E* **72**, 056319 (2005).
- (13) Coburn, B. & Salary, R. R. Mechanical Characterization of Porous Bone-like Scaffolds with Complex Microstructures for Bone Regeneration. *Bioengineering* **12**, 416 (2025).
- (14) Vijayavenkataraman, S., Zhang, L., Zhang, S., Hsi Fuh, J. Y. & Lu, W. F. Triply Periodic Minimal Surfaces Sheet Scaffolds for Tissue Engineering Applications: An Optimization Approach toward Biomimetic Scaffold Design. *ACS Appl. Bio Mater.* **1**, 259–269 (2018).
- (15) Li, K. *et al.* Optimizing mass transfer performance in triply periodic minimal surface porous scaffolds through isosurface offset. *Thin-Walled Struct.* **208**, 112800 (2025).
- (16) Zeinali, B., Soleimani Jafarabadi, H. & Bazargan, M. Novel Equations for Permeability and Elastic Modulus of Triply Periodic Minimal Surface Scaffolds via Experimental and Numerical Analysis. *Adv. Eng. Mater.* **27**, 2500414 (2025).
- (17) Zhu, J., Zou, S., Mu, Y., Wang, J. & Jin, Y. Additively Manufactured Scaffolds with Optimized Thickness Based on Triply Periodic Minimal Surface. *Materials* **15**, 7084 (2022).
- (18) Clarke, D. A., Dolamore, F., Fee, C. J., Galvosas, P. & Holland, D. J. Investigation of flow through triply periodic minimal surface-structured porous media using MRI and CFD. *Chem. Eng. Sci.* **231**, 116264 (2021).
- (19) da Silva, M. T. Q. S., do Rocio Cardoso, M., Veronese, C. M. P. & Mazer, W. Tortuosity: A brief review. *Mater. Today Proc.* **58**, 1344–1349 (2022).
- (20) Alketan, O. & Abu Al-Rub, R. MSLattice: A free software for generating uniform and graded lattices based on triply periodic minimal surfaces. *Mater. Des. Process. Commun.* **3**, (2020).
- (21) COMSOL Multiphysics® v. 6.0. COMSOL AB (2024).
- (22) Image Processing and Data Analysis Software - Dragonfly. <https://dragonfly.comet.tech/>.
- (23) Xian, G. *et al.* Gyroid-Structured Scaffolds Guide Uniform Ossification and Modulate Vascular Morphology During Rat Calvarial Bone Defect Regeneration. *J. Biomed. Mater. Res. A* **113**, e37978 (2025).
- (24) Alketan, O. & Abu Al-Rub, R. MSLattice: A free software for generating uniform and graded lattices based on triply periodic minimal surfaces. *Mater. Des. Process. Commun.* **3**, (2020).
- (25) Thai, T. H. *et al.* 3D Printed Triply Periodic Minimal Surface Scaffolds Enhanced Human Exfoliated Deciduous Teeth Stem Cell Proliferation and Differentiation for Oral and Maxillofacial Bone Tissue Engineering. *Polym. Eng. Sci.* **65**, 4374–4386 (2025).
- (26) Muñoz, A., De Paolis, A., Cardoso, L. & Carriero, A. Mechanical deformations of bone generate interstitial fluid flow at nanoscale velocities around osteocytes. *Front. Bioeng. Biotechnol.* **13**, (2025).
- (27) McCoy, R. J. & O'Brien, F. J. Influence of Shear Stress in Perfusion Bioreactor Cultures for the Development of Three-Dimensional Bone Tissue Constructs: A Review. *Tissue Eng. Part B Rev.* **16**, 587–601 (2010).
- (28) Jaasma, M. J., Plunkett, N. A. & O'Brien, F. J. Design and validation of a dynamic flow perfusion bioreactor for use with compliant tissue engineering scaffolds. *J. Biotechnol.* **133**, 490–496 (2008).
- (29) Batchelor, G. K. *An Introduction to Fluid Dynamics*. (Cambridge University Press, Cambridge, 2000). doi: 10.1017/CBO9780511800955.
- (30) Maskery, I., Aboulkhair, N. T., Aremu, A. O., Tuck, C. J. & Ashcroft, I. A. Compressive failure modes and energy absorption in additively manufactured double gyroid lattices. *Addit. Manuf.* **16**, 24–29 (2017).
- (31) Jarrar, Z. A., Al-Raoush, R. I., Hannun, J. A. & Alshibli, K. A. New model for estimating geometric tortuosity of variably saturated porous media using 3D synchrotron microcomputed tomography imaging. *Soil Sci. Soc. Am. J.* **85**, 1867–1879 (2021).
- (32) Silva, J. L. da & Santana, H. S. Residence time distribution in reactive and non-reactive flow systems in micro and millidevices. *Chem. Eng. Sci.* **248**, 117163 (2022).
- (33) Rodrigues, A. E. Residence time distribution (RTD) revisited. *Chem. Eng. Sci.* **230**, 116188 (2021).
- (34) Wang, Y., Li, C., Xue, K., Qu, X. & Liu, Y. Effect of Surface Tortuosity on Particle Dynamics in Rock Fractures. *Processes* **13**, 2702 (2025).
- (35) Zakirov, T. R. & Khranchenkov, M. G. Prediction of permeability and tortuosity in heterogeneous porous media using a disorder parameter. *Chem. Eng. Sci.* **227**, 115893 (2020).
- (36) Abdi, R., Krzaczek, M. & Abdi, M. From Darcy to turbulent flow: Investigating flow characteristics and regime transitions in porous media. *Phys. Fluids* **36**, 103372 (2024).

Hydrothermally Synthesized Ni/Co-Doped MoS₂/MoO₃ Nanosheets for NO₂ Sensing: Experimental and DFT Insights

Zahra Ali Alharbi, Retno Dwi Astuti, Md Saif Hasan, and Md Delowar Hossain*

Cite <https://doi.org/10.64589/juri/219159>

Submitted: September 12, 2025 Revised: February 22, 2026 Accepted: March 11, 2026

ABSTRACT

Two-dimensional (2D) MoS₂/MoO₃ nanosheets have attracted significant attention for room-temperature NO₂ gas sensing owing to their large surface areas and tunable electronic properties. In this work, pristine and Co-/Ni-doped MoS₂/MoO₃ nanosheets were synthesized at 240 °C via a hydrothermal method. Structural and morphological characterizations confirmed the formation of a semiconducting 2H phase in pristine MoS₂/MoO₃, while Co and Ni doping induced slight lattice distortions. Gas-sensing measurements showed that at room temperature, pristine MoS₂/MoO₃ exhibited a response of ~28% at 1 ppm, which increased to ~46% at 50 ppm of NO₂ gas concentrations. The response improved further when the operating temperature was raised to 55 °C. The doping-induced lattice strain promoted NO₂ adsorption and led to slower recovery compared with the pristine sample. Selectivity tests indicated a stronger response to NO₂ than to ammonia (NH₃), acetone (C₃H₆O), ethylene (C₂H₄), and methane (CH₄) under identical conditions. Density functional theory (DFT) calculations revealed that doping modified the Mo-S bond lengths and electronic structure, thereby affecting the adsorption and charge-transfer behavior, consistent with our experimental observations. These findings provide insights into how subtle structural modifications influence the gas-sensing performance, guiding the design of high-performance MoS₂-based sensors and other transition-metal-sulfide systems.

Keywords: transitional metal chalcogenides, hydrothermal synthesis, gas sensor, density functional theory, nitrogen dioxide, electronic structure

1. INTRODUCTION

Air pollution is an escalating threat to human health and the environment. In 2019, the World Health Organization (WHO) estimated that outdoor air pollution contributed to approximately 4.2 million premature deaths globally, underscoring the urgent need for continuous monitoring of key air pollutants, such as nitrogen dioxide (NO₂). As a highly reactive and corrosive gas, NO₂ poses significant risks to human respiratory and cardiovascular systems, even at very low concentrations. Exposure to NO₂ in children has been linked to increased incidence of eczema, pneumonia, and other respiratory illnesses, as well as high mortality¹⁻³. In addition, NO₂ adversely affects terrestrial and aquatic ecosystems and is a major atmospheric pollutant contributing to environmental challenges such as the greenhouse effects⁴. These health and environmental risks highlight the urgent need for rapid and accurate NO₂ monitoring systems.

To address this need, a wide range of gas-sensing technologies have been developed in recent decades, including electrochemical^{5,6}, optical^{7,8}, acoustic^{9,10}, field-effect-transistor-based¹¹⁻¹³, and metal-oxide semiconductor chemoresistive platforms¹⁴⁻¹⁶. Despite their progress, each approach faces inherent limitations: electrochemical sensors suffer from short lifetimes, signal drift, and sensitivity to humidity; optical methods remain costly and complex, requiring precise alignment and long optical paths;

acoustic sensors demand intricate fabrication and sensitive packaging; and field-effect transistor-based devices encounter fabrication complexity, stability and hysteresis issues, and challenges related to gas accessibility⁵⁻¹⁶. Among available technologies, metal oxide-based chemiresistive gas sensors have been widely used owing to their low cost, compact design, and simple electrical readouts. In most cases, the active layers in such sensors are metal oxide semiconductors, such as SnO₂¹⁷⁻¹⁹, In₂O₃²⁰⁻²², and ZnO²³⁻²⁵, which offer high sensitivity, wide detection ranges, and good stability. However, these sensors require elevated operating temperatures (often up to 400 °C), which demand additional heating systems that increase cost, power consumption, and safety concerns. Consequently, conventional metal-oxide-semiconductor-based chemiresistive sensors are not suitable for portable and low-power applications²⁶.

To overcome these challenges, two-dimensional (2D) transition metal chalcogenides (TMDs), such as MoS₂, have gained significant attention. A recent study shows that hydrothermally synthesized 2D MoS₂ exhibits tunable bandgap, strong adsorption affinity, and abundant defect sites, enabling highly sensitive detection of various oxidizing gases²⁷. Moreover, defect engineering further improves room-temperature sensitivity by increasing the number of active adsorption sites and strengthening the electron transfer pathways²⁸. Consequently, hydrothermally synthesized MoS₂-based nanomaterials have demonstrated

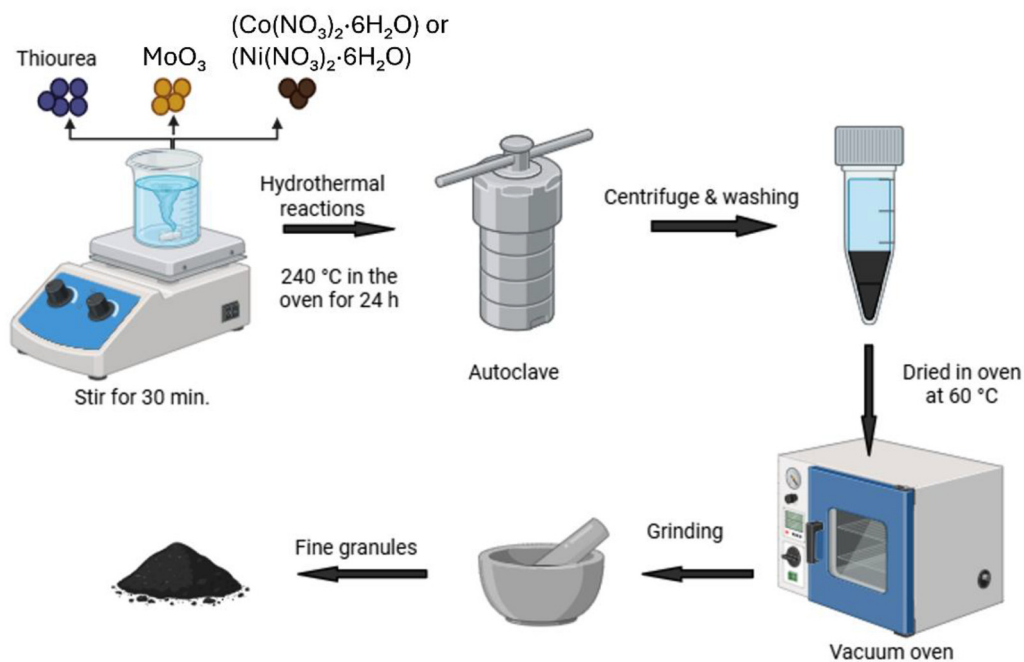


Figure 1. Schematic illustration of the hydrothermal synthesis of pristine and doped $\text{MoS}_2/\text{MoO}_3$

strong potential for room-temperature NO_2 sensing. During hydrothermal synthesis, the MoO_3 precursor may not be completely converted to MoS_2 , leading to the coexistence of MoS_2 and residual MoO_3 . Such mixed $\text{MoS}_2/\text{MoO}_3$ systems can modify electron coupling and adsorption behavior, potentially enhancing the sensitivity to oxidizing gases^{29,30}. Recent studies have further demonstrated that engineering $\text{MoS}_2/\text{MoO}_3$ heterostructures is an effective strategy to boost low-temperature NO_2 sensing by strengthening the interfacial charge transfer and increasing the number of accessible adsorption sites. For example, creating $\text{MoS}_2/\text{MoO}_3$ heterojunctions via controlled surface oxidation can enhance room-temperature sensitivity of $\sim 18.9\%$ response at 1 ppm (parts per million) NO_2 , provide fast response/recovery, and enable detection down to the ppb (parts per billion) range. This highlights the importance of an optimized heterostructure and preserved 3D morphology in accelerating carrier transport and gas-surface interactions³¹. Complementary

results have also been reported for $\text{MoS}_2/\text{MoO}_3$ hybrid architectures, which show a p-type response and improved selectivity attributed to interfacial charge redistribution and the formation of a hole-rich region in MoS_2 during NO_2 exposure³². In addition, hydrothermally derived $\text{MoS}_2/\text{MoO}_3$ nanocomposites have been used for NO_2 detection at mild temperatures (50°C), where heterojunction effects and defect-assisted adsorption/desorption pathways strongly influence sensing polarity, magnitude, and reversibility³³. Furthermore, $\text{MoS}_2/\text{MoO}_3$ composites have been integrated into self-powered sensing configurations, indicating that the heterostructure design can be coupled with stable signal transduction for practical monitoring under ambient conditions³⁴. The structural phase of MoS_2 is significantly affected by the reaction temperature and sulfur availability. Higher temperatures favor a stable 2H (semiconducting) phase, whereas lower temperatures or sulfur-deficient conditions promote the formation of a metastable 1T (conducting) phase³⁵.

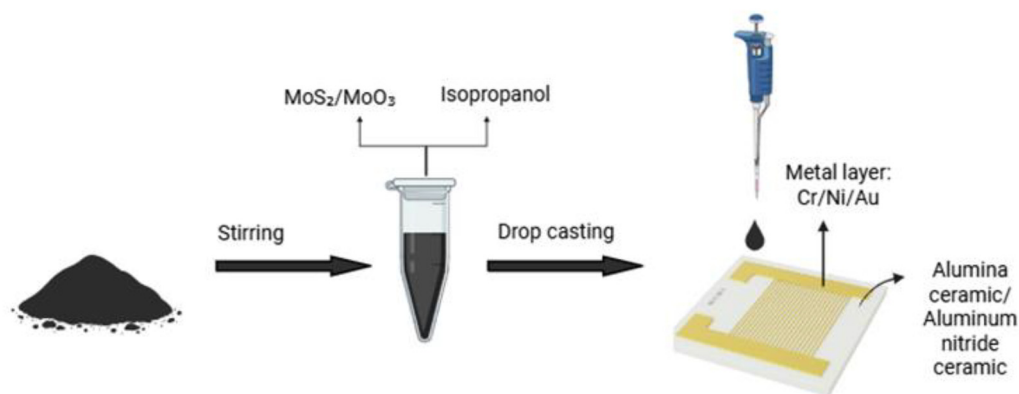


Figure 2. Preparation of $\text{MoS}_2/\text{MoO}_3$ -based gas sensors via drop-casting onto Au-IDEs

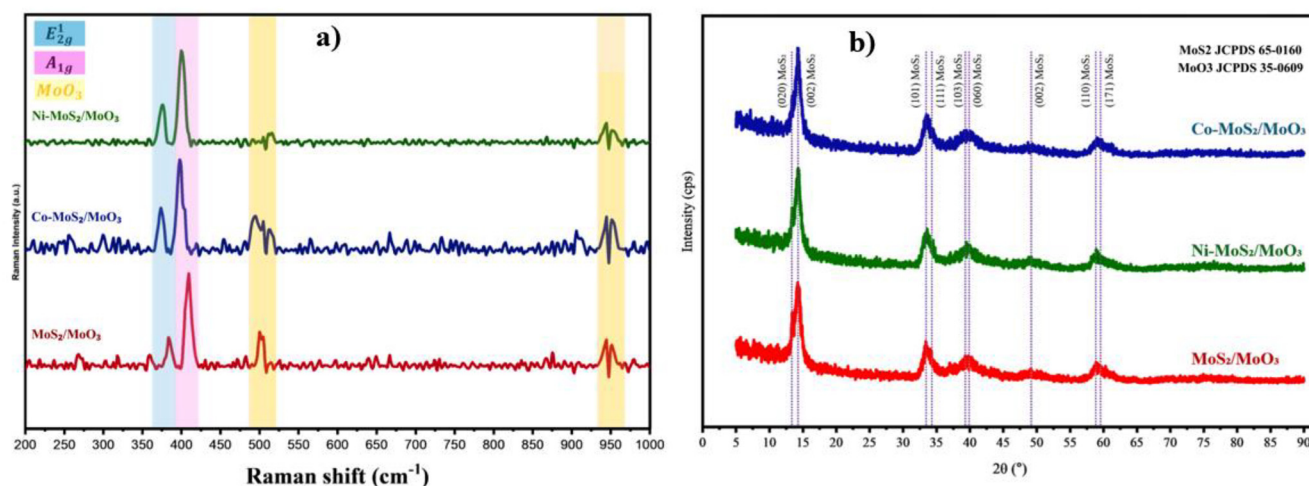


Figure 3. (a) Raman spectra and (b) XRD patterns of pristine (red), Ni-doped (green), and Co-doped MoS₂/MoO₃ (blue)

In parallel with experimental advances, computational approaches have been extensively utilized to elucidate the gas-sensing behavior of TMDs. Density functional theory (DFT) calculations on pristine and defect-engineered MoS₂ consistently show that S vacancies markedly enhance NO₂ adsorption and charge transfer, resembling most active sites. These predictions are consistent with experimental findings^{36,37}. Another DFT study investigated oxygen-incorporated monolayer MoS₂ using both finite cluster and periodic calculations, and the results show that progressive S→O substitution reduces the band gap, while maintaining charge transfer pathways. The subsequent bond rearrangements in MoS_{2-x}O_x layers, which lead to MoO₃ formation, relieve lattice strain and stabilize the resulting MoS₂/MoO₃ heterostructures³⁸.

Despite the above experimental and computational advances, several aspects of MoS₂/MoO₃-based NO₂ sensing systems remain unclear. First, although MoS₂/MoO₃ heterostructures have been widely studied for low-temperature NO₂ sensing, most reports focus on oxidation-derived interfaces, surface functionalization, or device integration. The effects of transition-metal doping (such as Co and Ni) on adsorption strength, selectivity, charge transfer, and overall room-temperature response have not been systematically investigated. Second, the sensing performance reported experimentally is not always directly linked to atomistic-level mechanisms, and a unified picture that correlates dopant-induced structural/electronic changes (from DFT) with measured NO₂ sensing performance remains limited. Addressing these gaps is important for the rational design of MoS₂-based sensors with predictable low-temperature performances.

In this study, we combined experimental and computational approaches to investigate the NO₂ sensing performance of hydrothermally synthesized pristine, Co-doped, and Ni-doped MoS₂/MoO₃. The obtained powders were comprehensively characterized using Raman spectroscopy, X-ray diffraction (XRD), scanning electron microscopy (SEM), and transmission electron microscopy (TEM). The gas-sensing performance is evaluated at various NO₂ concentrations and two different temperatures (25° and 55 °C). Moreover, the selectivity of the MoS₂/MoO₃ sensors toward NO₂ is compared with other interfering gases. To complement the experiments, DFT calculations

were performed to determine the NO₂ adsorption configurations and associated electronic structure changes, including band gap modulation and density of state (DOS) evolution before and after NO₂ adsorption. Taken together, these results provide an integrated understanding of low-temperature NO₂ sensing on MoS₂/MoO₃ and offer a robust foundation for the design of next-generation 2D gas sensors.

2. EXPERIMENTAL AND COMPUTATIONAL METHODOLOGY

2.1. Hydrothermal Synthesis of MoS₂/MoO₃.

2.1.1. Preparation of pristine MoS₂/MoO₃ without doping. To prepare a homogenous precursor solution, 0.046 g of MoO₃ powder (≥99.5%, ACS Reagent Grade) and 0.062 g of thiourea (≥99.0%, ACS Reagent Grade) were dissolved in 32.5 mL of deionized water and stirred for 30 min. Then, the obtained mixture was transferred to a 50 mL stainless-steel autoclave and heated in an oven at 240 °C for 24 h. After naturally cooling to room temperature, the resulting black precipitate was collected by centrifugation at 38,000 rpm and repeatedly washed using ethanol and deionized water until the pH was neutral (Fig. 1). Finally, the products were dried in an oven at 60 °C and ground into a fine powder.

2.1.2. Synthesis of MoS₂/MoO₃ with Co and Ni doping. Before hydrothermal treatment, 5 wt.% of cobalt nitrate hexahydrate (Co(NO₃)₂•6H₂O) or nickel nitrate hexahydrate (Ni(NO₃)₂•6H₂O) was directly added to the raw material solution. The hydrothermal reaction for both doped samples was conducted at 240 °C, and subsequent centrifugation, washing, drying, and grinding were performed following the same procedure as used for the pristine sample.

2.1.3. Preparation of gas sensors. A drop-casting method was employed to fabricate MoS₂/MoO₃-based gas sensors. First, 0.006 g of doped or undoped-MoS₂/MoO₃ was dispersed in 100 μL of isopropanol-2 and sonicated for 30 min to obtain a well-dispersed solution. Then, 10 μL of the obtained solution was drop-cast onto a gold interdigitated electrode (Au-IDE) and

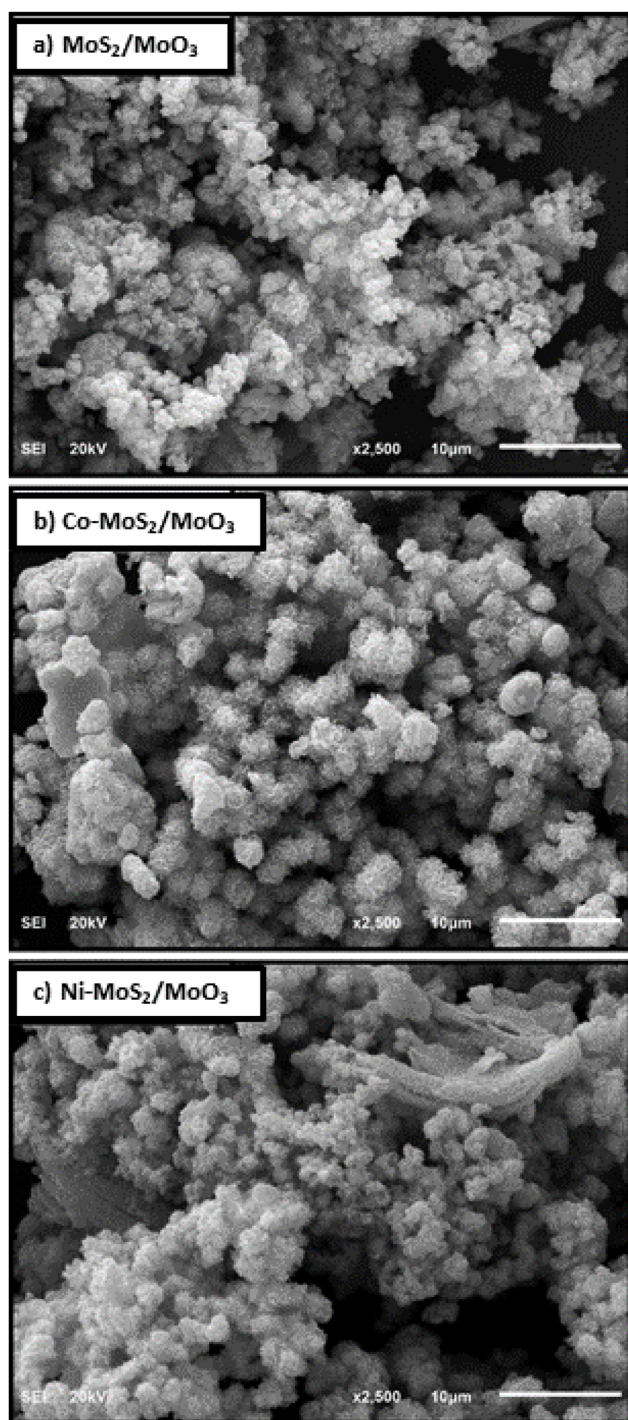


Figure 4. SEM images of a) $\text{MoS}_2/\text{MoO}_3$, b) $\text{Co-MoS}_2/\text{MoO}_3$, and c) $\text{Ni-MoS}_2/\text{MoO}_3$

dried at room temperature to form a homogeneous sensing layer (Fig. 2).

2.2. Computational Details. All spin-polarized DFT calculations were performed using the Vienna Ab initio Simulation Package (VASP) code, version 5.4.4^{39,40}. The electron-ion interactions were described using the projector augmented-wave (PAW) method⁴¹, and the exchange-correlation functional was treated within the generalized gradient approximation (GGA)

using the Perdew (Burke) Ernzerhof parameterization. A plane-wave basis set with a kinetic energy cutoff of 500 eV was used for all calculations. Geometry optimization was conducted using a conjugated gradient algorithm with a $3 \times 3 \times 1$ Monkhorst-Pack k-point mesh for the Brillouin zone sampling. The convergence criteria for electronic and ionic relaxations were set to 10^{-5} eV and -0.02 eV/Å, respectively. The vacuum region above the MoS_2 slab was set to 20 Å. Gaussian smearing was used for all systems. The DFT-D3 method with Becke–Johnson damping was included to account for the van der Waals interactions, which are critical for describing the adsorption phenomena.⁴² Electronic structure analyses, including the calculation of the density of states (DOS), projected density of states (pDOS), and electron localization function (ELF), were performed on the fully optimized structures using a denser $13 \times 13 \times 1$ k-point mesh. The band structure (pDOS) was calculated using the VASP kit (version 1.5).⁴³

3. RESULTS AND DISCUSSION

3.1. Structural Analysis of Pristine and Ni/Co-doped $\text{MoS}_2/\text{MoO}_3$. The structural characteristics of the synthesized samples were examined using Raman spectroscopy and XRD. As shown in Fig. 3a, the Raman spectrum of undoped $\text{MoS}_2/\text{MoO}_3$ exhibits two distinct peaks at ~ 378 and ~ 403 cm^{-1} , corresponding to the E_{2g}^1 and A_{1g} vibrational modes of a 2H- MoS_2 phase, respectively. This confirms that the hydrothermal temperature is sufficient to produce the 2H semiconducting phase. The difference between the two peaks ($\Delta\omega$) is approximately 25 cm^{-1} , indicating the formation of a few layers of MoS_2 , rather than bulk MoS_2 .⁴⁴ An additional peak at ~ 500 cm^{-1} and a broad low-intensity band from 600 to 900 cm^{-1} correspond to the M-O vibrations of MoO_3 , indicating partial retention of the MoO_3 precursor.

Under the same hydrothermal conditions, we subsequently prepared 5 wt.% Co- and 5 wt.% Ni-doped $\text{MoS}_2/\text{MoO}_3$ to investigate the effect of transition-metal doping on the structure. The spectrum of the Co-doped sample (blue line) displays two characteristic peaks at 373.6 and 397.7 cm^{-1} , which are assigned to the in-plane E_{2g}^1 mode and out-of-plane A_{1g} mode of 2H- MoS_2 , respectively. For the Ni-doped sample (green line), these modes were observed at 375.59 and 399.6 cm^{-1} . In both cases, the presence of the E_{2g}^1 and A_{1g} modes confirm that the 2H- MoS_2 structure was preserved after Co and Ni incorporation, demonstrating the successful synthesis of doped 2H- MoS_2 . The slight shifts in peak positions and changes in linewidth relative to those of the pristine sample reflect the influence of dopant incorporation on the local lattice environment and electron-phonon interactions, providing further evidence of effective Co and Ni doping into the 2H- MoS_2 lattice. Additionally, both spectra exhibit characteristic bands at around ~ 500 and ~ 950 cm^{-1} , verifying the coexistence of MoO_3 and thus the formation of $\text{MoS}_2/\text{MoO}_3$.^{45,46}

The crystalline phases of the synthesized $\text{MoS}_2/\text{MoO}_3$ were examined by XRD, and the corresponding patterns are presented in Fig. 3b. All three samples exhibit a set of diffraction peaks that can be indexed to the standard JCPDS patterns of MoS_2 and MoO_3 .⁴⁷ The intense reflection in the low-angle region around

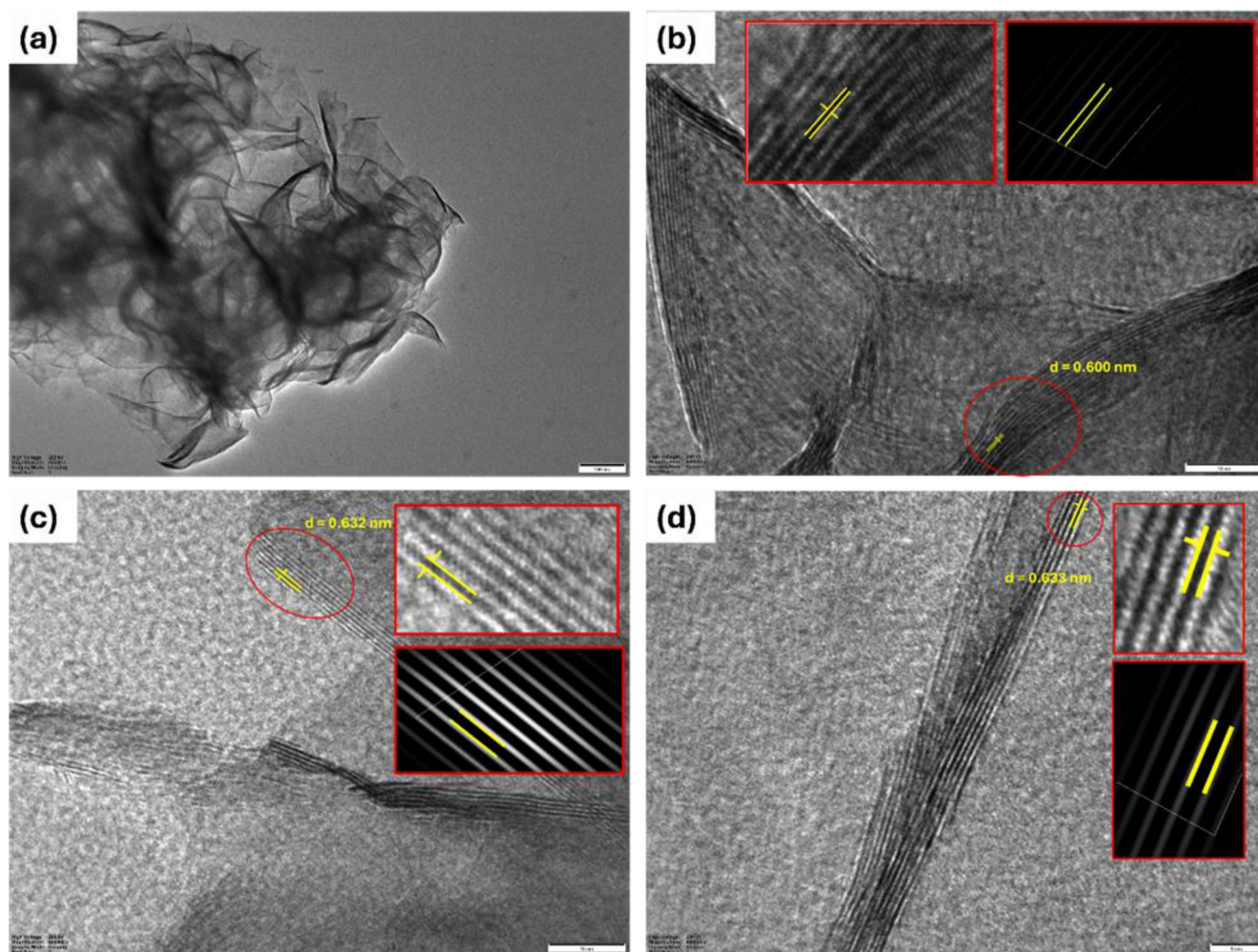


Figure 5. (a) TEM image of a pristine sample showing thin, wrinkled nanosheets. (b) HRTEM images of pristine MoS₂ with a lattice spacing of ~0.600 nm. (c) Co-doped sample with a slightly expanded spacing of ~0.632 nm. (d) Ni-doped sample with a spacing of ~0.633 nm

$2\theta \approx 14^\circ$ is assigned to the (002) plane of MoS₂, while additional peaks at higher angles correspond to the (101), (103), and (110) planes of MoS₂. The diffraction peaks ascribed to the (020), (060), (002), and (171) planes of the MoO₃ phase are also observed. The diffraction patterns of Co- and Ni-doped samples closely resemble that of undoped MoS₂/MoO₃, with no additional peaks related to secondary or impurity phases, indicating that the MoS₂/MoO₃ crystal structure was preserved upon doping. The observed minor changes in peak intensity and line broadening may be associated with lattice distortion or crystallite size effects due to the incorporation of Co and Ni species⁴⁷. Overall, XRD analysis confirms the successful formation of MoS₂/MoO₃ heterostructures, which remain intact upon Co and Ni doping.

3.2. Morphological and Microstructural Analysis.

SEM was employed to examine the morphologies of the as-synthesized samples (Fig. 4a–c). All three samples (pristine, Co-doped, and Ni-doped) exhibit similar morphological features, characterized by agglomerated, porous, and granular structures at a magnification of 2500. The particles appear as loosely stacked, flower-like assemblies, indicating the formation of multilayered MoS₂/MoO₃.

To complement the above observations, the fine structural features and crystallinity of the synthesized samples were investigated by TEM (Fig. 5). Pristine MoS₂/MoO₃ (Fig. 5a) consists of few-layered nanosheets with transparent areas and stacked, slightly folded morphologies, indicating the formation of layered structures with large surface areas, which can provide more active sites for gas adsorption during sensing. The high-resolution TEM image of pristine MoS₂/MoO₃ (Fig. 5b) also confirms the layered structure. The lattice fringes are observed clearly with an interlayer spacing of 0.6 nm corresponding to the (002) plane of MoS₂, which confirms the high crystallinity of the material. The presence of connected and uniform fringes supports the formation of a 2H (semiconducting) phase at 240 °C, consistent with the Raman results.

The overall morphologies of the doped samples (Fig. 5c–d) remain similar to that of the pristine MoS₂/MoO₃, despite slight variations in the interlayer spacing (0.632 nm for Co-doped MoS₂/MoO₃ and 0.633 nm for Ni-doped MoS₂/MoO₃). The increased d-values are attributed to localized stress or distortion in the Mo-S lattice, which weakens the bonds. Such distortion often occurs when foreign metal ions are incorporated into the structure, creating higher-activity sites that enhance the reaction

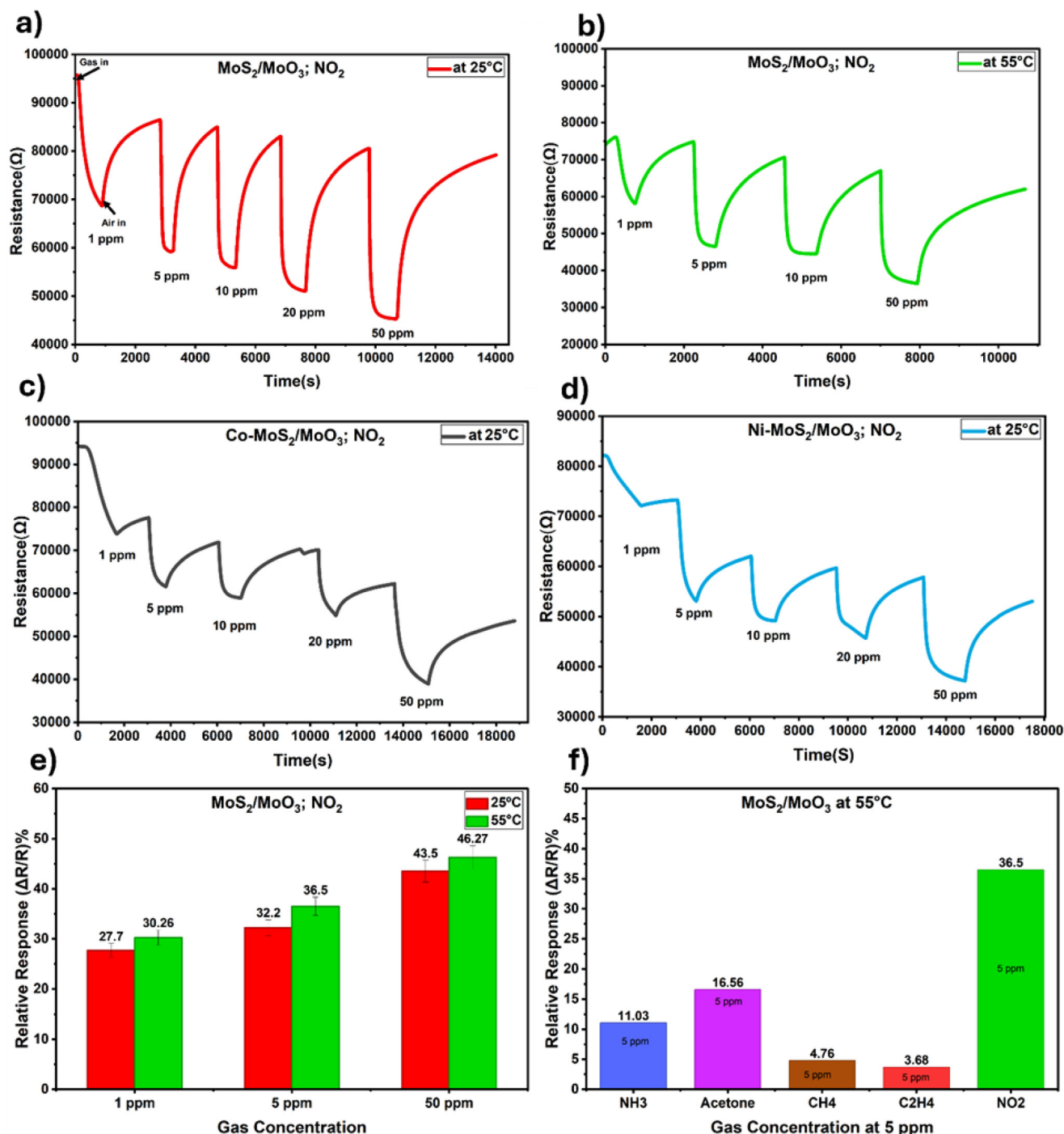


Figure 6. Gas-sensing analysis and performance of the $\text{MoS}_2/\text{MoO}_3$ sensor. (a–b) Dynamic NO_2 response of pristine $\text{MoS}_2/\text{MoO}_3$ at 25 °C and 55 °C. (c–d) Responses of Co- and Ni-doped samples at 25 °C showing enhanced sensitivity. (e) Relative response of pristine sample at 25 °C and 55 °C at various NO_2 concentrations. (f) Selectivity at 55 °C toward different gases (5 ppm)

with the gas and reduce the recovery time. Despite this distortion, the layered structure remains intact and highly crystalline.

In summary, TEM images confirm that all samples fabricated at 240 °C exhibit layered nanostructures, with minor lattice expansion observed upon the incorporation of Co or Ni, which may affect the gas-sensing performances of doped samples.

3.3. Gas Sensing Analysis.

3.3.1. Response calculation and mechanism. In this study, the gas-sensing performance was evaluated using a

chemiresistive configuration, in which sensing is achieved by monitoring the change in the electrical resistance of $\text{MoS}_2/\text{MoO}_3$ layer when gas molecules chemically interact with the surface. All measurements were performed under atmospheric pressure. The prepared $\text{MoS}_2/\text{MoO}_3$ was deposited onto an Au-IDE substrate inside a sealed test chamber and exposed to air until its resistance stabilized. In this study, an approximate stabilization period of 5000 s was used. After stabilization, NO_2 at different concentrations was introduced into the chamber. Once the gas reacts with the sensor surface, the resistance

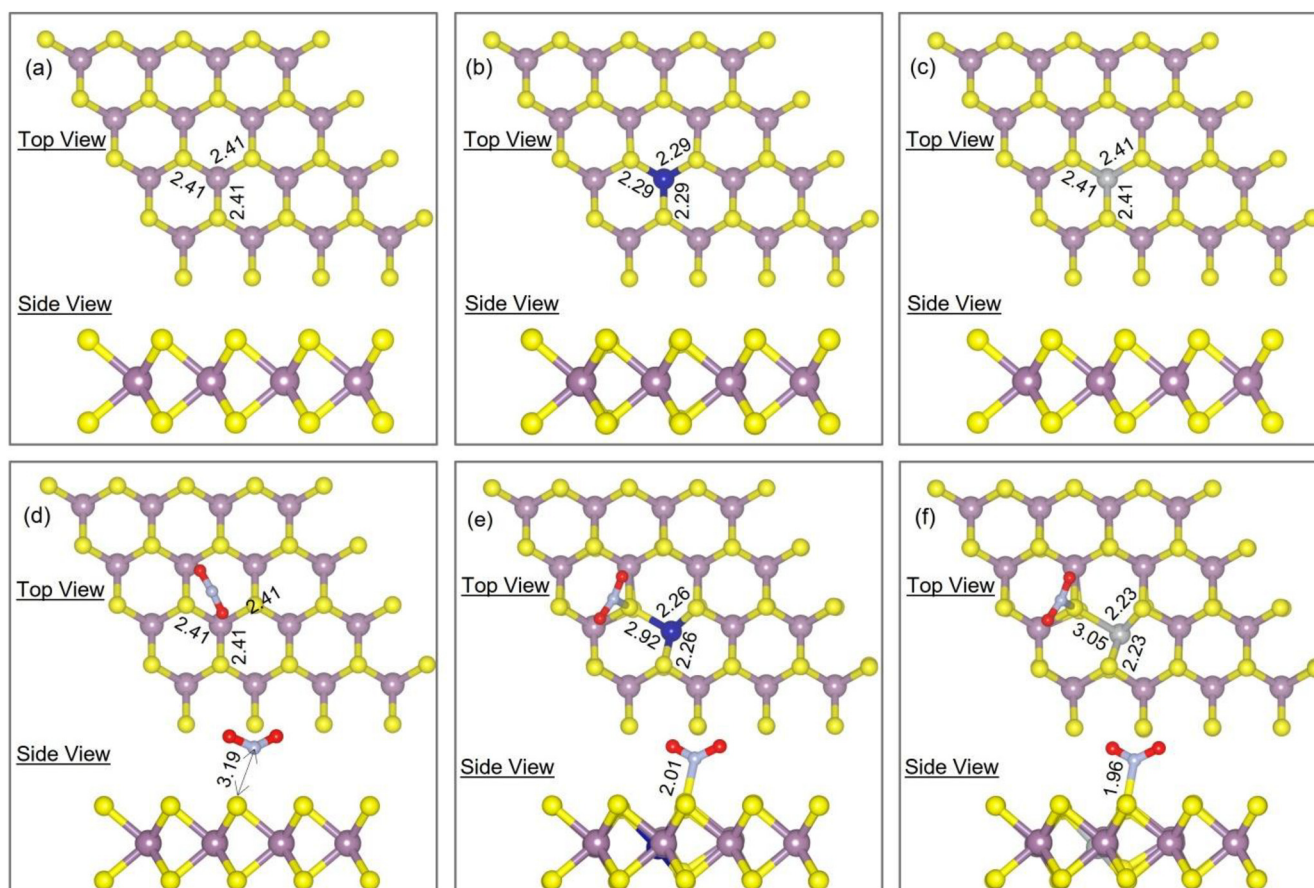


Figure 7. Optimized atomic structures and adsorption-induced bond reconstruction. (a–c) Top and side views of the optimized geometries before NO₂ adsorption for (a) pristine 2H-MoS₂, (b) Cobalt-doped MoS₂ (Co-MoS₂), and (c) Nickel-doped MoS₂ (Ni-MoS₂). Key bond lengths (in Å) are indicated. (d–f) Corresponding structures after NO₂ adsorption

begins to decrease, indicating p-type behavior. The resistance was recorded when it reached a constant level. When the gas was turned off, the chamber was cleaned with a flow of air to desorb gas molecules and allow the sensor to return to its original resistance. This cycle was repeated for each NO₂ concentration to measure the repeatability and recovery of the sensor. The relative sensor response was calculated using the following equation:

$$\text{Response (\%)} = \frac{R_{\text{air}} - R_{\text{gas}}}{R_{\text{air}}} \times 100, \quad (1)$$

where R_{air} is the resistance of the sensor under air flow, and R_{gas} is the resistance of the sensor in the presence of gas analytes. A higher relative response value indicates a strong interaction between the sensor and the gas. All the relative response graphs were calculated using the above equation.

3.3.2. Response behavior of the pristine MoS₂/MoO₃ sample. The pristine MoS₂/MoO₃ sample (Fig. 6a) shows a dynamic response at room temperature. When an oxidizing gas (NO₂) was introduced, it was absorbed on the available active sites. As a strong electron-withdrawing gas, NO₂ accepts electrons from the sensor substrate and forms adsorbed NO₂⁻ species. Electron withdrawal from the p-type material increases the hole concentration, leading to a significant decrease in resistance. The rapid interaction between gas molecules and the sensor surface

results in a prompt response at room temperature, consistent with the expected behavior of a p-type material exposed to an oxidizing gas.

The response gradually increased with gas concentration, starting from ~28% at 1 ppm to ~46% at 50 ppm. This proves that the pristine sample can detect NO₂ across different concentrations and that the response increases proportionally with the gas concentration.

3.3.3. Response and recovery kinetics. The response and recovery times were extracted from the dynamic resistance–time curves, as shown in Fig. 6(a–d). The response time (t_{90}) was defined as the time required for the sensor resistance to reach 90% of its total resistance after NO₂ exposure. Similarly, the recovery time (t_{90}) was defined as the time required for the resistance to return to 90% of its original baseline value after the NO₂ flow was stopped and the chamber was purged with air.

For the pristine MoS₂/MoO₃ sensor at room temperature (Fig. 6a), the response time was approximately 70–90 s at 1 ppm NO₂, which decreased to ~50–60 s at 50 ppm, indicating faster adsorption kinetics at higher gas concentrations. The corresponding recovery time increased from ~120–150 s at 1 ppm to ~200–230 s at 50 ppm, reflecting stronger gas–surface interaction at elevated NO₂ levels. Similar tests were conducted at a high temperature (55 °C) to examine its effect on sensing performance (Fig. 6b, discuss in 3.3.4 section).

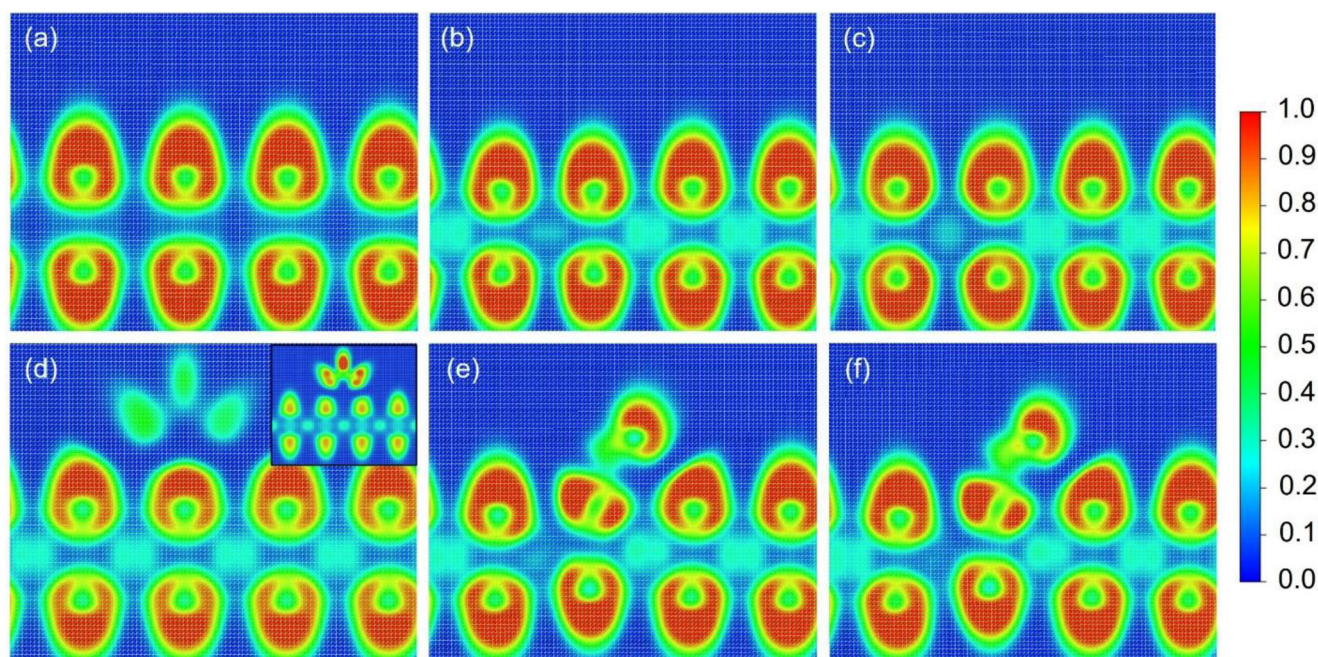


Figure 8. Electron Localization Function (ELF) maps of pristine and doped MoS₂ structures before and after NO₂ adsorption. (a–c) ELF iso-surfaces for (a) pristine 2H-MoS₂, (b) Co-doped MoS₂, and (c) Ni-doped MoS₂. (d–f) ELF maps after NO₂ adsorption on (d) pristine, (e) Co-doped, and (f) Ni-doped MoS₂. All slices correspond to the (1 0 0) plane at a distance of 3.98 Å from the origin. To clearly visualize the ELF of the NO₂ molecule in (d), the inset shows the (1 0 0) slice at 5.53 Å from the origin

In comparison, the Co- (Fig. 6c) and Ni-doped (Fig. 6d) sensors exhibited similar response times but noticeably longer recovery times at room temperature. At 1 ppm NO₂, the recovery times were approximately 200–240 s for Co-doped and 230–270 s for Ni-doped samples, which further increased at higher concentrations. This behavior suggests that transition-metal doping strengthens NO₂ adsorption, leading to slower desorption and prolonged recovery.

3.3.4. Effect of operating temperature. The effect of temperature has studied to examine sensing performance. The pristine MoS₂/MoO₃ sensor exhibits a shorter response time and improved recovery at 55 °C (Fig. 6b) compared with 25 °C (Fig. 6a). This enhancement can be attributed to increased surface activation for faster charge transfer at a higher temperature, which facilitates the interaction between NO₂ molecules and the active sites on the gas sensor. However, the response magnitude does not increase significantly when the temperature is elevated from 25 to 55 °C (Fig. 6e), indicating that the sample is sufficiently active at room temperature and that thermal activation plays a limited role. This suggests that increasing the operating temperature provides no additional advantages for sensing, with temperature primarily affecting the reaction kinetics.

3.3.5. Effect of Co and Ni doping on sensing behavior. The Ni- and Co-doped samples were tested at room temperature (Fig. 6c–d) to study the effects of doping. Both samples retained the p-type response pattern and repeatable cycles observed in the pristine sample, and the response magnitude increased with the NO₂ concentration. Although the doped samples did not show a higher response magnitude than the undoped one, their recovery was slower. As shown in the TEM images, the incorporation

of Co or Ni caused a slight expansion in lattice spacing, generating additional defect sites. These structural changes alter local charge distribution and create new adsorption sites, affecting the adsorption kinetics of NO₂. The additional defect sites bind to NO₂ molecules more strongly and trap them for a longer period, increasing the stability of adsorbed NO₂[−] species and making desorption less efficient.

3.3.6. Selectivity. To examine the selectivity of the MoS₂/MoO₃ sensor, it was exposed to several common toxic gases, including NH₃, acetone, C₂H₄, and CH₄, each at a constant concentration of 5 ppm and a test temperature of 55 °C (Fig. 6f). The response to NO₂ was significantly higher than that to the other gases. This selectivity toward NO₂ can be explained by the strong oxidizing nature and high electron affinity of NO₂, which allow more effective reactions with active sites on the MoS₂/MoO₃ surface, particularly sulfur edge sites and oxygen-containing groups. These sites tend to attract electrons, increasing the hole concentration in the p-type sensing layer and resulting in greater changes in resistance. By contrast, reducing gases, such as NH₃, or weakly interacting molecules, such as CH₄, C₂H₄, and acetone, induce less charge transfer, resulting in weaker sensing responses. These observations indicate that the prepared sensor exhibits good selectivity, even at low temperatures. This is a crucial requirement for portable and low-power NO₂ detection.

Notably, the selectivity comparison in Fig. 6f is based on the responses measured under the same experimental conditions (including operating temperature, gas concentration, and baseline environment) for all analytes. Systematic evaluation under controlled relative humidity (RH) levels is an important topic for future work to quantify humidity-induced baseline variations and

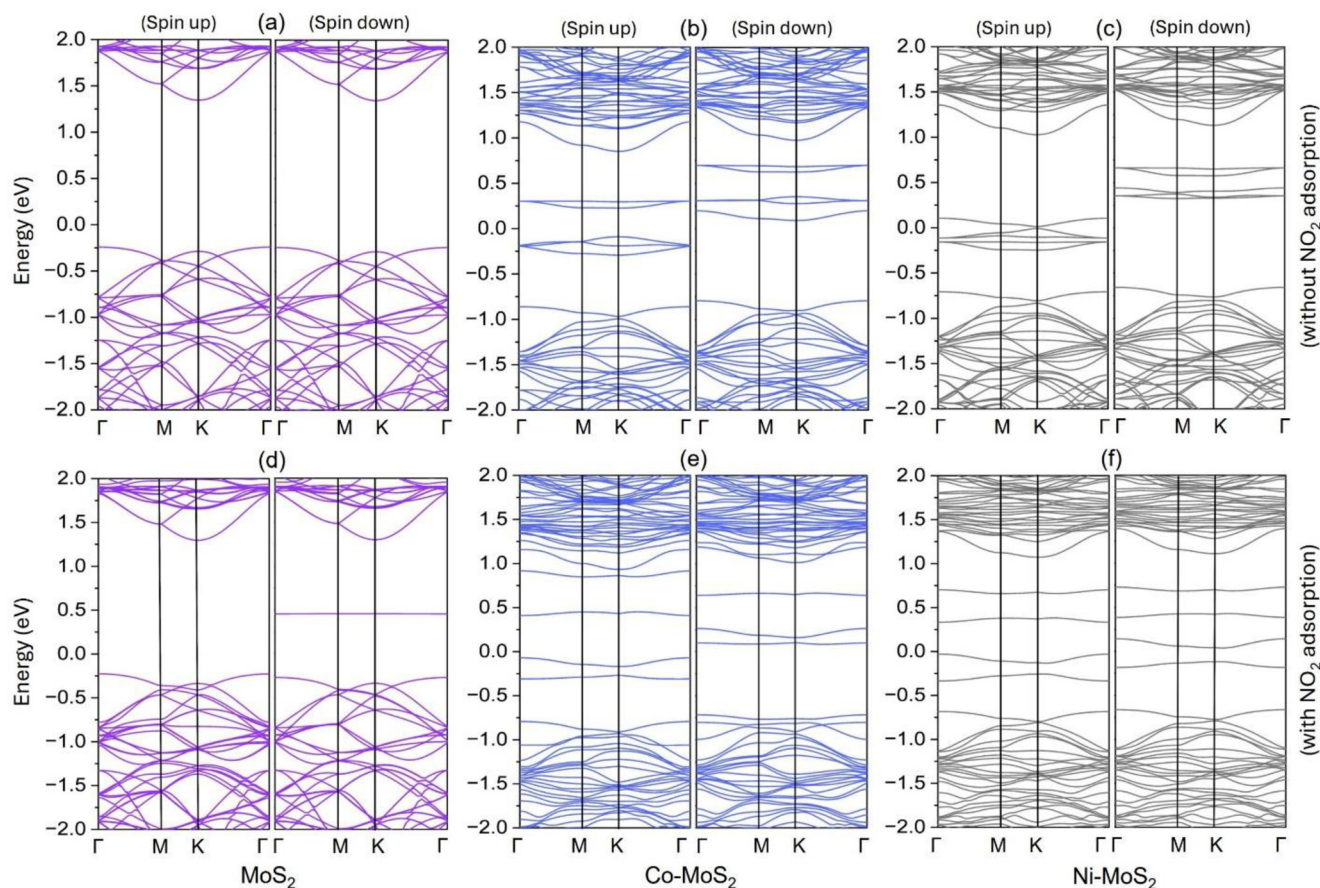


Figure 9. Electronic band structure analysis of pristine and doped MoS₂ structures before and after NO₂ adsorption. (a–c) Band structures of (a) pristine, (b) Co-doped, and (c) Ni-doped MoS₂ before NO₂ adsorption. Doping introduces spin-polarized mid-gap states, narrowing the band gap and inducing magnetic moments, with Ni doping showing more pronounced spin splitting with bands crossing the Fermi level. (d–f) Band structures after NO₂ adsorption on (d) pristine, (e) Co-doped, and (f) Ni-doped MoS₂

assess long-term environmental robustness of the MoS₂/MoO₃ sensor.

3.4. Computational Analysis of Pristine and Doped MoS₂: Structural, Electronic, and Adsorption Behavior.

Geometric optimization of the pristine and doped 2H-MoS₂ systems revealed distinct structures due to transition metal doping (Fig. 7a–c). The pristine 2H-MoS₂ (Fig. 7a) lattice served as a stable reference, with the Mo-S bond lengths consistently measured at ~2.41 Å. Substitution with Co (Fig. 7b) caused profound local lattice distortion. The resulting Co-S bond length was contracted to ~2.29 Å because of the smaller covalent radius of Co compared with that of Mo. This contraction induced a strong, localized lattice strain, thereby creating a potent defect site. By contrast, substitution with iso-valent Ni (Fig. 7c) did not contract the Ni-S bond lengths, which remained at ~2.41 Å, nearly identical to the original Mo-S bonds, despite the smaller size of Ni. The lack of measurable contraction suggests a complex electronic reorganization at the Ni site that effectively compensates for the expected geometric strain.

The adsorption of NO₂ molecules (Fig. 7d–f) also depends on the doping elements in 2H-MoS₂. For pristine 2H-MoS₂ (Fig. 7d), NO₂ adsorption was weak (physisorption), inducing a negligible change in the Mo-S framework (the bond length remained at ~2.41 Å). In the Co-doped system (Fig. 7e), the

already contracted Co-S bonds (2.29 Å) underwent a substantial elongation to ~2.92 Å upon NO₂ adsorption. This increase of ~0.63 Å is a hallmark of strong chemisorption. The NO₂ molecule interacts strongly with the S site, weakening the Co-S bonds as a new N-O bond is formed. The Co-doped site acts as a sacrificial anchor, and the adsorption energy is partially compensated through the relaxation of pre-existing strain. By contrast, the Ni-doped system (Fig. 7f) shows adsorption-induced structural instability. Upon NO₂ adsorption, the local symmetry was completely broken. Specifically, one Ni-S bond elongated dramatically to 3.05 Å, while the remaining two Ni-S bonds contracted sharply to 2.23 Å. This bifurcation of bond lengths transforms the Ni site from symmetric coordination to a highly asymmetric reconstructed geometry. The newly formed N-S bonds replace the existing Ni-S bonds, implying the stronger chemisorption of NO₂ on the Ni-doped 2H-MoS₂ structure. Compared with pristine and Co-doped MoS₂, the distance between the NO₂ molecule and the basal plane of Ni-doped MoS₂ is the shortest (1.96 Å), confirming the stronger adsorption.

To investigate the electronic structure of the MoS₂ basal plane and its evolution upon Co or Ni doping and subsequent gas adsorption, we performed ELF analysis. For pristine 2H-MoS₂ (Fig. 8a), the ELF map displays high electron localization (red/yellow regions) around the S atoms, reflecting their electron-rich nature and the presence of strong, polarized

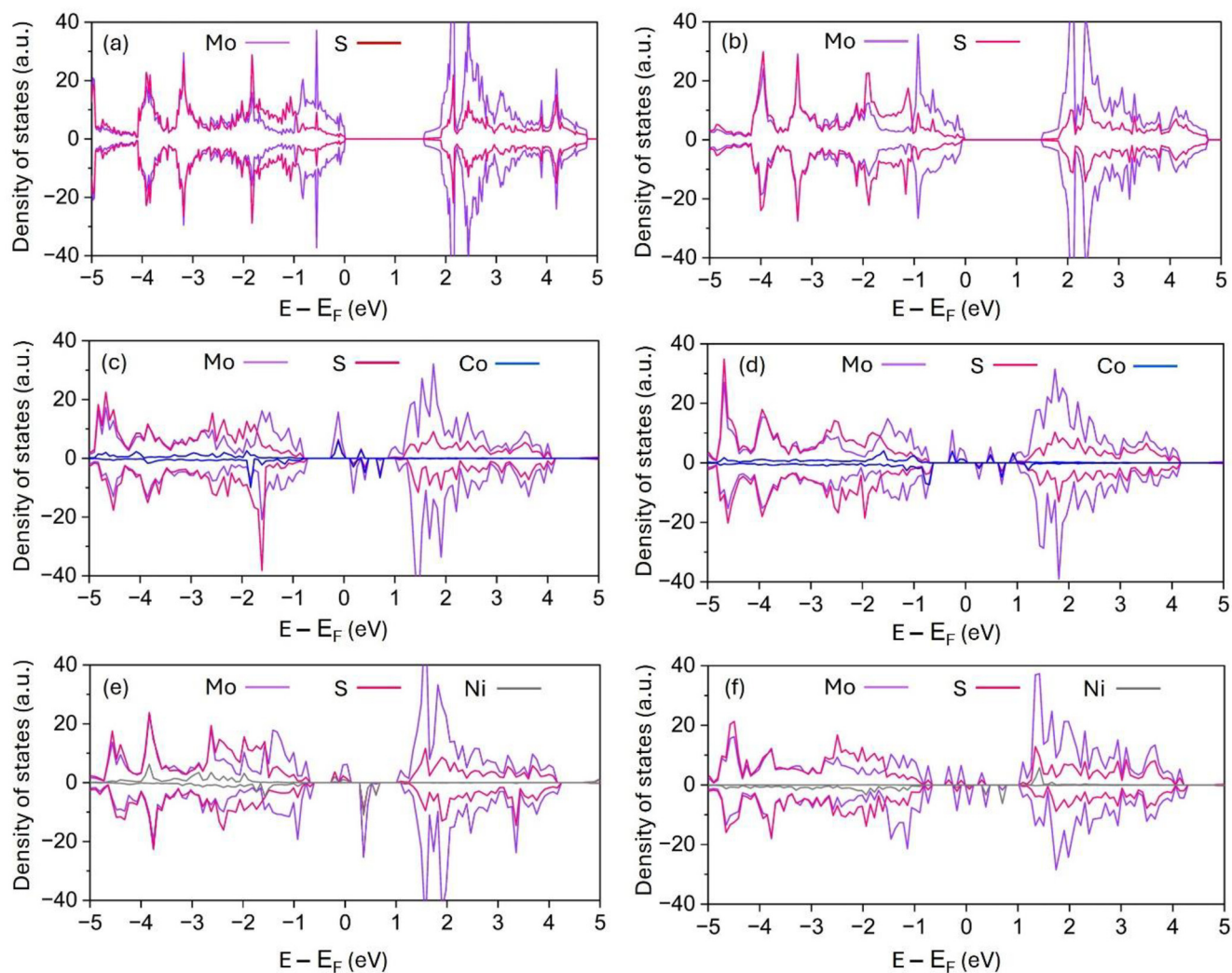


Figure 10. Projected density of states (pDOS) of elements of undoped and doped MoS₂ structures before and after NO₂ adsorptions. (a) pDOS of pristine MoS₂ shows characteristic S-3p valence band and Mo-4d conduction band. (b) NO₂ physisorption on pristine MoS₂ induces a negligible change. (c) Co-doping introduces hybridized Co-3d/Mo-4d/S-3p mid-gap states. (d) NO₂ chemisorption on Co-MoS₂ causes broadening, shifting, and redistribution of these states, indicating charge withdrawal and orbital hybridization. (e) Ni-doping creates prominent spin-polarized mid-gap states. (f) Strong chemisorption on Ni-MoS₂ nearly obliterates these distinct defect peaks, splitting them into a continuum of hybridized states

covalent bonds with the underlying Mo atoms. Introducing a Co or Ni dopant (Fig. 8b-c) fundamentally altered this electronic property. The area surrounding the Co/Ni atom shows a noticeably different ELF profile compared with that of the Mo site in Fig. 8a, with a reduction in high-localization features. This indicates a modulation of the local charge density, which creates a distinct electronic environment. The functional consequences of this electronic pre-activation on NO₂ adsorption are shown in Fig. 8(d-f). For pristine MoS₂ (Fig. 8d), the ELF shows virtually no shared electron density between the surface and the NO₂ molecule, confirming a weak physisorptive interaction. By contrast, the adsorption on Co-MoS₂ (Fig. 8e) reveals a new, shared region of moderate electron localization (green) between the Co atom and the NO₂ molecule. This ELF feature indicates a nascent chemical bond, confirming the transition from physisorption to chemisorption. For the Ni-MoS₂ system (Fig. 8f), a more robust ELF basin forms a definitive "bridge" between the basal plane and the NO₂ molecule, providing direct evidence of strong covalent interactions.

To complement the ELF analysis of chemical bonding, we investigated the electronic band structures to elucidate the impact of Co/Ni doping on the electronic and magnetic properties of MoS₂. The band structure of pristine 2H-MoS₂ (Fig. 9a) exhibits a characteristic semiconductor profile with a clear band gap, consistent with its well-known inertness. The introduction of Co or Ni induces a dramatic electronic transition. Cobalt doping (Fig. 9b) introduces mid-gap states, narrowing the bandgap and rendering the system semi-metallic features. Significant spin polarization emerges, and the electronic states of the spin-up and spin-down channels were asymmetric, generating a local magnetic moment and transforming the originally nonmagnetic MoS₂ sheet into a magnetically active material. This effect was even more pronounced with Ni-doping (Fig. 9c), where extensive spin splitting occurs in the bands around the Fermi level (E_F). The resulting highly dispersive, spin-polarized bands crossing E_F suggest enhanced charge carrier density and mobility, which are critical for promoting charge transfer during surface reactions.

The electronic band structure after NO₂ adsorption is directly correlated with the bonding interactions visualized by ELF. Upon NO₂ physisorption onto pristine MoS₂ (Fig. 9d), the band structure remains nearly unperturbed, retaining its semiconducting character, which confirms the absence of strong chemical interactions expected for physisorption. By contrast, chemisorption of NO₂ onto Co-MoS₂ (Fig. 9e) reorganized the electronic states. The original mid-gap states are broadened and shifted, indicating (i) strong hybridization between the S-3p states and the molecular orbitals of NO₂, (ii) the energy shifts of Co-3d and Mo-4d orbitals after the NO₂ chemisorption on the structure. This hybridization, coupled with the preserved spin polarization, confirms a covalent bonding formation mechanism involving electron transfer and spin selection. The most profound change was observed for the Ni-MoS₂ system (Fig. 9f). Strong chemisorption completely reconfigured the band structure near the Fermi level. The distinct spin-polarized bands of the pristine system are replaced with a new set of hybridized states, consistent with the electronic signatures observed in the ELF analysis (Fig. 8).

Projected density of states (pDOS) analysis was performed to correlate the electronic states and the nature of the adsorbate-substrate interaction (Fig. 10). The pDOS of pristine 2H-MoS₂ (Fig. 10a) shows that its characteristic valence band maximum (VBM) is dominated by S-3p orbitals, and its conduction band minimum (CBM) is dominated by Mo-4d orbitals. The physisorption of NO₂ (Fig. 10b) induces a negligible perturbation of this electronic landscape, with only minimal changes near the S-3p VBM, consistent with a weak, nonbonding interaction. However, the introduction of transition-metal dopants creates profound mid-gap states. For Co-doped MoS₂ (Fig. 10c), these mid-gap states arise from the Co-3d, S-3p, and Mo-4d orbitals, which probably form a covalent "defect complex." The presence of these hybridized states is the orbital-level origin of the semi-metallic character observed in the band structure (Fig. 10b). The chemisorption of NO₂ on the Co-doped sample (Fig. 10d) triggers significant electronic reorganization. The broadening and energy shifts of these defect states, particularly the redistribution of spin-down Mo and S character into a narrower energy range (-0.5 to +1.1 eV), provide direct evidence of orbital hybridization between the adsorbate and the substrate. Specifically, localized electrons in the S atoms interact with the unoccupied molecular orbitals of NO₂, leading to electron density withdrawal from the S-3p states. This charge redistribution results in energetic stabilization and a reduced density of S-3p states. The transfer of electron density toward the NO₂ molecule is also supported by the elongated Co-S bond length after NO₂ adsorption. (Fig. 7e).

This phenomenon was more pronounced in the Ni-doped system. Before adsorption (Fig. 10e), the mid-gap states feature a particularly strong spin-down peak at +0.4 eV (Mo-4d, S-3p, and Ni-3d). The adsorption of NO₂ (Fig. 10f) almost eliminates these distinct defect peaks. The reduced Mo-4d, S-3p, and Ni-3d peak intensities in the mid-gap states and their splitting over a close energy range (-0.5 to +0.8 eV) can be explained using the same hybridization mechanism proposed for the Co-MoS₂ structure. However, the bond strength (i.e., the interaction between the localized electrons of the S atoms and the unoccupied molecular orbitals of NO₂) is stronger than that of the Co-doped structure. This conclusion aligns with our previous

findings from the geometric reconstruction (Fig. 7f) and ELF mapping (Fig. 8f), which demonstrate massive charge transfer and orbital mixing between the S-site and NO₂.

4. CONCLUSIONS

In this work, we combined structural characterization, chemiresistive gas-sensing measurements, and first-principle simulations to investigate pristine and Co- or Ni-doped MoS₂/MoO₃ heterostructures as room-temperature NO₂ sensors. XRD, Raman, and TEM analyses confirmed the formation of layered 2H-MoS₂/MoO₃ nanosheets at 240 °C, with Co and Ni incorporation preserving the main crystal framework while inducing subtle lattice distortions. The pristine MoS₂/MoO₃ sensor exhibited a clear p-type response, good sensing performance over the 1–50 ppm NO₂ range, and pronounced selectivity toward NO₂ over other common gases, with only a modest dependence on operating temperature. DFT calculations further revealed that Co and Ni dopants strongly modified the local bonding, electronic structure, and adsorption behavior, promoting chemisorption and enhancing charge transfer at the doped sites, which helped rationalize the enhanced sensing performance of doped MoS₂/MoO₃. Although repeatable response-recovery cycling was observed within the measurement series, a systematic evaluation of long-term aging stability was not included in the present study. Future work will therefore focus on controlled humidity (%RH) testing to quantify moisture-induced baseline variations, along with multiweek aging studies and post-aging structural characterization to further assess the environmental robustness and long-term stability relevant to practical deployment.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Zahra Ali Alharbi – Materials Science and Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia;  0009-0009-1595-592X
Email: s202274820@kfupm.edu.sa

Authors

Retno Dwi Astuti – Materials Science and Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia;  0009-0004-9898-8260
Email: g202413960@kfupm.edu.sa

Md Saif Hasan – Department of Chemical and Biological Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, 999077 P.R. China;  0000-0002-1702-8814
Email: mdsaiifhasan@connect.ust.hk

Corresponding Author

Md Delowar Hossain – Research Mentor, Assistant Professor Materials Science and Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia; & Interdisciplinary Research Center for Hydrogen Technologies and Carbon Management, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia;  0000-0003-3440-8306
Email: mddelowar.hossain@kfupm.edu.sa

■ DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT to improve the language. After using this tool/service, the authors reviewed and edited the content as required and took full responsibility for the publication.

■ CONFLICT OF INTEREST

The authors declare no conflict of interest.

■ ACKNOWLEDGEMENTS

The authors acknowledge the support provided by the Department of Materials Science and Engineering and the Interdisciplinary Research Center (IRC) Hydrogen Technology and Carbon Management (HTCM) at King Fahd University of Petroleum and Minerals (KFUPM). This research was funded by Project No. INHT2509. The authors also extend their appreciation to the technical staff and laboratory team of the Core Research Facilities (CRF) KFUPM for their assistance in specimen preparation and testing, particularly for XRD, SEM, and TEM characterization.

■ REFERENCES

- (1) Chen, X. *et al.* A Novel Artificial Neuron-Like Gas Sensor Constructed from CuS Quantum Dots/Bi₂S₃ Nanosheets. *Nano-Micro Letters* 2021 14:1 **14**, 8- (2021).
- (2) Liu, W. *et al.* Prenatal and postnatal exposures to ambient air pollutants associated with allergies and airway diseases in childhood: A retrospective observational study. *Environ. Int.* **142**, 105853 (2020).
- (3) Bizak, Z., Faleiros, M. C., Vijjapu, M. T., Yaqoob, U. & Salama, K. N. Highly Sensitive Wireless NO₂ Gas Sensing System. *IEEE Sens. J.* **23**, 15667–15674 (2023).
- (4) Tamvakos, A. *et al.* NO₂ Gas Sensing Mechanism of ZnO Thin-Film Transducers: Physical Experiment and Theoretical Correlation Study. *ACS Sens.* **1**, 406–412 (2016).
- (5) Schmitz, S., Caseiro, A. & von Schneidmesser, E. How electrochemical sensors measure up to reference-grade nitrogen dioxide monitors across temporal scales. *Science of The Total Environment* **980**, 179476 (2025).
- (6) Mawrence, R., Munniks, S. & Valente, J. Calibration of Electrochemical Sensors for Nitrogen Dioxide Gas Detection Using Unmanned Aerial Vehicles. *Sensors (Basel)*. **20**, 7332 (2020).
- (7) Zhao, Y. *et al.* Development of High-Precision NO₂ Gas Sensor Based on Non-Dispersive Infrared Technology. *Sensors* 2024, Vol. 24, Page 4146 **24**, 4146 (2024).
- (8) Wijesinghe, D. R., Zobair, M. A. & Esmaelpour, M. A Review on Photoacoustic Spectroscopy Techniques for Gas Sensing. *Sensors* 2024, Vol. 24, Page 6577 **24**, 6577 (2024).
- (9) Lagutin, A. S. & Vasil'ev, A. A. Solid-State Gas Sensors. *Journal of Analytical Chemistry* 2022 77:2 **77**, 131–144 (2022).
- (10) Xiong, S. *et al.* High Performance Acoustic Wave Nitrogen Dioxide Sensor with Ultraviolet Activated 3D Porous Architecture of Ag-Decorated Reduced Graphene Oxide and Polypyrrole Aerogel. *ACS Appl. Mater. Interfaces* **13**, 42094–42103 (2021).
- (11) Yuvaraja, S. *et al.* Fully Integrated Organic Field-Effect Transistor Platform to Detect and to Quantify NO₂ Gas. *Physica Status Solidi - Rapid Research Letters* **14**, 2000086 (2020).
- (12) Sacco, L., Forel, S., Florea, I. & Cojocaru, C. S. Ultra-sensitive NO₂ gas sensors based on single-wall carbon nanotube field effect transistors: Monitoring from ppm to ppb level. *Carbon N. Y.* **157**, 631–639 (2020).
- (13) Shin, S. Y. *et al.* Improved NO₂ Gas-Sensing Performance of an Organic Field-Effect Transistor Based on Reduced Graphene Oxide-Incorporated Nanoporous Conjugated Polymer Thin Films. *Chemistry of Materials* **35**, 7460–7474 (2023).
- (14) Patil, S. V. *et al.* Chemiresistive NO₂ gas sensor based on RuO₂/MoS₂ composite. *J. Alloys Compd.* **1036**, 181748 (2025).
- (15) Hong, H. S. *et al.* Ultra-low detection limit chemoresistive NO₂ gas sensor using single transferred MoS₂ flake: an advanced nanofabrication. *RSC Adv.* **12**, 33403 (2022).
- (16) Li, Q., Zeng, W. & Li, Y. Metal oxide gas sensors for detecting NO₂ in industrial exhaust gas: Recent developments. *Sens. Actuators B Chem.* **359**, 131579 (2022).
- (17) Kumar, R., Mamta, Kumari, R. & Singh, V. N. SnO₂-Based NO₂ Gas Sensor with Outstanding Sensing Performance at Room Temperature. *Micromachines* 2023, Vol. 14, Page 728 **14**, 728 (2023).
- (18) Berwal, P. *et al.* Fabrication of durable and sensitive SnO₂-Based gas sensor for NO₂ detection. *Applied Physics A* 2025 131:9 **131**, 700- (2025).
- (19) Masuda, Y. Recent advances in SnO₂ nanostructure based gas sensors. *Sens. Actuators B Chem.* **364**, 131876 (2022).
- (20) Khort, A. *et al.* High-performance selective NO₂ gas sensor based on In₂O₃-graphene-Cu nanocomposites. *Scientific Reports* 2023 13:1 **13**, 7834- (2023).
- (21) Raju, P., Rambelosen, J., Ioannou, D. E., Motayed, A. & Li, Q. Noble Metal-Decorated In₂O₃ for NO₂ Gas Sensor: An Experimental and DFT Study. *Chemosensors* 2025, Vol. 13, Page 350 **13**, 350 (2025).
- (22) Mei, H., Zhang, F., Zhou, T. & Zhang, T. Pulse-Driven MEMS NO₂ Sensors Based on Hierarchical In₂O₃ Nanostructures for Sensitive and Ultra-Low Power Detection. *Sensors (Basel)*. **24**, 7188 (2024).
- (23) Ibrahim, K. M. & Saleh, W. R. ZnO Nanostructures as Low Concentration NO₂ Gas Sensor and Impact the Temperature on Sensing Properties. *AIP Conf. Proc.* **2922**, 240001 (2024).
- (24) Ren, X. *et al.* Conductometric NO₂ gas sensors based on MOF-derived porous ZnO nanoparticles. *Sens. Actuators B Chem.* **357**, 131384 (2022).
- (25) Han, T. H. *et al.* Decoration of CuO NWs Gas Sensor with ZnO NPs for Improving NO₂ Sensing Characteristics. *Sensors (Basel)*. **21**, 2103 (2021).
- (26) Kumar, S. *et al.* Enhancing NO₂ Gas Sensing: The Dual Impact of UV and Thermal Activation on Vertically Aligned Nb-MoS₂ for Superior Response and Selectivity. *ACS Sens.* **10**, 2191–2202 (2025).
- (27) Kumar, S., Meng, G., Mishra, P., Tripathi, N. & Bannov, A. G. A systematic review on 2D MoS₂ for nitrogen dioxide (NO₂) sensing at room temperature. *Materials Today Communications* vol. 34 Preprint at <https://doi.org/10.1016/j.mtcomm.2022.105045> (2023).
- (28) Tian, X. *et al.* Recent advances in MoS₂-based nanomaterial sensors for room-temperature gas detection: A review. *Sensors and Diagnostics* **2**, 361–381 (2023).
- (29) Chithambararaj, A. & Bose, A. C. Hydrothermal synthesis of hexagonal and orthorhombic MoO₃ nanoparticles. *J. Alloys Compd.* **509**, 8105–8110 (2011).
- (30) Pradeep Kumar, P. & Singh, V. Enhanced dual gas sensing performance of MoS₂/MoO₃ nanostructures for NH₃ and NO₂ detection. *Ceram. Int.* **50**, 21978–21988 (2024).
- (31) Yan, J., Wang, Y., Yang, C., Deng, H. & Hu, N. MoS₂/MoO₃ heterojunctions enabled by surface oxidization of MoS₂ nanosheets for high-performance room-temperature NO₂ gas sensing. *J. Alloys Compd.* **976**, (2024).

- (32) Li, W. *et al.* Highly Sensitive NO₂ Gas Sensors Based on MoS₂@MoO₃ Magnetic Heterostructure. *Nanomaterials* **12**, (2022).
- (33) Pradeep Kumar, P. & Singh, V. Enhanced dual gas sensing performance of MoS₂/MoO₃ nanostructures for NH₃ and NO₂ detection. *Ceram. Int.* **50**, 21978–21988 (2024).
- (34) Ji, Y. *et al.* Surface-functionalized MoS₂/MoO₃ heterostructure-based NO₂ sensors self-powered by triboelectric nanogenerators. *Sens. Actuators B Chem.* **446**, (2026).
- (35) Mahla, S. K., Kesarwani, S. & Pal, A. From 1T to 2H: Engineering Mo S₂ phases through tunable hydrothermal routes. *Journal of Alloys and Compounds Communications* **8**, 100109 (2025).
- (36) Kushwaha, A. *et al.* Enhanced NO₂ Gas Sensing in Nanocrystalline MoS₂ via Swift Heavy Ion Irradiation: An Experimental and DFT Study. *ACS Sens.* **9**, 5966–5975 (2024).
- (37) Gerbi, Z. D. Electronic and optical properties of molybdenum disulfide (MoS₂) mono layer using density functional theory (DFT) calculations. *AIP Adv.* **15**, 25313 (2025).
- (38) Kozdra, S., Wójcik, A., Das, T. & Michałowski, P. P. From DFT investigations of oxygen-implanted molybdenum disulfide to temperature-induced stabilization of MoS₂/MoO₃ heterostructure. *Appl. Surf. Sci.* **631**, 157547 (2023).
- (39) Kresse, G. & Hafner, J. *Ab initio* molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251 (1994).
- (40) Kresse, G. & Furthmüller, J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169 (1996).
- (41) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953 (1994).
- (42) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **27**, 1787–1799 (2006).
- (43) Wang, V., Xu, N., Liu, J. C., Tang, G. & Geng, W. T. VASP-KIT: A user-friendly interface facilitating high-throughput computing and analysis using VASP code. *Comput. Phys. Commun.* **267**, 108033 (2021).
- (44) Panasci, S. E. *et al.* Multiscale Investigation of the Structural, Electrical and Photoluminescence Properties of MoS₂ Obtained by MoO₃ Sulfurization. *Nanomaterials* **12**, (2022).
- (45) De Melo, O. *et al.* Optical and electrical properties of MoO₂ and MoO₃ thin films prepared from the chemically driven isothermal close space vapor transport technique. *Journal of Physics Condensed Matter* **31**, (2019).
- (46) Faizan, M. *et al.* MoO₃ @MoS₂ Core-Shell Structured Hybrid Anode Materials for Lithium-Ion Batteries. *Nanomaterials* **12**, (2022).
- (47) Yan, H., Song, P., Zhang, S., Yang, Z. & Wang, Q. Facile fabrication and enhanced gas sensing properties of hierarchical MoO₃ nanostructures. <https://doi.org/10.1039/c5ra13036k> (2015).

Design and Characterization of Green Cementitious Composites for Structural Health Monitoring

Mohammed Al Abdulmohsin and Asad Hanif*

Cite <https://doi.org/10.64589/juri/219224>

Submitted: November 30, 2025 Revised: January 19, 2026 Accepted: January 29, 2026

ABSTRACT

This study experimentally characterized various green cementitious composites with inherent self-sensing capabilities for use in structural health monitoring. Four composite mortar mixes were prepared using different combinations of ordinary Portland cement, reactive magnesia, slag, and graphene oxide (GO), then cured in water (for cement mixes) or air (for magnesia–slag mixes). Compressive strengths were determined by testing 50 mm cubic specimens cured for 28 d, and electrical properties were determined over a broad alternating current frequency range (10 Hz to 100 kHz) using four probes embedded in prismatic specimens cured for 1–28 d or subjected to monotonic and cyclic loading after 28 d. The cement-based reference specimen exhibited the highest compressive strength (~56.9 MPa) at 28 d; the strength of the GO-modified cement specimen (CGS) was only slightly lower (52.3 MPa), indicating that the addition of GO had little effect on mechanical performance. By contrast, the magnesia–slag specimens exhibited lower strengths of 10.2 MPa (without GO) and 9.0 MPa (with GO). The four-probe measurements revealed that the incorporation of GO significantly reduced electrical resistance: the CGS specimen exhibited the lowest resistivity across all frequencies, approximately 30% and 250% lower than the resistivities of the cement and magnesia–slag specimens without GO, respectively. Notably, the CGS specimen also exhibited stable and repeatable piezoresistive responses under both monotonic and cyclic loading. These results confirm that GO-modified cementitious composites can simultaneously provide adequate structural capacity and reliable self-sensing functionality, thereby offering a promising solution for real-time structural health monitoring.

Keywords: self-sensing composites, graphene oxide, magnesia–slag mortar, electrical resistivity, sustainable materials, structural health monitoring

1. INTRODUCTION

1.1. Study Background. Concrete is among the most critical and widely used construction materials worldwide and is second only to water in terms of global consumption^{1,2}. Its value lies in its ease of production, relatively low manufacturing cost, and versatility in applications, which range from foundations and columns to bridges, roads, and dams³. The ability to cast concrete in various shapes and sizes makes it particularly valuable for diverse construction applications. However, despite these advantages, concrete structures are subject to numerous issues, including corrosion of embedded steel reinforcement, deterioration due to environmental exposure, cracking under service loads, and degradation from chemical attack⁴. These issues necessitate continuous research to develop advanced concrete systems that ensure enhanced durability, extended service life, and improved performance.

Methods for monitoring the health of concrete structures have attracted significant interest from researchers and designers working in civil engineering^{5,6}. Structural health monitoring (SHM) comprises the assessment of structure condition using sensor data collected over coordinated time intervals to

predict potential damage in advance, facilitate timely maintenance, and limit repair costs⁷. Indeed, the early detection of structural distress can significantly extend service life and prevent catastrophic failures. Conventional SHM systems have been successfully applied in major infrastructure projects, including high-rise towers, long-span bridges, and large dams, where safety requirements are particularly stringent^{8,9}. However, conventional SHM approaches rely on external sensors, such as strain gauges, accelerometers, and fiber optic sensors, which require costly and complex installation procedures as well as regular maintenance^{10,11}. These limitations have motivated researchers to explore alternative solutions for improving the monitoring efficiency, accessibility, and cost-effectiveness of this technology, particularly in small-scale projects and routine infrastructure applications.

Despite considerable advances in SHM technologies, most recent studies have concentrated on large-scale projects and employed advanced fillers and measurement systems, making these solutions costly and impractical for smaller projects or simple laboratory conditions. Therefore, this study characterized the electrical and mechanical performances of various cementitious mixes incorporating graphene oxide, magnesia, and slag

to design a composite with intrinsic self-sensing for practical, low-cost SHM.

1.2. Self-sensing Cementitious Composites. Conventional concrete structures require external sensors to monitor their condition and detect problems or damage. By contrast, self-sensing cementitious composites combine structural and sensing functions within the same matrix^{12,13} to monitor stress, strain, and cracking through changes in electrical resistivity. This principle relies on the piezoresistive effect, in which the electrical resistance changes in response to mechanical load^{14–16} following Ohm's law, which is given by

$$R = \frac{V}{I} \quad (1)$$

where R denotes the electrical resistance (Ω), V denotes the voltage drop across the specimen (V), and I denotes the applied current (A).

The resistivity (ρ) of a material under applied stress is calculated by

$$\rho = R \times \frac{A}{L} \quad (2)$$

where A denotes the electrode surface area and L denotes the distance between electrodes.

Finally, the sensitivity of material resistivity/resistance to stress can be expressed as the fractional change in resistivity/resistance (FCR) as follows:

$$FCR = \frac{\rho - \rho_0}{\rho_0} \times 100 \quad (3)$$

where ρ_0 denotes the initial resistivity (or resistance) at zero or reference stress ($\Omega \cdot m$), and ρ denotes the resistivity (or resistance) measured under strain.

This piezoresistive mechanism enables an appropriately designed cementitious composite to function not only as a load-bearing material but also as an integrated sensor. Such self-sensing functionality has been achieved in previous studies by incorporating conductive fillers into cementitious matrices. The use of carbon fibers for this purpose has been investigated extensively: Han et al.¹⁷ reported that carbon fiber-reinforced concrete exhibits excellent piezoresistivity at gauge factors of up to 500. The use of carbon nanotubes (CNTs) has also shown considerable promise, as reported by D'Alessandro et al.¹⁸ and Konsta-Gdoutos et al.¹⁹, although challenges related to dispersion uniformity remain to be addressed. Graphene and graphene oxide have emerged as particularly attractive materials for enabling self-sensing because of their exceptional electrical conductivities, large surface areas, and cost-effectiveness compared with CNTs^{20–22}. The use of metallic powders such as nickel for self-sensing applications has also been explored^{23,24}.

The incorporation of such fillers creates conductive networks that alter electron flow in response to mechanical deformation^{12–15,17,18}, allowing the composite to serve as both a structural element and an integrated sensor. This study focuses on green cementitious composites reinforced with graphene oxide and magnesia-slag binders, as they represent the most sustainable and cost-effective materials suitable for SHM.

1.3. Use of Graphene Oxide, Magnesia, and Slag in Cementitious Materials. Graphene oxide (GO) possesses many unique physical and chemical properties and has shown considerable promise when used with cementitious materials^{25–27}. Among its most valuable features are its large surface area (typically 500–1000 m²/g) and abundant surface functional groups (e.g., hydroxyl, epoxide, and carboxyl), which allow it to disperse effectively within aqueous solutions and interact with cement hydration products²⁸. Another key characteristic of GO is its ability to form conductive pathways within the cement matrix, which is essential for enabling internal connectivity and sensing capability²⁹. Furthermore, studies by Lv et al.³⁰ and Pan et al.³¹ demonstrated that GO can enhance the mechanical strength of cementitious materials at optimal dosages (0.01–0.05% by weight of binder), improve microstructural density, and extend material service life through its nanoscale reinforcing effect. Finally, GO is more cost-effective and easier to disperse than alternatives such as CNTs, further increasing its practical value³².

Magnesia (MgO) has received increasing attention in recent years from researchers investigating green cement and sustainable construction^{33–35}. When used as a replacement for ordinary Portland cement, MgO plays a significant role in reducing the carbon emissions associated with cement production, which account for approximately 8% of global CO₂ emissions³⁶. Upon hydration, reactive MgO forms magnesium hydroxide (Mg(OH)₂), which can contribute to strength development and exhibits the potential for carbonation hardening^{37,38}. Research conducted by Mo et al.³⁹ and Unluer and Al-Tabbaa⁴⁰ has shown that MgO-based binders can achieve comparable performance to Portland cement under appropriate curing conditions while offering environmental benefits. Thus, MgO represents a valuable material for sustainable construction.

Ground granulated blast furnace slag (GGBFS) is a by-product of the steel industry that provides environmental advantages when used in cementitious composites^{41,42}. Indeed, its utilization as a supplementary binder in cement-based materials not only entrains industrial waste but also improves durability and long-term performance owing to its pozzolanic and latent hydraulic properties⁴³. When blended with MgO, GGBFS enhances hydration through the formation of calcium-silicate-hydrate and magnesium-aluminum-silicate-hydrate phases, improving mechanical properties while maintaining sustainability benefits^{44,45}.

This study aims to evaluate whether a new type of green cementitious composite developed by combining graphene oxide with magnesia-slag binders can achieve both durability and self-sensing functionality. While GO imparts sensing capability through enhanced electrical conductivity, magnesia and slag improve sustainability and reduce carbon footprint. Together, these materials provide a cost-effective alternative to conventional cementitious composites and represent a suitable approach for SHM.

2. METHODOLOGY

2.1. Raw Materials. In this study, different materials were used to prepare various mortar mixes, and their performance was



Figure 1. (a) Mixing and preparation of mortar specimens; (b) ultrasonication used to disperse GO in water before mixing

evaluated under the same conditions to determine the optimal mix for SHM. These mixes comprised cement and sand (CS); cement, GO, and sand (CGS); MgO, GGBFS, and sand (MS); and MgO, GGBFS, GO, and sand (MGS) in the proportions provided in Table 1.

2.2. Specimen Preparation. The specimen preparation method varied according to the materials used in each mix. First, the required quantities of all materials were measured accurately. Next, the dry components (sand with cement or MgO and GGBFS) were dry-mixed in a mechanical mixer until a uniform blend was obtained, then water (or a GO–water dispersion) was added gradually. Mixtures containing GO were prepared by initially dispersing the GO in water using a probe-type ultrasonicator for 2 min at an amplitude of 60% to ensure adequate exfoliation and enhance the dispersion of GO within the cementitious composite; this process also reduced nanoparticle agglomeration and minimized scattering during subsequent mixing. The dry and wet components were mixed until a homogeneous consistency was achieved; the average total mixing time was approximately 6 min.

Six specimens were prepared from each mix. Three 50 mm cubic specimens of each mix were cast for compressive strength testing. Three 40 mm × 40 mm × 160 mm prismatic specimens of each mix were cast for electrical property measurement using four stainless steel mesh electrodes inserted into the fresh mortar at a 30 mm spacing to ensure consistent electrical contact with the cementitious matrix through proper bonding and thereby prevent contact degradation during subsequent curing.

Figure 1 illustrates the specimen mixing and preparation processes, including the ultrasonic dispersion of GO. Figure 2 shows and the casting procedure used to produce both the cubic and prismatic specimens, including the embedded electrodes.

Table 1. Mix proportions for the mortar specimens evaluated in this study

Mix	Binder (g)				Sand (g)	Water (g)
	Cement	MgO	GGBFS	GO (g)		
CS	1250	0	0	0	1250	625
CGS	1250	0	0	7.81	1250	625
MS	0	500	500	0	100	900
MGS	0	500	500	7	100	900

3. EXPERIMENTAL WORK

The prismatic specimen weight, voltage, current, and resistance were measured after curing for 1, 7, 14, 21, and 28 d in water (for cement-based specimens) or air (for MgO-based specimens) to monitor their changes over time. The cubic specimen compressive strength was determined after 28 d of curing in water or air to determine the effects of the considered materials on the composite strength. Finally, the prismatic specimens were subjected to monotonic and cyclic loading after 28 d of curing to evaluate their piezoresistive responses.

3.1. Four-probe Electrical Property Measurement. A four-probe measurement method was used to assess the electrical properties of the prismatic specimens at each curing age. This technique has been extensively applied to cementitious and composite materials because of its ability to limit contact resistance and precisely determine the intrinsic electrical resistivity of a specimen. In this experimental setup, shown in Figure 3, the two outer embedded electrodes were connected to a constant alternating current (AC) power supply, and the two inner embedded electrodes were connected to a digital ammeter and voltmeter to record the current and voltage drop across the embedded electrodes simultaneously. This separation between the current injection and voltage measurement points can effectively mitigate the influence of the electrode–specimen contact resistance on the measured voltage.

To gain insights into how the conductivity and sensing performance of the specimens vary across low- and high-frequency domains, the electrical properties of the composite specimens were examined under different conditions by conducting measurements at AC frequencies of 10 Hz, 100 Hz, 500 Hz, 1 kHz, 30 kHz, 100 kHz, 500 kHz, and 1 MHz. The collected data were used to determine the changes in resistance and resistivity



Figure 2. Casting of mortar specimens showing cubic and prismatic molds with metallic mesh electrodes embedded to enable electrical measurements

of each specimen and thereby evaluate the self-sensing capabilities of the composite specimens. These changes were quantified using the FCR, calculated using resistivity in Equation (3), which is a critical parameter for evaluating the self-sensing behavior of cementitious composites. Note that the electrodes remained embedded throughout the curing and testing periods to facilitate repeated and consistent electrical measurements at different ages without disturbing the specimen. This configuration permitted the reliable monitoring of the electrical property evolution according to hydration progress, curing conditions, and material composition.

3.2. Compressive Strength Testing. The compressive strength of each cubic mortar specimen was determined in accordance with ASTM C109/C109M. The fresh mortar was cast into steel cubic molds in two layers, each of which was compacted using 32 tamps to ensure proper consolidation before the surface was leveled. Next, the specimens were covered and stored at $23 \pm 2^\circ\text{C}$ for 24 ± 2 h, then demolded and cured in water (or air) at $23 \pm 2^\circ\text{C}$ until testing after 28 d of curing. Before each test, the specimen was removed from its curing location, and its surface moisture was wiped off. Next, the specimen was centrally

aligned between the platens of a calibrated compression-testing machine. The compressive load was applied continuously and without shock at a load-controlled rate of 0.9 ± 0.2 MPa/s until failure. The compressive strength was calculated as the ratio of the maximum load to the loaded area; the values reported in this paper represent the average result obtained by testing at least three specimens for each mix.

3.3. Piezoresistive Response Testing. The piezoresistive responses of the prismatic specimens when subjected to cyclic and monotonic compression were evaluated after 28 d of curing using the four-probe electrical measurement configuration (detailed in Section 2.3.1) coupled with a uniaxial compression-testing machine (described in Section 2.3.2). Before each specimen was tested, electrical contacts were attached to the embedded electrodes, and the baseline electrical resistance was recorded. During the monotonic compression test, the load was continuously applied at a constant displacement-controlled rate until the target strain or failure was reached. During the cyclic compression test, each specimen was subjected to repeated loading-unloading cycles between predefined stress (or strain) limits at a constant loading rate as the mechanical strain and

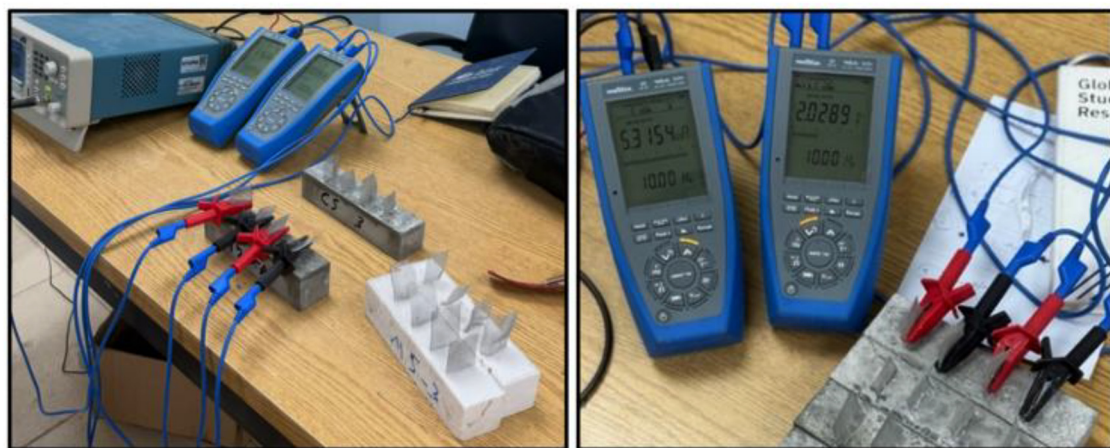


Figure 3. Electrical testing setup using the four-probe method to measure current and voltage across the mortar specimens at multiple frequencies for resistivity analysis

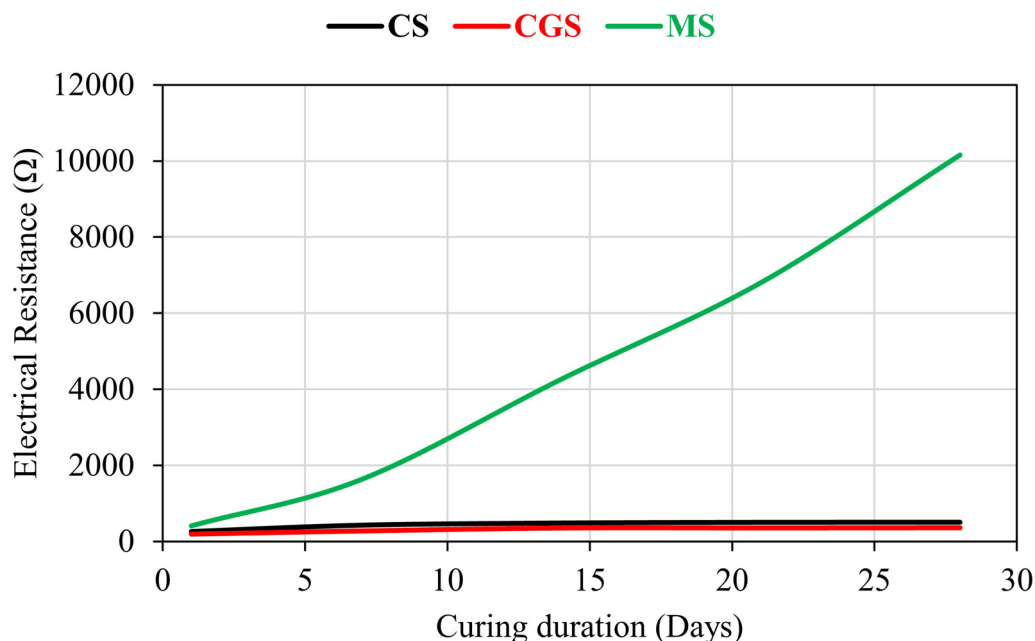


Figure 4. Electrical resistance of the considered specimens according to curing age when tested at 1 kHz AC

electrical resistance were continuously recorded. The piezoresistive performance of each specimen was assessed by correlating its FCR, calculated using resistance in Equation (3), with the applied strain to evaluate its sensitivity, repeatability, and signal stability under cyclic and monotonic loading conditions.

4. RESULTS AND DISCUSSION

4.1. Electrical Properties of Mortar Specimens. Figure 4 presents the variation in electrical resistance over time, and

Figure 5 presents the electrical resistance according to AC frequency for the CS, CGS, and MS specimens. A consistent trend can be observed across all specimens in which resistance decreases as the frequency increases; this behavior is expected in cementitious composites, as capacitive reactance decreases at higher frequencies, resulting in a lower overall impedance and, consequently, a reduction in measured resistance. Additionally, the observed decrease in resistance with curing age reflects ongoing hydration and microstructural refinement, which improve internal connectivity and reduce ionic and pore resistance.

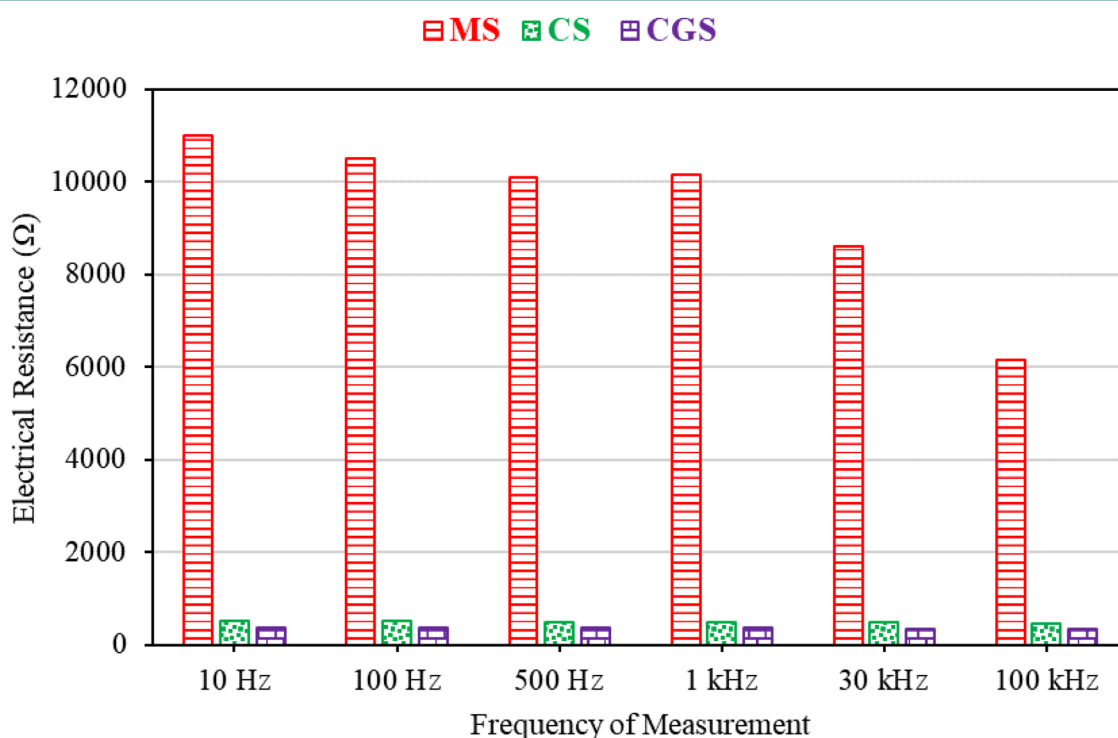


Figure 5. Electrical resistance of CS, CGS, and MS specimens after 28 d of curing according to the applied AC frequency

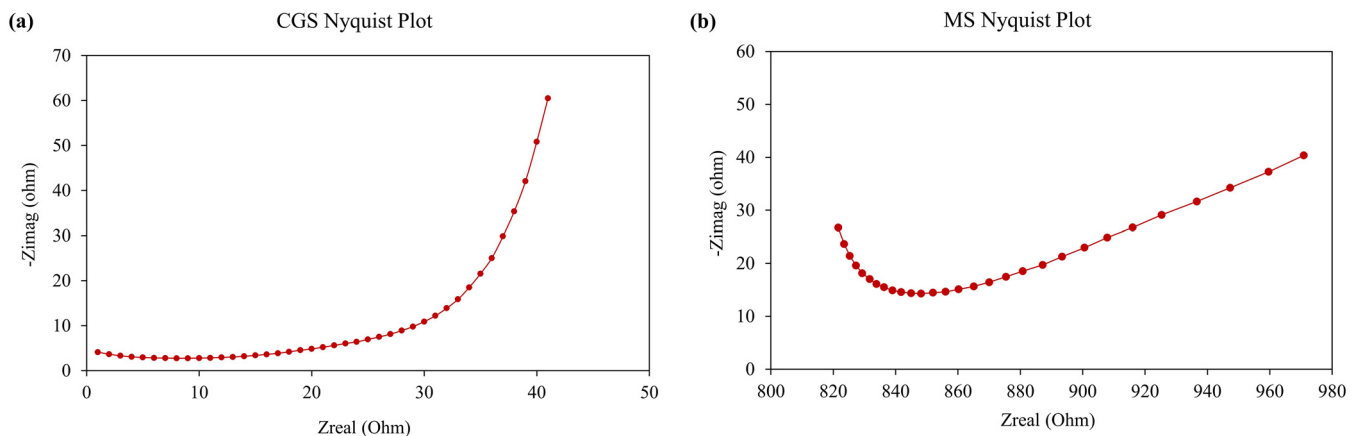


Figure 6. Nyquist plots ($-Z_{imag}$ vs Z_{real}) for the CGS and MS specimens after 28 d of curing

Among the evaluated specimens, CGS exhibited the lowest overall resistance, confirming the positive contribution of GO to the formation of conductive pathways within the cement matrix. Thus, the inclusion of GO can enhance electron mobility and improve sensing performance. By contrast, MS and MGS exhibited higher resistance values owing to the applied air-curing regime and the presence of MgO and GGBFS, both of which limit ionic conduction. Nevertheless, the gradual reduction in the resistance of these specimens over time suggests the continued formation of hydration and carbonation products that enhance the electrical response throughout the curing period.

Figure 6 shows the impedance responses of selected mortar specimens (CGS and MS) after 28 d of curing. In the Nyquist representation, the real impedance component (Z_{real}) reflects the bulk resistance of the specimen, whereas the imaginary component ($-Z_{imag}$) represents the capacitive behavior associated with charge storage and polarization within the specimen pore structure.

Both Nyquist plots clearly exhibit typical semicircular patterns that are characteristic of cementitious materials tested under AC. The semicircle in the Nyquist plot for the CGS specimen is smaller than that for the MS specimen, suggesting that the addition of GO reduces overall impedance and improves electrical conductivity by forming continuous conductive pathways that enhance charge transfer through the cement matrix. By contrast, the larger semicircle exhibited by the MS specimen confirms its higher internal resistance and weaker conductive network. Notably, the application of air curing and the use of MgO and GGBFS limited ionic transport in this specimen compared with that in the water-cured CGS specimen.

4.2. Compressive Strength Results. The compressive strength results provide insights into the effects of different binder systems and curing conditions on the overall mechanical performance of the mortar specimens. As shown in Figure 7, a distinct contrast between the compressive strengths of the

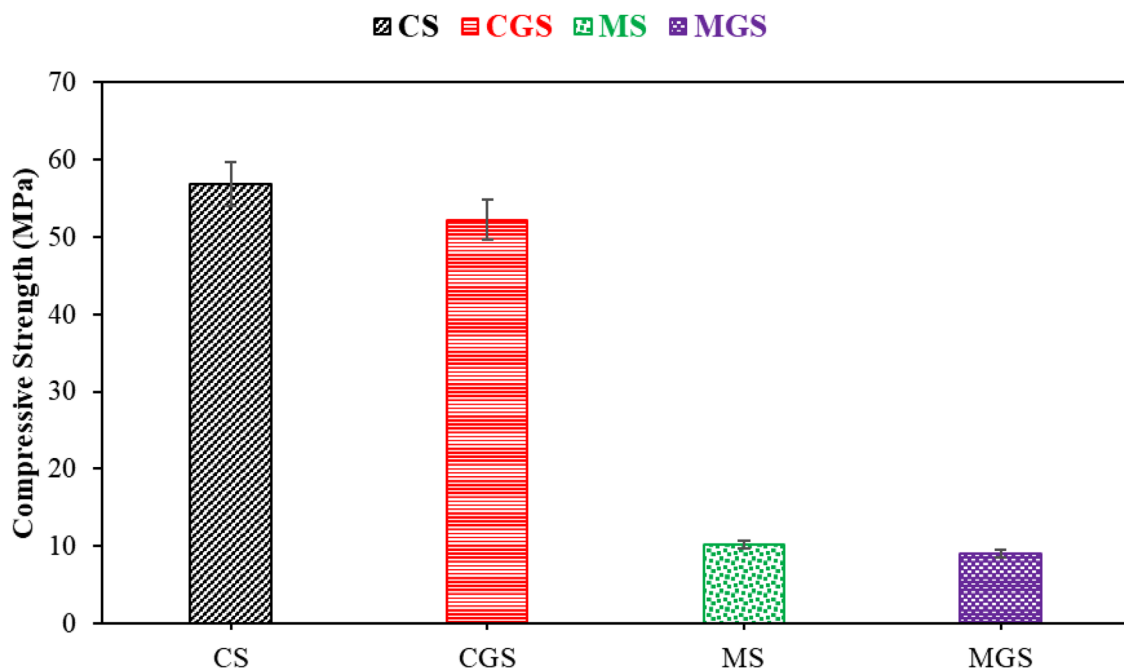


Figure 7. Compressive strengths of the evaluated specimens after 28 d of curing

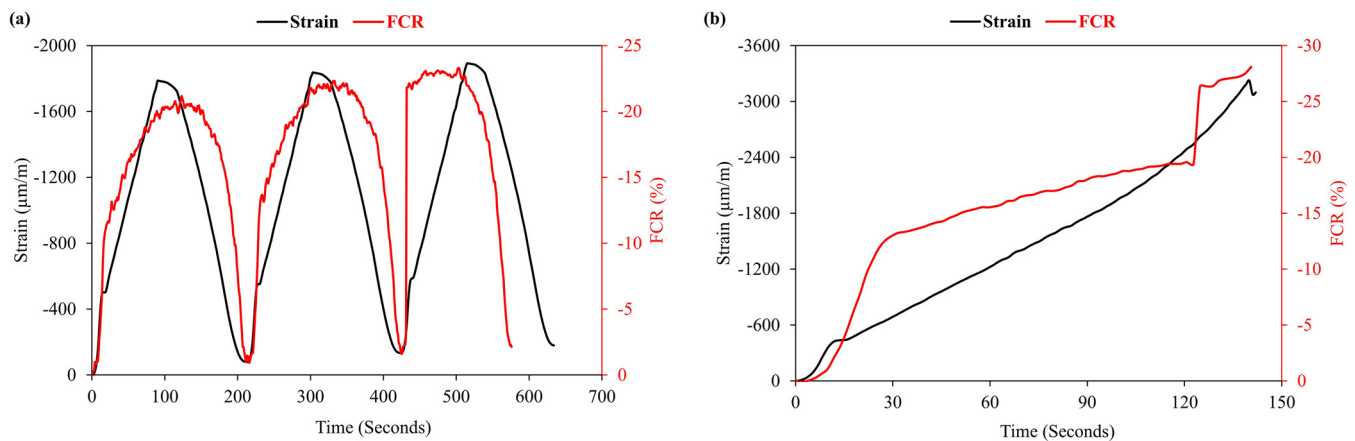


Figure 8. Strain-FCR responses of the 28 d CGS specimen under (a) monotonic and (b) cyclic loading

cement-based (CS and CGS) and MgO-GGBFS-based (MS and MGS) specimens was observed.

The CS specimen exhibited the highest strength (averaging 56.9 MPa), reflecting the strong bonding and dense structure resulting from the hydration of Portland cement during water curing. The strength of the CGS mix, which incorporated GO, was slightly lower (averaging 52.3 MPa); this reduction is common in graphene-modified cement systems because GO initially interferes with hydration before it enhances long-term performance through superior stress distribution and microstructural refinement.

The MS and MGS specimens exhibited considerably lower strengths averaging 10.2 MPa and 9.0 MPa, respectively. These results align with the expected behavior of air-cured MgO-GGBFS binders, in which limited hydration and the absence of continuous moisture lead to weaker mechanical bonds. The inclusion of GO in the MGS specimen did not noticeably improve its strength, suggesting that the conductivity benefits of GO do not translate into structural reinforcement when the mortar is dry-cured.

4.3. Piezoresistive Response Under Mechanical Loading. Piezoresistive response testing was conducted under monotonic and cyclic loading to evaluate the ability of GO-modified mortar to sense mechanical deformation in real time based on the FCR.

4.3.1. Monotonic loading response. Figure 8(a) shows the strain-FCR response of the CGS specimen under monotonic loading. The FCR increased continuously with the applied compressive strain, indicating a strong correlation between mechanical deformation and electrical resistance. Indeed, the observed increase in FCR reflects the piezoresistive mechanism of the composite, in which applied stress alters the conductive pathways formed by the GO within the cement matrix. The consistent trend and absence of sudden fluctuations confirm stable sensing performance during continuous compressive loading.

4.3.2. Cyclic loading response. Figure 8(b) shows the response of the same composite during three loading-unloading cycles. Note that the FCR curve closely follows the strain pattern in each cycle, with both strain and FCR increasing during loading and decreasing during unloading in a nearly repeatable manner.

This repeatability across cycles confirms that the CGS mortar exhibited reversible piezoresistive behavior, indicating that its sensing response returned to the baseline value after each unloading stage. This characteristic is essential for real-time SHM applications in components subjected to repeated load cycles, such as bridge members and pavements.

5. CONCLUSIONS

This study developed and evaluated various green cementitious composites incorporating ordinary Portland cement, reactive MgO, GGBFS, and GO for use in SHM applications. The cement-based reference specimen exhibited the highest 28-d compressive strength (56.9 MPa), but the incorporation of GO into this mix resulted in only a modest strength reduction (52.3 MPa), confirming that self-sensing functionality can be introduced without significantly compromising mechanical performance. By contrast, the air-cured MgO-GGBFS specimens exhibited substantially lower strengths owing to limited hydration and higher porosity. The addition of GO markedly enhanced the electrical conductivity and impedance characteristics of the specimens, reducing their resistance, shrinking their Nyquist semicircles, and improving charge transfer within, with resistance decreasing as curing progressed and the applied AC frequency increased. Notably, the GO-modified cement specimen exhibited stable and repeatable piezoresistive behavior under both monotonic and cyclic loading, producing a strong strain-FCR correlation with minimal hysteresis; by contrast, the plain cement specimen exhibited negligible sensing capability. Water curing promoted continuous hydration and strength gain in the cement systems, whereas air curing led to moisture loss and inferior performance in the MgO-GGBFS systems. Overall, GO-modified Portland cement was identified as the optimal self-sensing mortar composite, as it balances adequate structural strength against enhanced electrical conductivity, reliable self-sensing response, and practical workability. These findings confirm the strong potential of GO-modified cementitious composites for use in intrinsic, low-cost SHM.

Although this study demonstrated the feasibility of green self-sensing cementitious composites, several limitations should be acknowledged. First, this investigation primarily focused on mechanical and electrical responses; advanced microstructural

characterization techniques (e.g., scanning electron microscopy, energy dispersive X-ray spectroscopy, X-ray diffraction, and mercury intrusion porosimetry) were not employed to assess the GO dispersion, hydration/carbonation products, or pore structure evolution directly. Second, piezoresistive behavior was evaluated only for the optimal CGS mix under a limited number of loading cycles; long-term cyclic performance, fatigue behavior, drift, and hysteresis were not examined. Third, all tests were conducted under controlled laboratory conditions without considering the effects of temperature, humidity, or aggressive environments. Fourth, this study was restricted to small-scale specimens; scale effects relevant to structural applications were not addressed. Finally, the MgO–GGBFS specimens exhibited lower compressive strengths than the cement-based specimens, reflecting the need for further optimization to capitalize on the notable sustainability advantages of these materials. Therefore, future research should comprehensively investigate specimen microstructures, undertake piezoresistive testing of all mixes under multiple loading modes and long-term cyclic loading regimes, evaluate the environmental and durability effects, optimize the dosage of GO and hybrid conductive fillers, scale specimens up to structural element size to evaluate the performance of the sensing networks embedded within, and optimize MgO–GGBFS formulations (including the application of carbonation curing and activators). Furthermore, multiphysics modeling and field-oriented application studies are recommended to support the practical deployment of sustainable self-sensing cementitious composites for SHM.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Mohammed Al Abdulmohsin – Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;
 ID 0009-0004-3476-7469
 Email: s202185590@kfupm.edu.sa

Corresponding Author

Asad Hanif – Mentor, Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia. Interdisciplinary Research Center for Construction and Building Materials, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; ID 0000-0002-4318-7712
 Email: asad.hanif@kfupm.edu.sa;

ACKNOWLEDGEMENTS

The authors express their sincere appreciation to the Civil and Environmental Engineering Department at King Fahd University of Petroleum & Minerals (KFUPM) for providing laboratory facilities and continuous academic support throughout this research. Special thanks are also extended to the Deanship of Research for its valuable assistance and encouragement in advancing undergraduate research projects at the university.

REFERENCES

- (1) Mehta, P. K. & Monteiro, P. J. *Concrete Microstructure, Properties, and Materials*. (McGraw-hill, 2006).
- (2) Worrell, E., Price, L., Martin, N., Hendriks, C. & Meida, L. O. CARBON DIOXIDE EMISSIONS FROM THE GLOBAL CEMENT INDUSTRY. *Annu. Rev. Energy Environ.* **26**, 303–329 (2001).
- (3) Neville, A. M. *Properties of Concrete*. vol. 4 (Longman London, 1995).
- (4) Bertolini, L., Elsener, B., Pedersen, P., Redaelli, E. & Polder, R. B. *Corrosion of Steel in Concrete: Prevention, Diagnosis, Repair*. (John Wiley & Sons, 2013).
- (5) Glisic, B. & Inaudi, D. *Fibre Optic Methods for Structural Health Monitoring*. (John Wiley & Sons, 2007).
- (6) Farrar, C. R. & Worden, K. An introduction to structural health monitoring. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **365**, 303–315 (2007).
- (7) Lynch, J. P. & Loh, K. J. A summary review of wireless sensors and sensor networks for structural health monitoring. *Shock Vib. Dig.* **38**, 91–130 (2006).
- (8) Ko, J. M. & Ni, Y. Q. Technology developments in structural health monitoring of large-scale bridges. *Eng. Struct.* **27**, 1715–1725 (2005).
- (9) Brownjohn, J. M. W. Structural health monitoring of civil infrastructure. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **365**, 589–622 (2007).
- (10) Measures, R. M. & Abrate, S. Structural monitoring with fiber optic technology. *Appl Mech Rev* **55**, B10–B11 (2002).
- (11) Spencer, B. F., Ruiz-Sandoval, M. E. & Kurata, N. Smart sensing technology: opportunities and challenges. *Struct. Control Health Monit.* **11**, 349–368 (2004).
- (12) Han, B., Yu, X. & Ou, J. *Self-Sensing Concrete in Smart Structures*. (Butterworth-Heinemann, 2014).
- (13) Dong, W., Li, W., Tao, Z. & Wang, K. Piezoresistive properties of cement-based sensors: Review and perspective. *Constr. Build. Mater.* **203**, 146–163 (2019).
- (14) Wen, S. & Chung, D. D. L. Double percolation in the electrical conduction in carbon fiber reinforced cement-based materials. *Carbon* **45**, 263–267 (2007).
- (15) Han, B., Ding, S. & Yu, X. Intrinsic self-sensing concrete and structures: A review. *Measurement* **59**, 110–128 (2015).
- (16) Shahzad, S., Toumi, A., Balayssac, J.-P. & Turatsinze, A. An experimental approach to assess the sensitivity of a smart concrete. *Buildings* **13**, 2034 (2023).
- (17) Han, B., Zhang, K., Yu, X., Kwon, E. & Ou, J. Electrical characteristics and pressure-sensitive response measurements of carboxyl MWNT/cement composites. *Cem. Concr. Compos.* **34**, 794–800 (2012).
- (18) D'Alessandro, A., Rallini, M., Ubertini, F., Materazzi, A. L. & Kenny, J. M. Investigations on scalable fabrication procedures for self-sensing carbon nanotube cement-matrix composites for SHM applications. *Cem. Concr. Compos.* **65**, 200–213 (2016).
- (19) Konsta-Gdoutos, M. S., Metaxa, Z. S. & Shah, S. P. Multi-scale mechanical and fracture characteristics and early-age strain capacity of high performance carbon nanotube/cement nanocomposites. *Cem. Concr. Compos.* **32**, 110–115 (2010).
- (20) Liew, K. M., Kai, M. F. & Zhang, L. W. Carbon nanotube reinforced cementitious composites: An overview. *Compos. Part Appl. Sci. Manuf.* **91**, 301–323 (2016).
- (21) Yoo, D.-Y., You, I. & Lee, S.-J. Electrical and piezoresistive sensing capacities of cement paste with multi-walled carbon nanotubes. *Arch. Civ. Mech. Eng.* **18**, 371–384 (2018).
- (22) Azhari, F. & Banthia, N. Cement-based sensors with carbon fibers and carbon nanotubes for piezoresistive sensing. *Cem. Concr. Compos.* **34**, 866–873 (2012).

- (23) Ding, S., Ruan, Y., Yu, X., Han, B. & Ni, Y.-Q. Self-monitoring of smart concrete column incorporating CNT/NCB composite fillers modified cementitious sensors. *Constr. Build. Mater.* **201**, 127–137 (2019).
- (24) Han, B., Zhang, K., Yu, X., Kwon, E. & Ou, J. Nickel particle-based self-sensing pavement for vehicle detection. *Measurement* **44**, 1645–1650 (2011).
- (25) Dreyer, D. R., Park, S., Bielawski, C. W. & Ruoff, R. S. The chemistry of graphene oxide. *Chem. Soc. Rev.* **39**, 228–240 (2010).
- (26) Zhu, Y. *et al.* Graphene and Graphene Oxide: Synthesis, Properties, and Applications. *Adv. Mater.* **22**, 3906–3924 (2010).
- (27) Compton, O. C. & Nguyen, S. T. Graphene Oxide, Highly Reduced Graphene Oxide, and Graphene: Versatile Building Blocks for Carbon-Based Materials. *Small* **6**, 711–723 (2010).
- (28) Babak, F., Abolfazl, H., Alimorad, R. & Parviz, G. Preparation and Mechanical Properties of Graphene Oxide: Cement Nanocomposites. *Sci. World J.* **2014**, 1–10 (2014).
- (29) Du, H., Gao, H. J. & Dai Pang, S. Improvement in concrete resistance against water and chloride ingress by adding graphene nanoplatelet. *Cem. Concr. Res.* **83**, 114–123 (2016).
- (30) Lv, S. *et al.* Effect of graphene oxide nanosheets of microstructure and mechanical properties of cement composites. *Constr. Build. Mater.* **49**, 121–127 (2013).
- (31) Pan, Z. *et al.* Mechanical properties and microstructure of a graphene oxide–cement composite. *Cem. Concr. Compos.* **58**, 140–147 (2015).
- (32) Sharma, S. & Kothiyal, N. C. Influence of graphene oxide as dispersed phase in cement mortar matrix in defining the crystal patterns of cement hydrates and its effect on mechanical, microstructural and crystallization properties. *Rsc Adv.* **5**, 52642–52657 (2015).
- (33) Mo, L., Deng, M., Tang, M. & Al-Tabbaa, A. MgO expansive cement and concrete in China: Past, present and future. *Cem. Concr. Res.* **57**, 1–12 (2014).
- (34) Unluer, C. & Al-Tabbaa, A. Impact of hydrated magnesium carbonate additives on the carbonation of reactive MgO cements. *Cem. Concr. Res.* **54**, 87–97 (2013).
- (35) Jin, F. & Al-Tabbaa, A. Strength and hydration products of reactive MgO–silica pastes. *Cem. Concr. Compos.* **52**, 27–33 (2014).
- (36) Andrew, R. M. & Global, C. O. Emissions from Cement Production. *Earth Syst. Sci. Data* **10**, 195 (2018).
- (37) Dung, N. T. & Unluer, C. Improving the performance of reactive MgO cement-based concrete mixes. *Constr. Build. Mater.* **126**, 747–758 (2016).
- (38) Ruan, S. & Unluer, C. Comparative life cycle assessment of reactive MgO and Portland cement production. *J. Clean. Prod.* **137**, 258–273 (2016).
- (39) Mo, L., Zhang, F. & Deng, M. Mechanical performance and microstructure of the calcium carbonate binders produced by carbonating steel slag paste under CO₂ curing. *Cem. Concr. Res.* **88**, 217–226 (2016).
- (40) Unluer, C. & Al-Tabbaa, A. Enhancing the carbonation of MgO cement porous blocks through improved curing conditions. *Cem. Concr. Res.* **59**, 55–65 (2014).
- (41) Özbay, E., Erdemir, M. & Durmuş, H. İ. Utilization and efficiency of ground granulated blast furnace slag on concrete properties—A review. *Constr. Build. Mater.* **105**, 423–434 (2016).
- (42) Escalante, J. I. *et al.* Reactivity of blast-furnace slag in Portland cement blends hydrated under different conditions. *Cem. Concr. Res.* **31**, 1403–1409 (2001).
- (43) Wang, A., Zhang, C. & Sun, W. Fly ash effects: I. The morphological effect of fly ash. *Cem. Concr. Res.* **33**, 2023–2029 (2003).
- (44) Wang, Q., Yan, P. & Feng, J. A discussion on improving hydration activity of steel slag by altering its mineral compositions. *J. Hazard. Mater.* **186**, 1070–1075 (2011).
- (45) Yi YaoLin, Y. Y., Liska, M., Unluer, C. & Al-Tabbaa, A. Carbonating magnesia for soil stabilization. <https://www.cabidigitallibrary.org/doi/full/10.5555/20133342341> (2013).

Arabic News Dataset Collection and Generation for Deepfake-Text Detection

Khulud Alotaibi, Amal Sunba, and Tarek Helmy*

Cite <https://doi.org/10.64589/juri/218957>

Submitted: November 23, 2025 Revised: February 12, 2026 Accepted: February 19, 2026

ABSTRACT

The spread of fake news and AI-generated deepfakes poses a significant threat to public trust, particularly in Arabic-speaking contexts where resources for misinformation detection remain limited. Although prior studies have examined Arabic fake-news detection, existing datasets often rely on simplistic synthetic content or manually collected articles that do not fully reflect real-world journalistic writing. This study addresses this gap by constructing a high-quality dataset of 21,000 authentic news articles from Al Arabiya and CNN Arabic and generating an equal number of fabricated articles using the GPT-4.1-mini model. Four deception strategies were employed: context shifting, exaggeration, contradiction, and misattribution. The resulting dataset, comprising more than 42,000 articles, was used to fine-tune the bidirectional encoder representations from transformers (BERT)-base-uncased model, achieving 96% accuracy in distinguishing real from fake news. This work introduces a large-scale Arabic news dataset, a reproducible deepfake generation methodology for Arabic news, and a transformer-based baseline for misinformation detection. However, the study is limited to modern standard Arabic and text-only content, highlighting the need for future research on cross-domain and dialectal robustness.

Keywords: fake news detection, deepfake generation, Arabic language processing, transformer models, BERT, misinformation detection

1. INTRODUCTION

Misinformation is a global concern that undermines public trust and social stability. Digital platforms enable fabricated information to spread at an unprecedented scale^{1,2}. In the Arabic-speaking world, this challenge is acute because misinformation research remains limited, and datasets are small and domain-specific³⁻⁵. These constraints hinder the development of reliable Arabic models for detecting deceptive content. Arabic's linguistic properties further complicate detection. As a morphologically rich language with optional diacritics and dialectal variation, it increases data sparsity and limits generalization, especially for long-form journalistic text written in modern standard Arabic (MSA).

Recent advances in natural language processing (NLP) have improved fake-news detection, especially in high-resource languages such as English. Benchmark datasets including LIAR⁶ and FEVER⁷, have supported transformer-based models that classify news using contextual patterns. However, Arabic remains challenging due to structural complexity and diglossia, encompassing MSA and regional dialects. Earlier studies relied on traditional machine learning models and were constrained by limited data. Recent work has introduced Arabic fake-news datasets and transformer-based approaches. However, most remain small, domain-restricted, or focused on short texts such as claims or social media posts⁸⁻¹⁰.

This study addresses these gaps by presenting a balanced Arabic fake-news dataset of over 42,000 articles, evenly divided between real and fake content. Unlike existing resources, the dataset preserves journalistic structure and applies controlled, domain-specific deception across domains. It also evaluates a fine-tuned bidirectional encoder representation from transformers (BERT)-base-uncased model for binary classification, establishing a benchmark for Arabic misinformation detection. Figure 1 summarizes the workflow from data collection to deepfake generation, dataset preparation, and model evaluation.

The paper is organized as follows. Section 2 reviews related work. Section 3 describes the methodology, including data collection, deepfake generation, and data cleaning. Section 4 outlines the experimental setup and model implementation. Section 5 presents the results and comparative analysis. Section 6 concludes with key findings and future research directions.

2. RELATED WORK

Several datasets have been proposed for Arabic fake-news and misinformation detection, differing in scale, domain coverage, and generation methodology. Early efforts focused on social media platforms, particularly X (formerly Twitter). For example, ArCOV19-Rumors collected Arabic COVID-19 claims for event-specific misinformation detection⁴. However, these datasets are

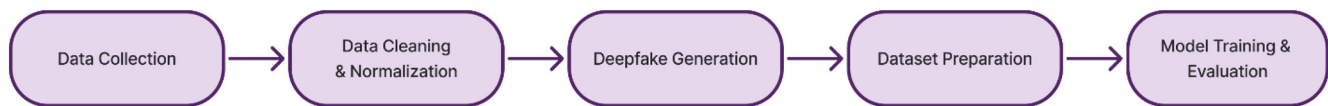


Figure 1. Workflow of the proposed Arabic deepfake-text detection study

limited to short informal texts and specific events, restricting their applicability to long-form journalistic content.

Other studies examined claim-level detection. Baly et al.³ introduced a stance-based corpus linking claims to evidence, emphasizing fact verification rather than full-article classification. Alnabrisi and Saad⁹ developed the ArabicFakeNews dataset with approximately 2,000 articles and reported strong AraBERT performance; however, the dataset remains small and domain-constrained.

Recent work also explored AI-generated Arabic text. Harrag et al.¹³ evaluated GPT-2-generated tweets using transformer models, focusing on short texts. Khairallah and Zubiaga¹⁰ proposed ALHD, a benchmark for detecting large language model (LLM)-generated Arabic text across genres; however, it is not limited to journalistic news and does not apply controlled deception at the article level.

These limitations reveal gaps in scale, domain diversity, and text length. The proposed dataset addresses these gaps by providing balanced, domain-diverse Arabic news articles, enabling more realistic evaluation of deepfake detection models. Table 1 summarizes representative datasets by size, text type, domain coverage, and generation methodology.

3. METHODOLOGY

3.1. Real Data Collection. We collected authentic Arabic news articles using WebScraper.io¹¹, a browser-based web scraping tool designed to extract structured data from official news websites. This method was selected for its efficiency in gathering large volumes of text without requiring custom coding. Using the tool's graphical interface, articles were scraped from two Arabic-language sources, Al Arabiya and CNN Arabic, both chosen for their editorial credibility and broad coverage of regional and international topics.

The scraper supported multi-page navigation and direct export to comma-separated values (CSV) format. This feature enabled the organized collection of metadata, including titles, content, and domain. The final dataset of real articles comprised 21,591 unique records across five domains: 7,224 in politics, 6,629 in technology, 3,876 in health, 2,920 in the Saudi economy, and 942 in sports. Figure 2 illustrates the domain distribution by

source website. Al Arabiya contributed the majority of articles, whereas CNN Arabic provided fewer but complementary articles, particularly in politics and sports. All articles were written in MSA.

Clear inclusion criteria guided article selection. Only articles with complete text were retained, whereas duplicate entries, incomplete pages, and non-editorial content were excluded. In addition, articles were filtered to ensure relevance to the pre-defined domains and balanced topical coverage across sources. Each record in the final CSV file contained the following fields: ID, Link, Title, Content, Date, Domain, Dialect (MSA), Source, Article Location, and Article Origin (Source Website).

After collection, all data were imported into Jupyter Notebook for cleaning and normalization. The raw text from both websites contained various web-related artifacts, including duplicate headlines, embedded social media posts, and advertisement placeholders. These were systematically removed using customized regular expression rules that targeted each site's structural patterns. For example, recurring section titles such as category chains (شركات – اقتصاد السعودية – مبادرة الاستثمار) were identified and removed, leaving only the main journalistic content.

Additional preprocessing steps were taken to ensure that all texts were consistent in orthography and linguistic form. HTML tags and non-Arabic characters were removed, and character variants were normalized to the *alef* and *hamza* forms. Non-Arabic characters, emojis, and excess punctuation were deleted to minimize tokenization noise. All real articles were then reviewed and annotated by a domain expert in Arabic news to validate domain assignments, remove miscategorized items, and confirm the linguistic and journalistic correctness of each text. This expert-driven validation step ensured the quality and reliability of the final dataset.

The adopted data collection and preprocessing strategy follows best practices commonly applied in low-resource language settings. Similar principles were employed by Rahman et al.¹⁴ in their work on large-scale Urdu document clustering, in which news articles were collected from credible sources and subjected to extensive cleaning and normalization before downstream analysis. Inspired by this methodology, the present study emphasizes source credibility, article completeness, and linguistic normal-

Table 1. Comparison of Arabic fake-news and deepfake-text datasets

Dataset	Year	Size	Text Type	Domains	Generation Method
ArCOV19-Rumors	2021	~10K	Tweets	Health	Social media rumors
ArabicFakeNews	2024	~2K	Articles	Mixed	Human-written
GPT-2 Arabic Tweets	2020	~5K	Tweets	General	GPT-2 generation
ALHD	2025	>50K	Mixed genres	Multi-domain	LLM-generated
This Work	2025	43K	Full articles	5 news domains	Controlled GPT-4.1-mini

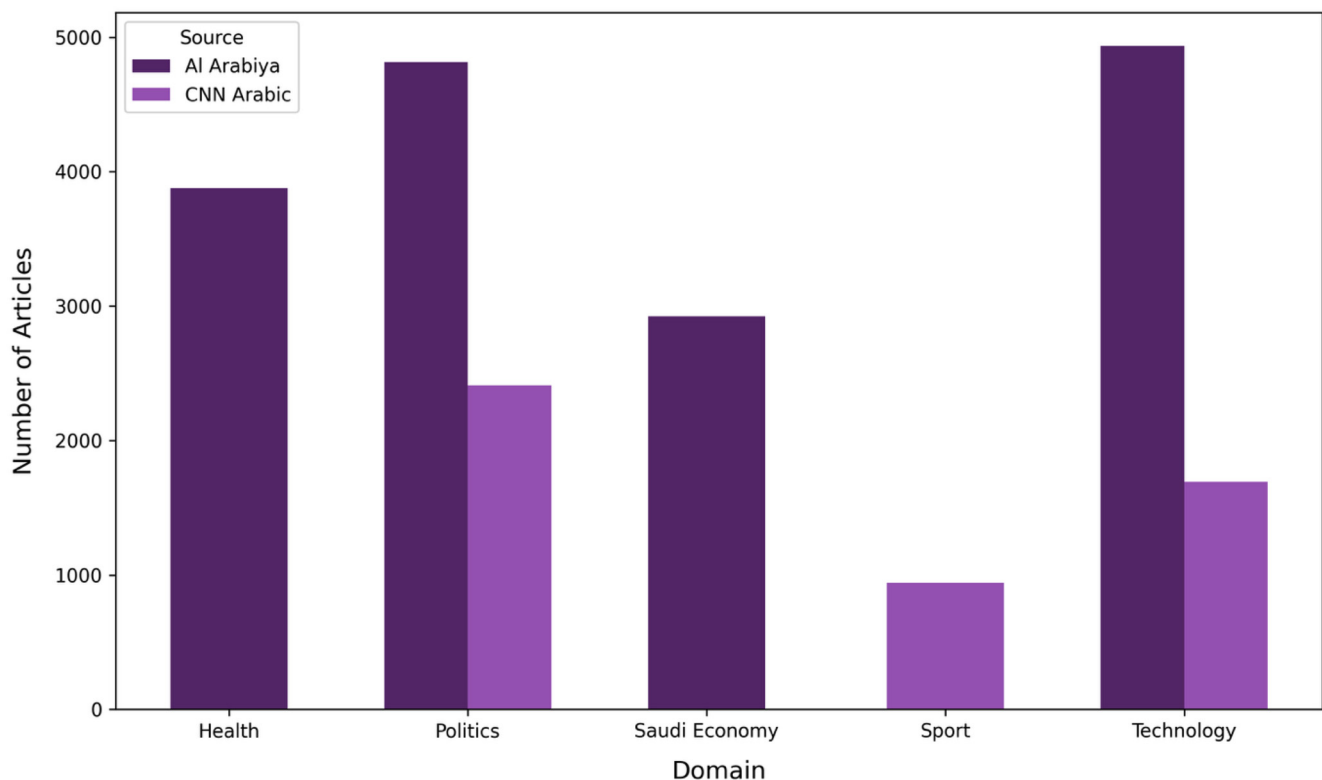


Figure 2. Domain distribution for the real data

ization to ensure the reliability and usability of the constructed Arabic news corpus. Table 2 summarizes the source-level statistics, including the estimated volume of available articles and the number selected for inclusion in the final dataset.

3.2. Deepfake Data Generation. An equal number of 21,591 synthetic articles was generated to establish a balanced dataset. Generation was performed in Jupyter Notebook using the OpenAI GPT-4.1 Mini language model via the OpenAI API (API version: March 2025). Each synthetic article was generated by conditioning the model on a corresponding real article and applying a predefined deception strategy, ensuring a one-to-one pairing between real and fake samples.

For reproducibility, a fixed prompt template was used across all articles, with only three variables changing per instance: the article text, its domain label, and the assigned deception strategy. Each prompt consisted of the following components: (1) input article text (a fully cleaned Arabic news article), (2) domain specification (politics, technology, health, Saudi economy, or sports), and (3) explicit deception instruction corresponding to the assigned strategy. The prompt template followed this structure:

“Rewrite the following Arabic news article as a realistic journalistic report. Preserve the writing style, tone, and structure of professional Arabic news. Apply the following deception strategy: [DECEPTION TYPE], while maintaining linguistic fluency and coherence. Do not introduce obvious contradictions or grammatical errors.”

Each domain used a single consistent deception strategy to ensure controlled generation: political articles were generated through misattribution, in which sources or officials were intentionally altered. Technology articles followed an exaggeration strategy, overstating innovations or performance claims to simulate promotional bias. Health articles employed contradiction, introducing claims that contradict verified medical information. Saudi economy articles employed context shifting, modifying the locations, dates, or circumstances of real events, whereas sports articles used contradiction by altering results or athletes’ performances.

All generations used the same model configuration: temperature = 0.9, top-p = 1.0, and maximum output length = 500 tokens. No additional frequency or presence penalties were applied, and the model was prompted with a single user role

Table 2. Source-level statistics of real news articles

Source	Available Articles*	Selected Articles	Domains Covered
Al Arabiya	~500,000 +	14,872	Politics, Economy, Health, Technology, Sports
CNN Arabic	~300,000 +	6,719	Politics, Health, Sports
Total	~800,000 +	21,591	5 domains

*Available article counts are approximate estimates inferred from publicly accessible site archives, sitemap coverage, and large-scale academic news corpora.

without system-level instructions. Each generated article was constrained to 200–300 words, matching the length distribution of the real articles. The generation process was automated in batches using Python scripts, with progress tracked via the tqdm library. Short delays were inserted between API calls to comply with OpenAI rate limits and ensure stable execution.

Post-processing included removing newline artifacts and extra formatting. A subset of the generated samples was manually reviewed to ensure fluency and topical relevance, and any incoherent or illogical outputs were regenerated. The final collection contained 21,591 fake articles that closely resembled authentic news, each paired with its corresponding real article.

3.3. Dataset Preparation. The real and deepfake datasets were merged to form a unified corpus of 43,182 Arabic news articles, with balanced representation of real and fake samples. Each record contained the text and its label, where 0 indicated a real article, and 1 indicated a fake article. The dataset was randomly shuffled to ensure reproducibility. To prepare for model training, the dataset was split into three subsets: 70% for training, 15% for validation, and 15% for testing. A stratified sampling technique was used to preserve class balance across all partitions. This ensured that real and fake samples were proportionally represented in each subset. Although the dataset is balanced between real and fake samples, it is naturally imbalanced across domains, as confirmed by the domain expert. Because evaluation is based on a global binary label rather than domain-level classification, this imbalance does not affect the overall performance metric, which is reported as a single unified score.

4. EXPERIMENTAL WORK

4.1. Model Architecture. The fine-tuning process was based on the BERT-base-uncased architecture¹². BERT captures contextual dependencies in both directions, left-to-right and right-to-left. This bidirectional approach allows the model to understand the full context of a sentence rather than isolated words, making it particularly suitable for detecting stylistic and contextual irregularities that may appear in deepfake text.

BERT was selected over recurrent neural network (RNN)-based architectures such as LSTM and GRU because of its superior ability to model long-range dependencies. While RNNs process text sequentially and may suffer from vanishing gradients or limited context retention in long documents, BERT processes tokens in parallel using self-attention, allowing it to learn global contextual relationships across entire articles. This advantage is especially important for long-form Arabic news texts, which often exhibit complex syntactic structures and nuanced discourse patterns. Moreover, transformer-based models have consistently outperformed RNNs in misinformation detection and text classification tasks, establishing BERT as a strong baseline for this study.

Although BERT-base-uncased was originally pre-trained on large English corpora, it was selected for its demonstrated generalization capability for normalized MSA text. The model consists of 12 transformer encoder layers, each containing 12 attention heads and producing 768-dimensional hidden representations. This architecture enables rich semantic encoding and multi-head

attention across diverse linguistic features, making it suitable for modeling long-form Arabic news articles. The term “uncased” indicates that the model processes text without distinguishing letter case, which aligns with Arabic script, as it does not encode case distinctions. The overall architecture used in this study is illustrated in Figure 3.

During fine-tuning, the contextualized [CLS] token represents the entire article. This representation is fed into a linear classification layer, followed by a softmax function to predict whether the article is real or fake. All model parameters were updated during training, enabling the model to adapt to Arabic sentence structures, writing styles, and domain-specific characteristics in the dataset.

4.2. Model Training. The fine-tuning was performed on Google Colab Pro using its built-in NVIDIA Tesla T4 GPU, which provided sufficient computational capacity to process thousands of long-text samples efficiently. The Trainer API from Hugging Face was used to automate dataset loading, gradient computation, and evaluation after each epoch. The AdamW optimizer was used with a learning rate of 2×10^{-5} and a weight decay of 0.01 to balance learning stability and model generalization. Both the training and validation batch sizes were set to 32 to balance performance and GPU memory constraints. Training was conducted for three epochs, chosen empirically after testing different configurations and observing convergence behavior and validation loss stability. Each epoch included periodic validation every 500 steps to track model improvement. The model automatically saved the best checkpoint based on the validation F1-score, ensuring that the optimal parameters were preserved. All logs, checkpoints, and intermediate results were stored in Google Drive for backup and reproducibility. Upon completing training, the fine-tuned model and tokenizer were uploaded to the Hugging Face Hub, allowing future use, sharing, and verification.

4.3. Evaluation Metrics. To evaluate model performance, four standard metrics were used: accuracy, precision, recall, and F1-score. These metrics were implemented through a custom evaluation function integrated within the Trainer API. Accuracy measured the overall proportion of correctly classified samples, precision quantified the ratio of correctly predicted fake articles to all predicted fake articles, recall measured the model’s ability to identify all actual fake samples correctly, and the F1-score provided a balanced harmonic mean of precision and recall. The F1-score was also employed as the primary criterion for selecting the best-performing checkpoint during training.

In addition to quantitative metrics, a confusion matrix and classification reports were generated for both the validation and test sets, providing insight into the distribution of correct and incorrect predictions across labels. These analyses also helped identify any systematic misclassification across deception strategies or domains. The detailed numerical outcomes and analysis are presented in the following section.

5. RESULTS AND DISCUSSION

5.1. Training and Validation Performance. During training, the model demonstrated rapid, stable convergence

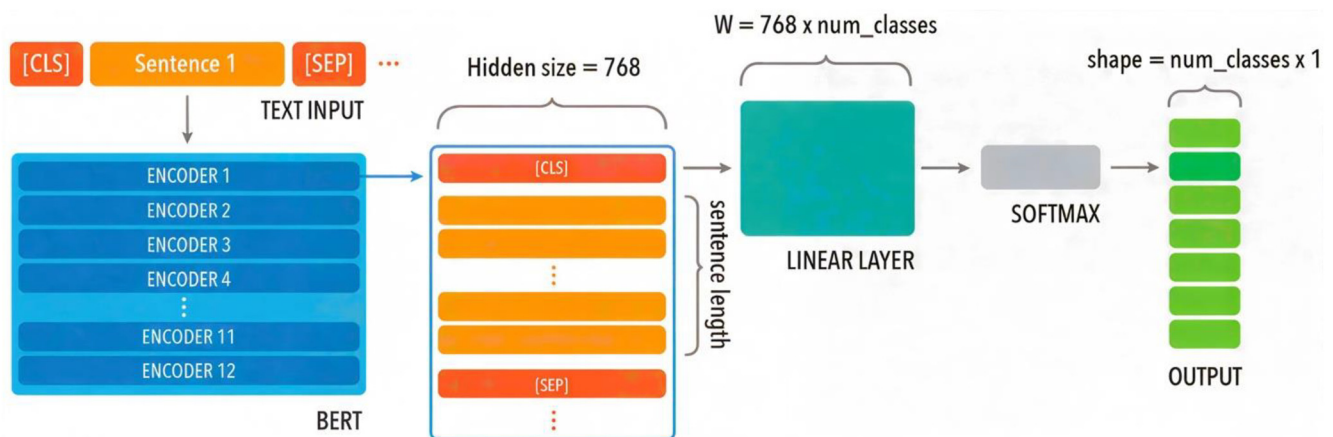


Figure 3. BERT-base-uncased architecture used for Arabic deepfake-text classification. Adapted from Elbatarny et al.¹⁵

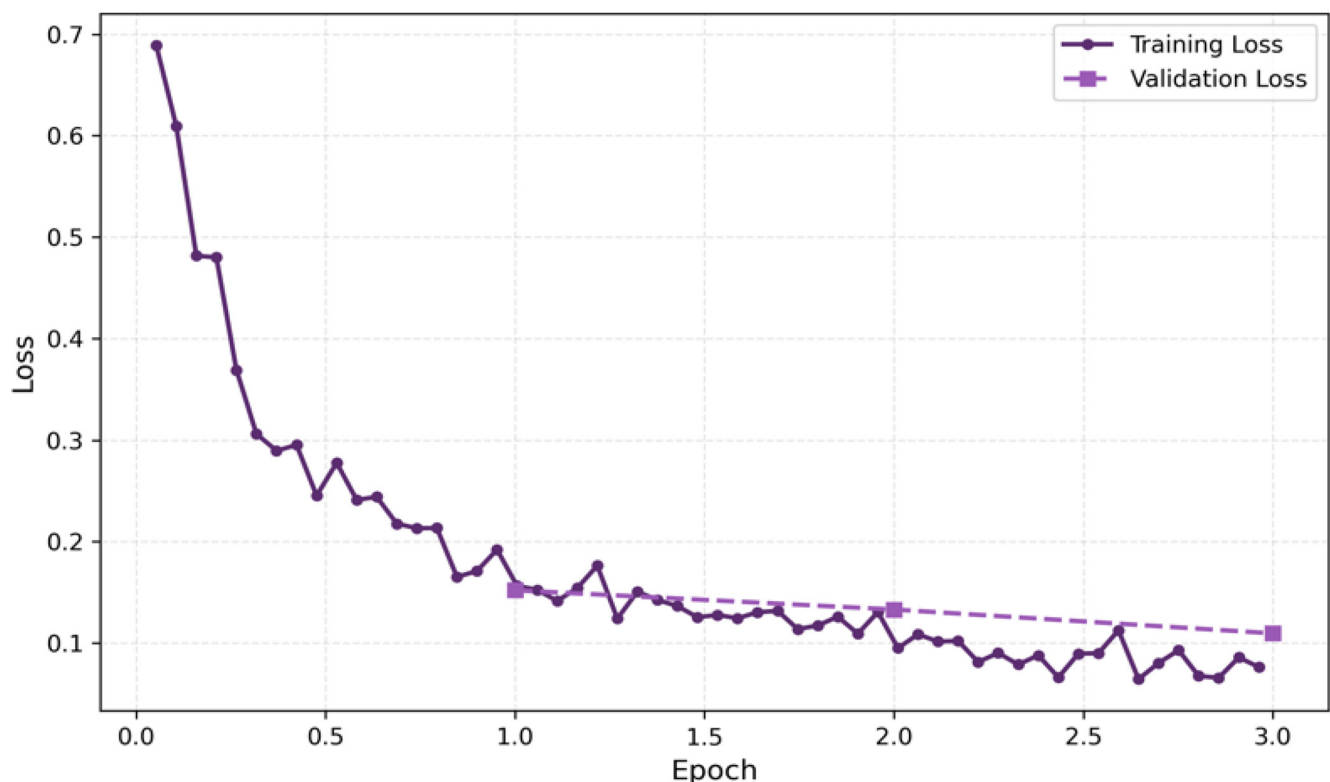


Figure 4. Training and validation loss curves across three epochs

within three epochs, with consistent decreases in training and validation loss. Figure 4 illustrates this trend, showing that both loss curves converge steadily and exhibit minimal overfitting during fine-tuning. The validation accuracy reached 96.3%, indicating that the model learned consistent patterns without signs of overfitting. As listed in Table 3, precision and recall values are nearly identical across the real and fake classes, confirming balanced performance. The fake class achieved slightly higher recall, indicating that the model was more sensitive in detecting deceptive articles, whereas the real class achieved higher precision, reflecting fewer false positives.

This balance indicates that the model did not favor one label over the other, which is crucial in misinformation detection, as labeling real news as fake may undermine credibility. The confusion matrix in Figure 5 illustrates this clearly, with most samples

falling along the main diagonal, indicating that the classifier correctly identified nearly all articles. The matrix corresponds to the third validation epoch, which achieved the best F1-score and was automatically saved as the optimal checkpoint during training. The remaining errors involved only a small number of borderline cases, typically articles that shared overlapping linguistic cues such as neutral tone.

5.2. Test-Set Performance. When evaluated on the test set of 6,477 unseen articles, the model achieved an accuracy of 96.62%, confirming its ability to generalize beyond the training data. As shown in Table 4, both real and fake categories achieved nearly equal F1-scores, maintaining the performance stability observed during validation. The precision for real articles at 98.21% indicates a low false detection rate. In contrast,

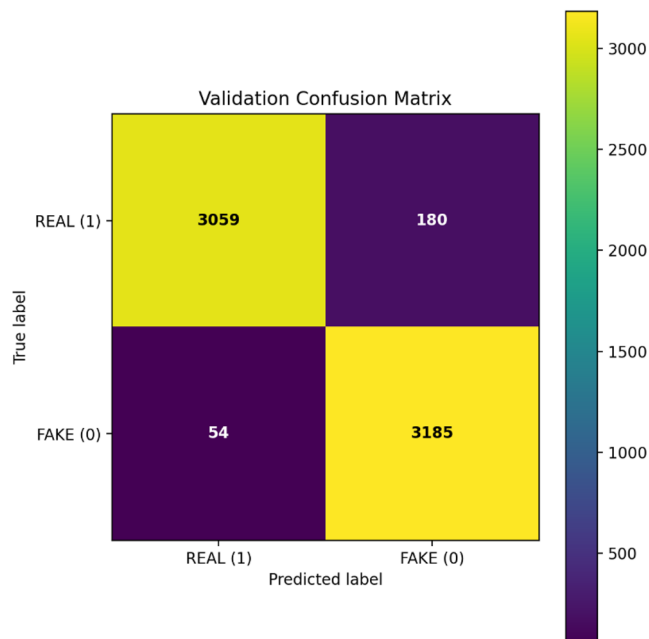


Figure 5. Confusion matrix for the validation set at the final training epoch

the recall for fake articles at 98.27% demonstrates the model's strong ability to identify deceptive samples. The confusion matrix in Figure 6 supports this finding, showing that nearly all predictions were correct, with few real articles misclassified as fake. This strong generalization implies that the model captured not only word-level patterns but also structural and contextual cues common in Arabic news. Such cues include emotional tone, sentence complexity, and reporting style, all subtle yet powerful signals that distinguish authentic journalism from AI-generated imitations.

5.3. Error Analysis. A closer examination of the misclassified samples reveals two main types of errors. First, some real articles were classified as fake, especially short factual reports or formal summaries that lacked emotional or descriptive language. Their neutral tone resembled machine-generated writing, leading the model to misclassify them as inauthentic. Second, some fake articles were classified as real when the generated text underwent minor edits, such as replacing names or numbers while maintaining the journalistic tone. These patterns show that although the model is highly capable of recognizing linguistic irregularities, it struggles when the deception is minimal or stylistically sophisticated. This limitation indicates the model's reliance on textual surface patterns rather than factual verification. In other words, the BERT model excels at detecting how something is written but not at determining whether it is factually accurate. Addressing this gap would require integrating factual knowledge bases,

Table 3. Validation result

	Precision	Recall	F1-score
FAKE	0.9465	0.9833	0.9646
REAL	0.9827	0.9444	0.9632
Accuracy	0.9639	0.9639	0.9639
Macro Avg	0.9646	0.9639	0.9639

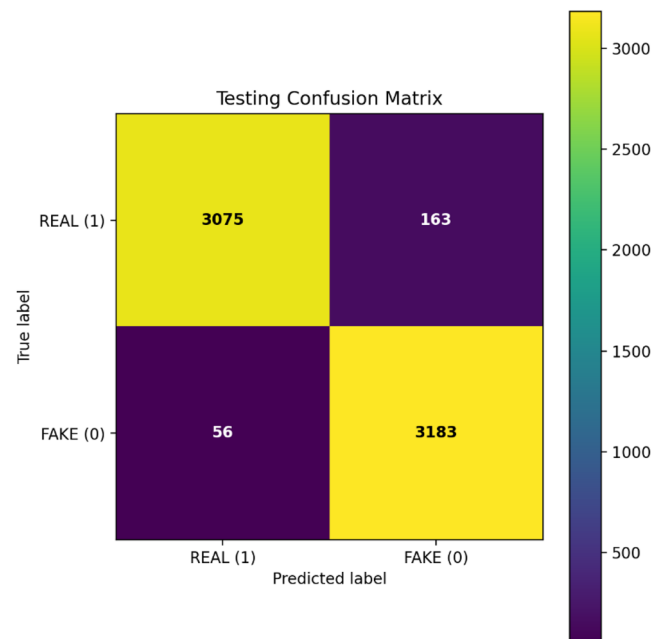


Figure 6. Testing confusion matrix

fact-checking models, or explicit reasoning mechanisms in future work.

5.4. Comparison with Related Work. Compared with previous Arabic fake-news detection studies reported in the literature, the current model achieved higher performance and broader domain coverage. The comparison focuses on transformer-based approaches applied to Arabic text, as these models constitute the current state of the art in misinformation detection. For instance, Baly et al.³ focused on small datasets linking news claims to stance labels, while Harrag et al.¹³ explored GPT-generated short texts using AraBERT on a few thousand samples. In contrast, this study fine-tuned a transformer on a larger, balanced corpus that included both human-written and AI-generated content. This scale enabled the model to learn nuanced representations of deception across politics, health, economy, technology, and sports, making it more robust than earlier models restricted to a single domain.

This study introduces a large article-level corpus of over 42,000 Arabic news samples, substantially larger than existing datasets, while preserving full journalistic structure rather than short claims or tweets. It also applies controlled, domain-specific deception strategies, combining large-scale data, structured deep-fake generation, and transformer-based benchmarking within a reproducible framework. Future work may compare transformer models such as BERT with traditional classifiers such as SVMs,

Table 4. Test-set results

	Precision	Recall	F1-score
FAKE	0.9513	0.9827	0.9667
REAL	0.9821	0.9497	0.9656
Accuracy	0.9662	0.9662	0.9662
Macro Avg	0.9667	0.9662	0.9662

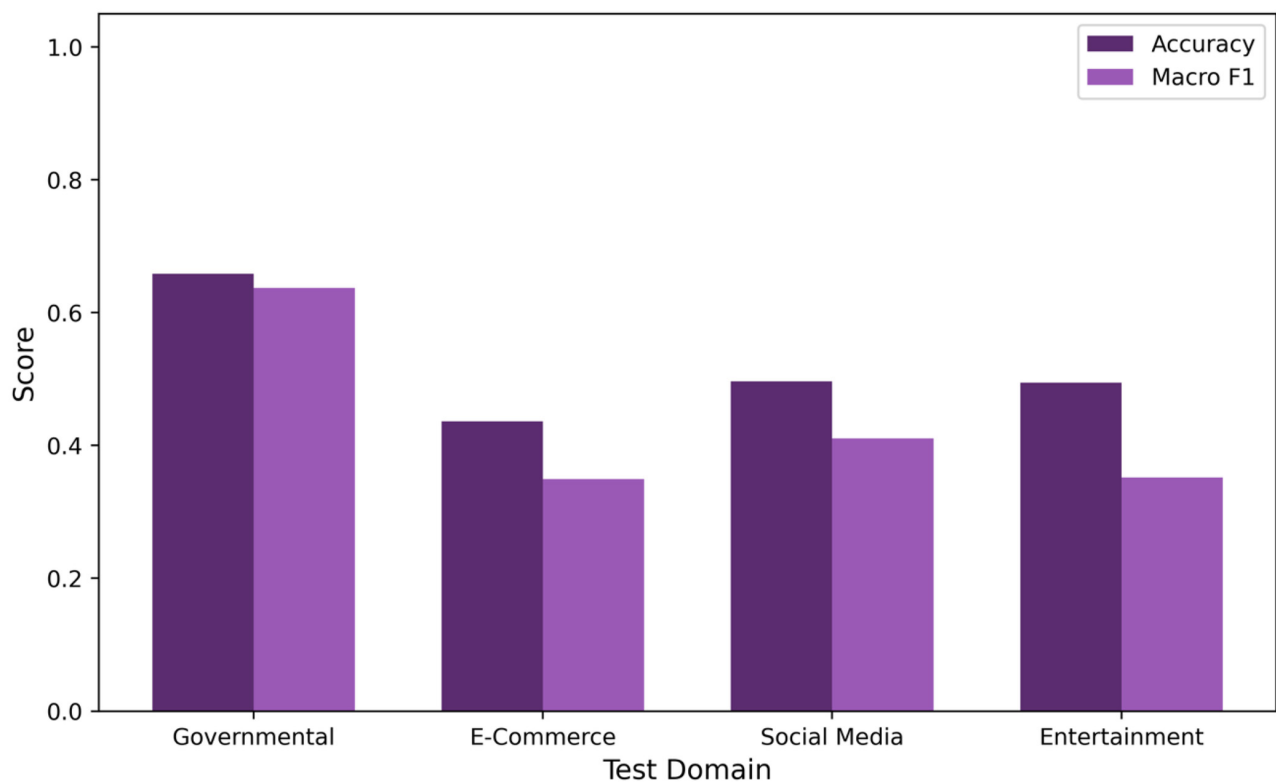


Figure 7. Cross-domain model performance comparison

which rely on manual features and often struggle to capture Arabic semantic dependencies.

For comparability, we report results from transformer-based models fine-tuned under similar training conditions, including three training epochs and an AdamW optimizer with a learning rate of 2×10^{-5} , as specified in the original studies. Where direct retraining on external datasets was not feasible due to differences in dataset availability, annotation schemes, or task formulation, we report published benchmark results following common practice in comparative NLP studies.

As summarized in Table 5, the proposed dataset, combined with a BERT-base-uncased model, achieves competitive performance relative to existing Arabic misinformation benchmarks while operating on longer texts and broader domains. Unlike prior datasets that primarily focus on short tweets, claims, or single-domain content, this study evaluates full-length journalistic articles and applies controlled, domain-specific deception strategies. These characteristics increase task difficulty and better reflect real-world misinformation scenarios, making direct accuracy comparisons conservative rather than inflated.

Table 5. Quantitative comparison with Arabic fake and deepfake-text datasets

Dataset	Model	Accuracy (%)	F1-score
ArabicFakeNews	AraBERTv2	97.9	-
GPT-2 Tweets	AraBERT	91.2	0.9
ALHD	XLM-R	93.4	0.93
This Work	BERT-base-uncased	96.6	0.97

5.5. Cross-Domain Evaluation. To further assess the model's adaptability beyond traditional news data, a cross-domain evaluation was conducted using four external datasets: e-commerce, social media, entertainment, and government. As summarized in Table 6 and illustrated in Figure 7, performance declines on informal text, with accuracies below 50% for e-commerce, social media, and entertainment. This degradation reflects the domain shift problem. These domains typically use conversational language, emojis, and incomplete sentences, which differ from the news writing style. Consequently, the model's learned patterns no longer aligned with the input style.

In contrast, the governmental dataset achieved the highest cross-domain performance, with 65.8% accuracy and 0.64 macro-F1, showing better alignment with traditional news style. This suggests that BERT learned features closely tied to formal Arabic syntax and journalistic tone rather than semantic truthfulness. Future studies can address this limitation by applying domain adaptation methods, such as fine-tuning on diverse corpora that include both formal and informal text, or by using adversarial training, which helps models focus on deeper semantic understanding rather than surface-level writing style. Another potential direction is to include factual consistency checks or external

Table 6. Cross-domain evaluation results

Dataset	Accuracy	Macro F1
E-commerce	0.4362036455	0.3491038089
Social media	0.4963572268	0.4107288411
Entertainment	0.4946140036	0.3517277592
Governmental	0.6582294912	0.6369536261

knowledge graphs to improve the detection of subtle factual distortions that purely textual models may overlook.

6. CONCLUSIONS

This study presented a large Arabic news corpus and proposed a controlled GPT-generated deepfake methodology. A fine-tuned BERT-base-uncased model achieved 96% accuracy in distinguishing real from fake articles, confirming the effectiveness of transformer architectures for Arabic text classification. Results highlighted strong performance on formal news with 96.6% accuracy, but limited cross-domain generalization. Among the external datasets, the governmental domain achieved the highest cross-domain accuracy of 65.8%, emphasizing the need for dialectal data and domain adaptation strategies. Looking ahead, future systems could incorporate dialectal data, utilize multilingual transfer learning, and improve resistance to adversarial manipulation such as paraphrasing or disguised phrasing. Since this study has some limitations, such as being limited to textual data, future research could explore integrating multimodal features, such as images, videos, and author metadata, to enhance reliability and scalability. Finally, ethical transparency must remain a priority. Given the sensitivity of misinformation in the Arabic context, future models should be developed responsibly, with safeguards against bias and misuse.

The GPT-generated articles were created under controlled conditions, which may not cover the full range of real-world misinformation or cases involving images and videos. The dataset also focuses on MSA, limiting generalization to dialectal and informal language. From an ethical standpoint, all synthetic articles were stored safely, clearly labeled as artificial, and used only for research. No misinformation was publicly released, and the study adhered to responsible AI and data-handling practices to ensure its findings contribute positively to combating misinformation rather than spreading it. In conclusion, this study addresses an essential gap in Arabic NLP by providing a reproducible dataset and a baseline model for fake-news detection, laying the groundwork for more diverse, ethically aligned, and resilient approaches to reducing digital misinformation.

7. DECLARATION

AI-assisted tools such as ChatGPT and grammar checkers were used only for spelling, formatting, and language refinement. They did not assist in designing the research, creating datasets, analyzing results, or writing the scientific content. All research, data processing, and model development were conducted solely by the authors.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Khulud Alotaibi – Information and Computer Science Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0009-0007-5083-6032
Email: s202168730@kfupm.edu.sa

Author

Amal Sunba – Research collaborator, Information and Computer Science Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0009-0002-4712-2104
Email: g202203780@kfupm.edu.sa

Corresponding Author

Tarek Helmy – Research Mentor, Professor, Information and Computer Science Department & Interdisciplinary Research Center for Intelligent Secure Systems, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0000-0001-9683-682X
Email: helmy@kfupm.edu.sa

ACKNOWLEDGEMENTS

This research was conducted under a collaboration between the General Authority for Defense Development (GADD) and King Fahd University of Petroleum & Minerals (KFUPM), under agreement number CSSP2643, Saudi Arabia. The authors acknowledge the General Authority for Defense Development (GADD) in Saudi Arabia for funding this research through project number (GADD_2024_01_438). The authors also acknowledge the support provided by the Information and Computer Science Department and the Interdisciplinary Research Center for Intelligent Secure Systems at King Fahd University of Petroleum and Minerals, Saudi Arabia.

REFERENCES

- (1) Shu, K., Sliva, A., Wang, S., Tang, J. & Liu, H. Fake news detection on social media: a data mining perspective. *ACM SIGKDD Explor. Newsl.* **19**, 22–36 (2017).
- (2) Zhou, X. & Zafarani, R. A survey of fake news: fundamental theories, detection methods, and opportunities. *ACM Comput. Surv.* **53**, 1–40 (2020).
- (3) Baly, R., Mohtarami, M., Glass, J., Márquez, L., Moschitti, A. & Nakov, P. Integrating stance detection and fact checking in a unified corpus. *NAACL-HLT*, 21–27 (2018).
- (4) Haouari, F., Hasanain, M., Suwaileh, R. & Elsayed, T. ArCOV19-Rumors: Arabic COVID-19 Twitter dataset for misinformation detection. *Arabic Nat. Lang. Process. Workshop*, 72–81 (2021).
- (5) Alam, F. *et al.* Fighting the COVID-19 infodemic: modeling the perspective of journalists, fact-checkers, social media platforms, policymakers, and society. *Findings Assoc. Comput. Linguist. EMNLP*, 611–649 (2021).
- (6) Wang, W. Y. “Liar, Liar Pants on Fire”: a new benchmark dataset for fake news detection. *ACL*, 422–426 (2017).
- (7) Thorne, J., Vlachos, A., Christodoulopoulos, C. & Mittal, A. FEVER: a large-scale dataset for fact extraction and verification. *NAACL-HLT*, 809–819 (2018).
- (8) Yousef, M. A., ElKorany, A. & Bayomi, H. Fake-news detection: a survey of evaluating Arabic datasets. *Soc. Netw. Anal. Min.* **14**, 225 (2024).

- (9) Alnabrisi, I. & Saad, M. Detect Arabic fake news through deep learning models and transformers. *SSRN Electron. J.* (2024).
- (10) Khairallah, A. & Zubiaga, A. ALHD: a large-scale and multi-genre benchmark dataset for Arabic LLM-generated text detection. *arXiv 2510.03502* (2025).
- (11) WebScaper.io. WebScaper: a browser-based tool for structured web data extraction (2025).
- (12) Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. *NAACL-HLT*, 4171–4186 (2019).
- (13) Harrag, F., Dabbah, M., Darwish, K. & Abdelali, A. BERT transformer model for detecting Arabic GPT2 auto-generated tweets. *Arabic Nat. Lang. Process. Workshop*, 207–214 (2020).
- (14) Rahman, A. U., Khan, K., Khan, W., Khan, A. & Saqia, B. Unsupervised machine learning based documents clustering in Urdu. *EAI Endorsed Trans. Scalable Inf. Syst.* (2018).
- (15) Elbatarny, L., Do, R. K. G., Gangai, N. & Ahmed, F. Applying natural language processing to single-report prediction of metastatic disease response using the OR-RADS lexicon. *Cancers* **15**, 4909 (2023).

Arabic Review Dataset for Deepfake Text Detection: Collection and Generation

Malath Alhashem, Amal Sunba, and Tarek Helmy*

Cite <https://doi.org/10.64589/juri/219353>

Submitted: November 23, 2025 Revised: February 16, 2026 Accepted: February 18, 2026

ABSTRACT

It has become increasingly challenging to distinguish between real and deepfake texts, particularly for low-resource languages such as Arabic. This study aims to construct a reliable Arabic dataset to enable deepfake text detection by collecting authentic YouTube comments and generating synthetic text using OpenAI GPT-4.0 Mini. The collected comments span four thematic domains—entertainment, religion, health, and sports—to capture common discussion topics and linguistic variations present in Arabic online communities. Synthetic samples were generated using a structured prompt-based methodology that applies predefined deception techniques to simulate realistic misleading content. To validate the proposed dataset, a Bidirectional Encoder Representations from Transformers (BERT)-based model was fine-tuned for binary classification. Experimental results achieved an accuracy of 91.43%, demonstrating strong classification capability and confirming the effectiveness of the dataset for deepfake detection tasks. Although the dataset remains limited in size and dialectal diversity, the results demonstrate the effectiveness of the proposed methodology. The dataset and methodology are expected to support future research in Arabic natural language processing and improve the reliability of automated deepfake detection approaches.

Keywords: deepfake text detection, machine learning, cross-validation, data collection, data preprocessing

1. INTRODUCTION

With the rapid development of artificial intelligence (AI) and natural language processing (NLP), deepfake text has emerged as a significant challenge. Deepfake text refers to synthetically generated content that closely mimics human language, making it increasingly difficult to distinguish between authentic and manipulated written text. This phenomenon raises serious concerns, as deepfakes may erode trust in digital media, increase cognitive load for readers, and trigger negative psychological responses¹.

Owing to these risks, research interest in deepfake text detection has increased in recent years. However, most existing studies have focused primarily on the English language, leaving a considerable gap in the detection of Arabic deepfake text². A major limitation in this field is the absence of comprehensive and publicly available Arabic datasets. In particular, the scarcity of labeled benchmark datasets remains a key challenge in advancing Arabic deepfake detection research³. Existing research has mainly relied on English datasets or domains such as news articles, e-commerce reviews, and social media platforms. For example, Su et al. introduced the HC3 dataset, which contains both human-written and ChatGPT-generated English content and has been widely used in deepfake detection studies⁴. Nevertheless, Arabic remains a relatively low-resource language in this field. Zita et al. reported that, despite being one of the six official United Nations languages, Arabic has received limited attention in AI-generated

text detection research⁵. Similarly, a survey by Seyam et al. found that approximately 91% of fake news detection studies focus on English content⁴. In addition, prior studies on Arabic deepfake detection have often constructed small synthetic datasets or translated English datasets into Arabic. For instance, Barhoum et al. created a fake news dataset using authentic data collected from Kaggle combined with synthetically generated Arabic text⁶. Likewise, Alghamdi et al. investigated fake reviews on Arabic e-commerce platforms, although the original dataset was initially developed in English and later translated into Arabic⁷.

Another important limitation in the current literature is the limited exploration of streaming platforms, particularly YouTube. Unlike traditional sources such as news or reviews, YouTube provides rich user-generated content that reflects informal language and diverse Arabic dialects. Although one Arabic YouTube dataset was introduced by Alahamdi⁸, the data collection methodology was not clearly documented, especially regarding the tools and application programming interfaces (APIs) used. This lack of transparency makes it difficult to verify the dataset's reliability and reproducibility. Previous studies have shown that the YouTube API is an effective tool for collecting large-scale datasets. For example, Cruickshank et al. demonstrated its usefulness for gathering authentic user-generated content⁹, while Siersdorfer et al. collected more than six million comments from over 67,000 YouTube videos to analyze user

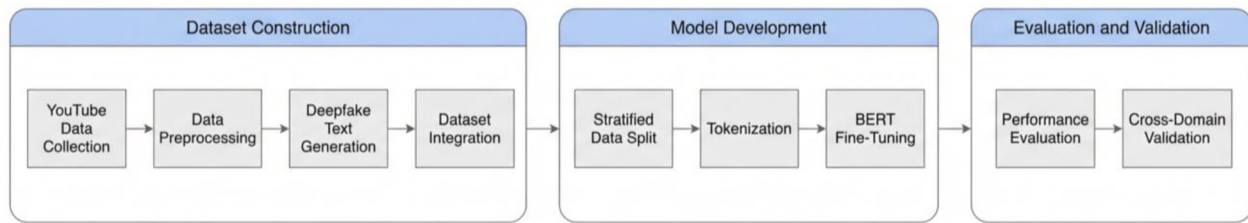


Figure 1. Workflow diagram

engagement¹⁰. However, these datasets focused exclusively on English content.

Overall, the literature highlights a clear gap: there is currently no verified, domain-specific, or publicly documented Arabic dataset derived from YouTube that supports deepfake text detection research. To address this limitation, this study aims to construct a robust Arabic dataset for deepfake text detection based on YouTube comments. The proposed dataset encompasses multiple Arabic dialects, including Modern Standard Arabic (MSA), Saudi, Egyptian, Kuwaiti, and Iraqi, as well as content from several domains such as sports, religion, entertainment, and health.

The research was conducted in three stages, as illustrated in Figure 1. The first stage involved dataset construction, including YouTube data collection, preprocessing, deepfake text generation, and dataset integration. The second stage focused on model development, including stratified data splitting, tokenization, and fine-tuning of the Bidirectional Encoder Representations from Transformers (BERT) model. The final stage involved model evaluation and validation using accuracy, precision, recall, F1-score, and a confusion matrix, along with cross-domain testing to assess the model's generalization capability.

This study contributes to the field by introducing a comprehensive Arabic YouTube-based dataset for deepfake text detection and demonstrating its effectiveness through cross-domain evaluation. The dataset and experimental framework provide a reliable foundation for future research in Arabic deepfake detection. Furthermore, this paper presents a detailed analysis of dataset construction and validation, including data collection, synthetic data generation, and model training and evaluation procedures.

2. METHODOLOGY

2.1. Data Collection and Preprocessing. The first stage of this research involved the collection and preprocessing of real Arabic data. YouTube was selected as the primary streaming

platform. Four domains were chosen before gathering the data—Health, Sports, Religion, and Entertainment. These domains were chosen because they represent popular topics among Arabic-speaking users and capture diverse dialectal expressions. The data were collected using a Python script executed in a Google Colab environment. The script utilizes YouTube API to ensure efficient, authentic, and user-generated content. YouTube video IDs were manually collected and saved in a list. The script utilized the list of IDs and collected the metadata from each video, including comments, timestamps, author names, and other relevant attributes. After executing the script, 14,854 Arabic comments were collected from the selected fields. As shown in Figure 2, the distribution between the fields was balanced. Next, the dataset was cleaned to ensure data accuracy and reliability. Duplicate, empty or meaningless, and inappropriate comments were removed from the dataset. Following data cleaning, the comments were manually reviewed by a domain expert and classified into the five Arabic dialects that best represented the content—MSA, Saudi, Iraqi, Egyptian, and Kuwaiti. The cleaned dataset was then organized and labeled based on these dialects, as shown in Figure 3. The distribution of dialects was not balanced and was biased towards MSA. This is because most YouTube comments do not reflect dialects.

2.2. Deepfake Text Generation. After collecting and cleaning the real dataset, synthetic Arabic deepfake text was generated using OpenAI GPT-4.0 Mini. Each deepfake instance was produced from a carefully designed prompt that generated a synthetic version of every real-text sample while preserving the assigned dialect.

Because the dataset was collected from YouTube, the generation process incorporated deception techniques commonly observed on that platform, including exaggeration, omission, contradiction, satirical tone, clickbait phrasing, and mixed truths. For each real text instance, one of these predefined techniques was randomly assigned while maintaining the corresponding dialect.

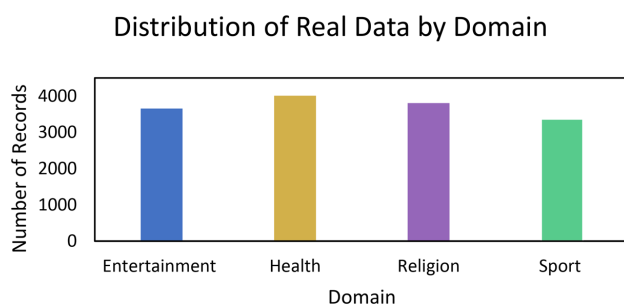


Figure 2. Number of real data records per domain

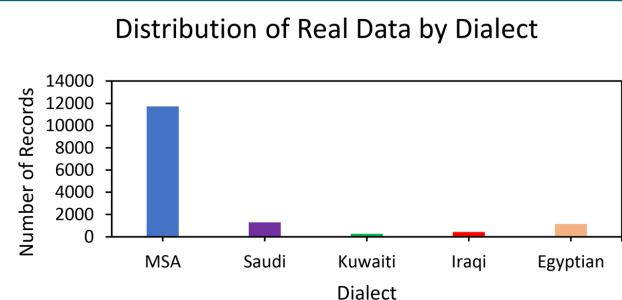


Figure 3. Number of real data records per dialect

Table 1. Distribution of real and synthetic text samples across Arabic dialects

Dialect	Real Samples	Fake Samples	Total
MSA	11739	11727	23466
Saudi	1281	1283	2564
Kuwaiti	272	274	546
Iraqi	424	425	849
Egyptian	1137	1146	2283

The prompt was designed to ensure that the generated text followed the selected dialect and produced a natural and conversational deceptive rewrite consistent with user-generated online discourse. To enhance realism, the generated text was required to sound human-like and reflect the informal language typically used in online discussions. Overly formal or highly structured language was discouraged, and minor imperfections or repetitions were permitted when they contributed to more authentic outputs. In addition, the generation process prevented verbatim copying of the original text to ensure that synthetic samples remained distinct from authentic comments. The model was also constrained to produce only the rewritten deceptive text without adding explanations, introductions, or meta-commentary.

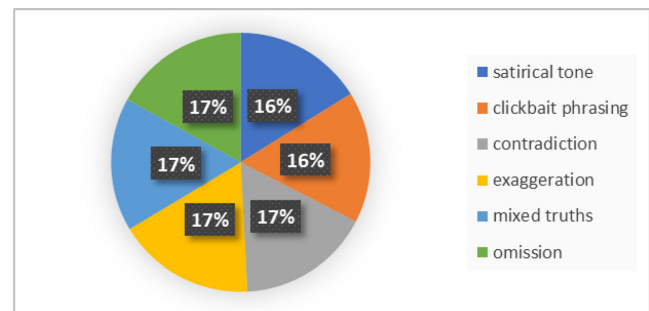
To automate this process, a Python script was developed and executed in the Google Colab environment. The script integrated the OpenAI API and processed the cleaned real dataset by injecting each sample into the prompt to generate its synthetic counterpart. This systematic design ensured consistent application of deception techniques and controlled generation across the dataset.

After executing the script, a synthetic deepfake dataset containing 14,857 rows was generated. Each entry included the original real text, its dialect, the applied deception technique, and the generated deepfake text. Figure 4 illustrates the distribution of deception techniques across the generated deepfake samples, which follows a balanced distribution. As shown in Figure 5, Modern Standard Arabic (MSA) was the most frequently occurring dialect.

After verifying the outputs of the generation process and ensuring dataset consistency, the real and deepfake datasets were merged and randomized. The resulting dataset contains several attributes, including a text column and a label column indicating whether a sample is real or fake. Additional columns include the dialect column (MSA, Saudi, Iraqi, Kuwaiti, and Egyptian), the field column (health, sports, entertainment, and religion), and a deception column specifying the deception technique used to generate fake samples. Table 1 presents a statistical summary of the dataset, showing the distribution of real and synthetic text samples across different Arabic dialects.

3. EXPERIMENTAL WORK

3.1. Model Architecture and Hyperparameters. This study employed a transformer-based architecture, specifically BERT, because of its strong capability to capture contextual and semantic relationships in text. Additionally, BERT has been

**Figure 4.** Distribution of deception techniques in the generated deepfake dataset

widely used in text-classification tasks, making it suitable for distinguishing between real and fake content¹¹.

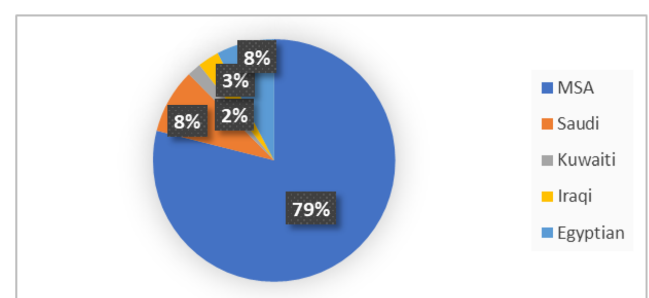
As illustrated in Figure 6, each Arabic text is first tokenized and augmented with special tokens [CLS] at the beginning and [SEP] at the end. The tokenized sequence is converted into embedding text and passed through 12 stacked transformer encoder layers. These layers utilize self-attention mechanisms to capture the bidirectional contextual relationships between words. After processing the entire sequence, the final hidden representation of the [CLS] token is used as the sentence-level representation. This representation is fed into a linear classification layer and is followed by a softmax function to predict whether the input text is real or fake¹². The entire model is fine-tuned using the constructed labeled dataset, which includes authentic Arabic YouTube comments and synthetically generated deepfake text.

Because BERT is a pre-trained model, it learns general language patterns, allowing effective fine-tuning with a limited number of training epochs. The model was configured using the following hyperparameters— a learning rate of 2e-5, batch size of 32, three training epochs, and a maximum sequence length of 512.

The maximum sequence length was set to 512, which is the maximum input length supported by BERT¹³. A learning rate of 2e-5 was selected to reduce the risk of catastrophic forgetting during fine-tuning¹³. For classification tasks, three to five epochs are commonly used; therefore, three epochs were chosen for the experiment¹⁴.

3.2. Training Environment and Evaluation Metrics.

The model was trained using an A100 GPU on the Google Colab platform, instead of the default CPU accelerator. GPUs are more efficient for machine-learning applications because they process data in parallel, whereas CPUs process data sequentially¹⁵. Given

**Figure 5.** Distribution of dialect in the generated deepfake dataset

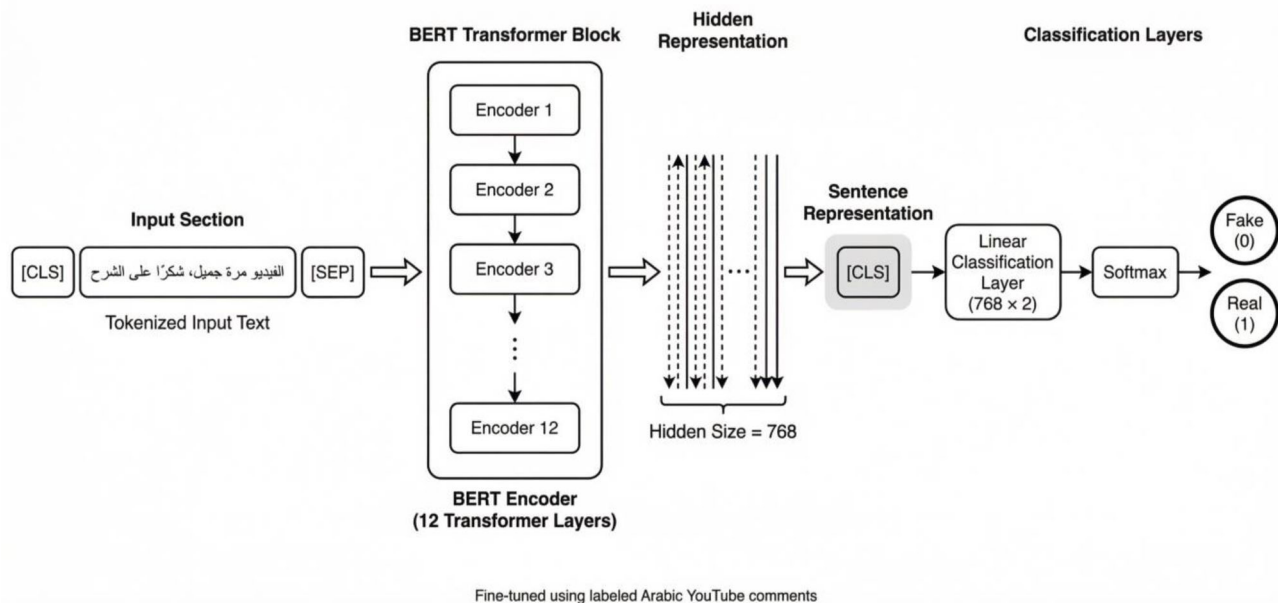


Figure 6. BERT-based architecture for Arabic deepfake text classification

the large dataset used in this study, GPU training significantly improved computational efficiency.

To prevent overfitting and ensure optimal model selection, the parameter “load_best_model_at_end” was set to True. This approach is recommended because the validation loss does not necessarily decrease in every epoch and may increase after a certain point, indicating that the model has stopped improving and started to overfit¹⁴. Therefore, the model corresponding to the epoch with the lowest validation loss was automatically saved and used in the final evaluation.

The model performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrices. To provide an overall evaluation of the model, the reported precision, recall, and F1-score values represent weighted averages across both labels (Real and Fake), where each label is weighted according to the number of samples belonging to it.

- Accuracy is computed as the ratio of correctly predicted samples to the total number of samples.
- The confusion matrix includes four elements: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).
 - TP represents real texts correctly identified as real.
 - TN represents fake texts correctly identified as fake.
 - FP represents fake texts incorrectly predicted as real.
 - FN represents real texts incorrectly predicted as fake.
- Precision was calculated as $TP / (TP + FP)$.
- Recall was calculated as $TP / (TP + FN)$.
- The F1-score was computed using the harmonic mean of the precision and recall $2 / ((1/Precision) + (1/Recall))$ ¹⁶.

3.3. Data Preparation and Training Procedure.

Before training, the dataset was normalized and preprocessed to ensure robust performance. The labels “Real” and “Fake” were mapped to binary values 1 and 0, respectively. Rows containing missing text or labels were excluded.

The dataset was divided into training (70%), validation (15%), and testing (15%) subsets. The training set was used to learn linguistic patterns, the validation set to monitor the performance and prevent overfitting, and the testing set to evaluate the model on unseen data¹⁷.

To enable the BERT model to distinguish between “Real” and “Fake” Arabic text, it was trained to identify the contextual and linguistic patterns that differentiate human-written content from text generated by OpenAI GPT-4.0 Mini. The model was implemented using the Hugging-Face Transformers library, which provides access to pretrained language models, such as BERT¹⁸.

As BERT cannot process raw text directly, a tokenizer was applied to convert Arabic sentences into numerical representations. Each sentence was segmented into words and mapped to the corresponding vocabulary IDs according to the BERT tokenizer¹⁸. The pretrained BERT model was then loaded and configured using the selected hyperparameters described in Section 3.1.

Training was conducted for three epochs, during which the model iteratively processed the input text, generated predictions, and compared them with the True labels to compute the loss values. After each epoch, the model was evaluated using the validation set, and the performance metrics, including training loss, validation loss, accuracy, precision, recall, and F1-score, were recorded.

Upon completion of the training, the final model was evaluated on a test set containing unseen data to assess its generalization performance. Through this training process, the model learned to capture semantic and contextual patterns in Arabic text, enabling effective discrimination between real human-generated content and fake text generated using OpenAI GPT-4.0 Mini.

4. RESULTS AND DISCUSSION

4.1. Training Result. As discussed in the previous section, the training process was performed over three epochs, and after

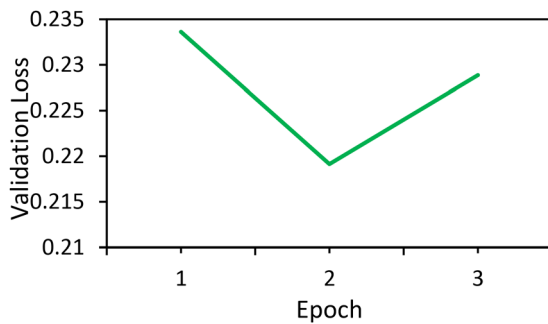


Figure 7. Learning curve-model validation loss per epoch

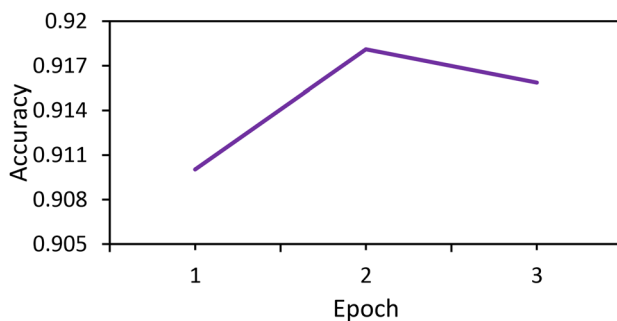


Figure 8. Learning curve-model accuracy per epoch

each epoch, the model used the validation set to assess its progress. Thus, a set of statistics was calculated for each epoch, as shown in Table 2. The validation loss decreased in each epoch but increased slightly in the third epoch, as shown in Figure 7. However, accuracy, precision, recall, and F1 all increased in each epoch, with a slight decrease in the third epoch. Figure 8 presents the change in model accuracy per epoch. This indicates that the model achieved its best performance and learned the most during the second epoch. Therefore, further training was not expected to improve the model. Thus, the model parameters from the second epoch were saved and used as the final model checkpoints.

4.2. Testing Results. After evaluating the model on the unseen test set, the evaluation loss was approximately 0.2267, and the accuracy reached 91.43%, as listed in Table 3. The test results imply that the model has a balanced and strong performance in distinguishing between real and fake text. The confusion matrix in Figure 9 shows that the model correctly identified 2,112 real texts and 1,980 fake texts, while misclassifying 116 real labels as fake and 249 fake labels as real. These results support the conclusion that the model provides reliable predictions, demonstrates good generalization capabilities, and does not exhibit signs of overfitting.

4.3. Cross-domain Evaluation Results. To assess the ability of the model to generalize across different datasets and domains, a cross-domain evaluation was conducted using four

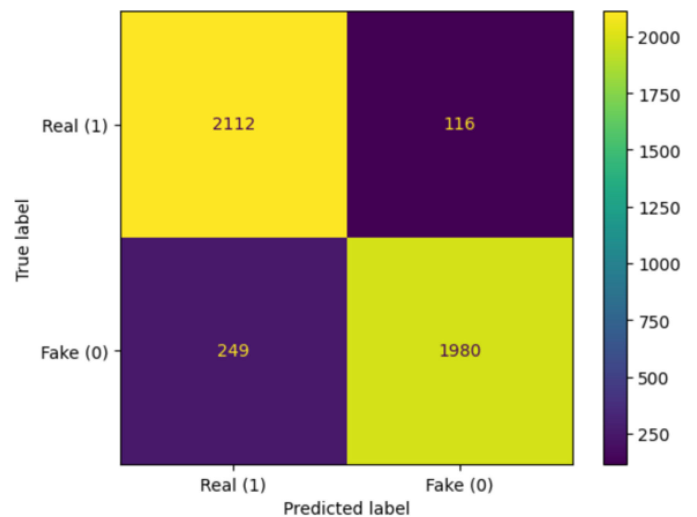


Figure 9. Confusion matrix of final model on the test set

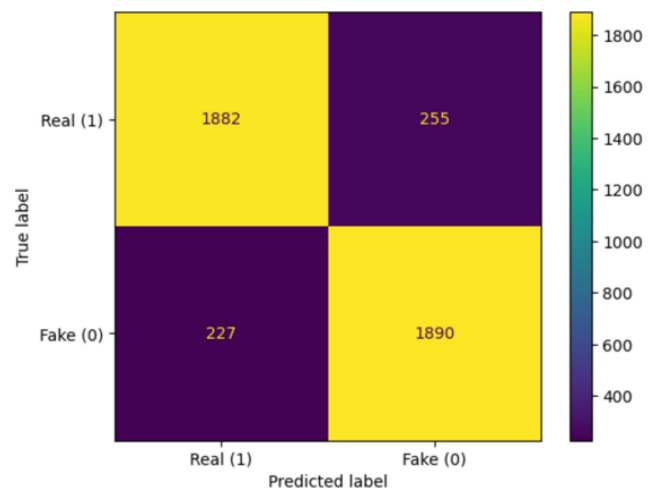


Figure 10. Confusion matrix – social media dataset

independent Arabic datasets: social media, e-commerce, news, and governmental.

A. Social Media Dataset

The first dataset contained user-generated content from social media platforms, including comments, tweets, and replies. The dataset included five dialects: Egyptian, Iraqi, Shami, Sudanese, and Modern Standard Arabic. With this dataset, the model achieved an accuracy of 88.66%, as shown in Table 4.

The confusion matrix in Figure 10 shows that the model correctly predicted 1,882 real and 1,890 fake texts. However, the model misclassified 255 real samples as fake and 227 fake samples as real.

B. E-Commerce Dataset

Table 2. Training and validation loss with corresponding performance metrics across epochs

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
1	0.242700	0.233621	0.910029	0.912076	0.910029	0.909918
2	0.216400	0.219132	0.918106	0.919602	0.918106	0.918034
3	0.210100	0.228866	0.915863	0.916081	0.915863	0.915852

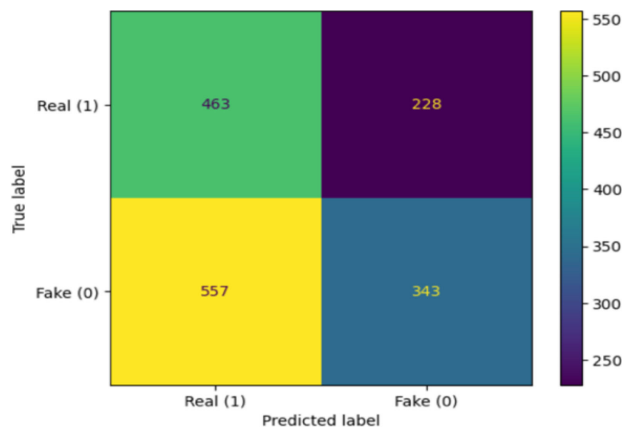


Figure 11. Confusion matrix - E-commerce dataset

The second dataset focused on the electronic domain, specifically smartphones and mobile devices. It contained descriptions of smartphones, including details such as storage capacity, battery size, camera specifications, display type, and connectivity features. This dataset contained only MSA text. The model achieved an accuracy of approximately 50.66%, as shown in Table 5. The model correctly classified 463 real texts and 343 fake texts, while incorrectly predicting 228 real samples as fake and 557 fake samples as real, as shown in the confusion matrix in Figure 11.

Table 3. Test set performance result of the final model

Metrics	Value
Evaluation Loss	0.22
Evaluation Accuracy	0.91
Evaluation Precision	0.91
Evaluation Recall	0.91
Evaluation F1 Score	0.91
Evaluation Runtime	28.87
Evaluation Samples / s	154.30
Evaluation Steps / s	4.84
Epochs	3

Table 4. Test results on the social media dataset

Metrics	Value
Accuracy	0.88
Precision	0.88
Recall	0.88
F1 Score	0.88

Table 5. Test results on the e-commerce dataset

Metrics	Value
Accuracy	0.50
Precision	0.53
Recall	0.50
F1 Score	0.49

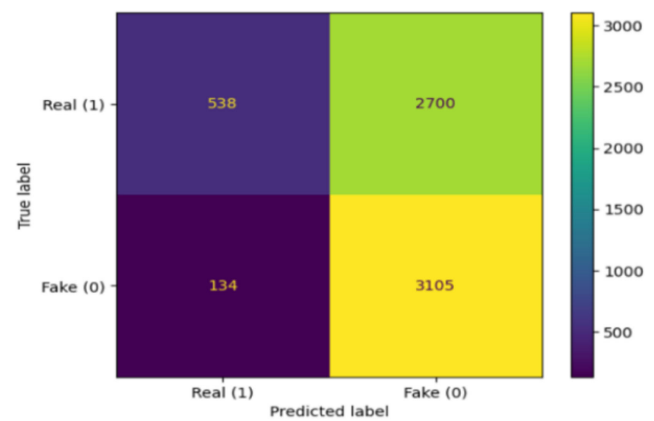


Figure 12. Confusion matrix-news dataset

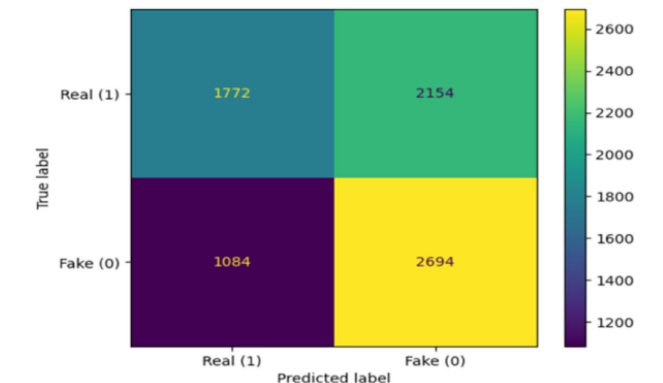


Figure 13. Confusion matrix - governmental dataset

C. News Dataset

The third dataset consisted of Arabic news covering politics, health, economy, and technology, and contained only MSA text. Table 6 indicates that the test results had an accuracy of approximately 56.24%. Figure 12 shows the confusion matrix of the model, which indicates that it could correctly predict 538 real texts and 3105 fake texts, while it predicted 2,700 real labels as fake and 134 fake labels as real.

D. Governmental Dataset

The final dataset included official Arabic texts from government sources such as the Ministries of Health, Education, and Justice. Thus, all texts were written in MSA. After testing the model on this dataset, its accuracy was approximately 57.97%, as shown in Table 7. The model identified 1,772 real texts and 2,694 fake texts. However, the model misclassified 2,154 real samples as fake and 1,084 fake samples as real, as shown in Figure 13.

Table 6. Test results on the news dataset

Metrics	Value
Accuracy	0.56
Precision	0.66
Recall	0.56
F1 Score	0.48

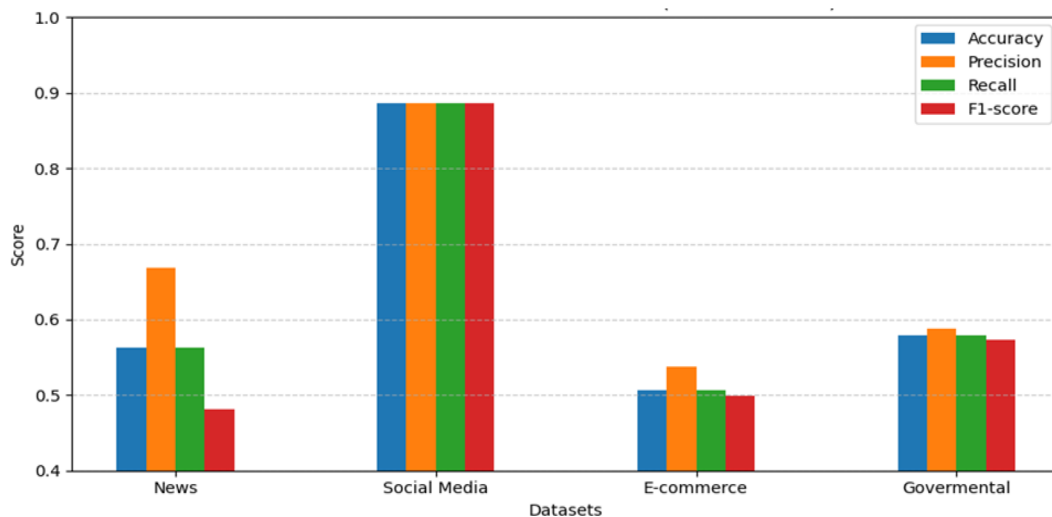


Figure 14. Model performance across domains

Table 7. Test results on the governmental dataset

Metrics	Value
Accuracy	0.57
Precision	0.58
Recall	0.57
F1 Score	0.57

Testing the model on different datasets of various domains and dialects revealed that the model demonstrated the highest accuracy, 88.66% when tested on the social media dataset, while the lowest accuracy was demonstrated with the e-commerce dataset (50.66%), as illustrated in Figure 14. As the social media domain closely resembles the YouTube domain, with similar linguistic and contextual patterns, the model maintained high performance and yielded highly accurate results. However, other news, government, and e-commerce datasets demonstrated more formal and distinct patterns. Thus, the accuracy of these sets ranged from approximately 50–58%. Therefore, the model could distinguish between real and fake text across different datasets. However, the closer the contextual and linguistic patterns are to YouTube, the higher the prediction accuracy.

5. CONCLUSION

This study aimed to build a comprehensive Arabic dataset and validate it by developing a model to detect Arabic deepfake text. To ensure dataset robustness, a collection of authentic Arabic text was gathered, cleaned, and complemented with synthetically generated deepfake text. The BERT model was then trained, validated, and tested, achieving an accuracy of approximately 91.43%. The results demonstrate the balanced and robust performance of the proposed model in detecting deepfake text in the constructed dataset. A cross-domain validation step was performed using external datasets to evaluate the model's ability to generalize and avoid overfitting. The constructed dataset and detection framework have important practical implications. The dataset can serve as a benchmark resource for future Arabic deepfake detection research and support the

development of automated moderation tools for Arabic online platforms. In addition, the proposed approach can assist social media platforms, streaming services, and content moderation systems in identifying AI-generated misleading text and reducing the spread of deceptive information in Arabic digital environments. Although the experimental results of this study are promising, there are several limitations. Specifically, the dataset size and diversity were limited, and only a subset of Arabic dialects and topics was considered. These factors may have affected the overall performance of the model. Future studies should consider larger datasets, more dialects, and a broader variety of fields to obtain stronger models and improved performance.

6. DECLARATION

Grammarly, QuillBot, and ChatGPT were used for grammar, spelling, and sentence structure corrections.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Malath Alhashem – Department of Information and Computer Science, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; s202174410@kfupm.edu.sa; ORCID: 0009-0001-3684-9319

Author

Amal Sunba – Research Collaborator, Department of Information and Computer Science, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; g202203780@kfupm.edu.sa; ORCID: 0009-0002-4712-2104

Corresponding Author

Tarek Helmy – Research Mentor, Professor, Department of Information and Computer Science, Interdisciplinary Research Center of Intelligent Secure Systems, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; helmy@kfupm.edu.sa; ORCID: 0000-0001-9683-682X

ACKNOWLEDGEMENTS

This research was conducted in collaboration between the General Authority for Defence Development (GAAD) and King Fahd University of Petroleum & Minerals (KFUPM) under agreement number CSSP2643, Saudi Arabia. The authors would like to acknowledge the General Authority for Defense Development (GADD) in Saudi Arabia for funding this research through project number (GADD_2024_01_438). The authors would like to acknowledge the support of the Department of Information and Computer Science and the Interdisciplinary Research Center of Intelligent Secure Systems at King Fahd University of Petroleum and Minerals

REFERENCES

- (1) Patel, T. & Kantawala, H. Psychological impacts of deepfakes: understanding the effects on human perception, cognition, and behavior. *Int. J. Digit. Soc.* **16**, 1–17 (2025).
- (2) Boutadjine, A. Fighting Deepfake Text: Towards Building Robust Defences against Arabic AI-generated Text (PhD thesis, Ferhat Abbas Sétif 1 University, 2025).
- (3) Seyam, A., Bou Nassif, A., Abu Talib, M., Nasir, Q. & Al Blooshi, B. Deep learning models to detect online false information: a systematic literature review. In *Proceedings of the 7th Annual International Conference on Arab Women in Computing*, 1–5 (ACM, 2021). <https://doi.org/10.1145/3485557.3485580>
- (4) Su, Z., Wu, X., Zhou, W., Ma, G. & Hu, S. HC3 Plus: a semantic-invariant human ChatGPT comparison corpus. Preprint at arXiv <https://doi.org/10.48550/arXiv.2309.02731> (2023).
- (5) Zita, K. et al. LMSA at AraGenEval Shared Task: Ensemble-Based Detection of AI-Generated Arabic Text Using Multilingual and Arabic-Specific Models. *Association for Computational Linguistics*, 26–31, (2025).
- (6) Barhoum, T., Ibrahim, M. & Al Bali, M. Detecting machine-generated Arabic text using AraBERT and LSTM: toward trustworthy NLP in low-resource languages. *Research Square* (2025).
- (7) Alghamdi, M., Siddiq, I., Alharthi, A. Detecting Fake Arabic Reviews in E-commerce using Machine and Deep Learning Approaches. *J Inform Secure Cybercrimes Res (JISCR)* **5**(1):32–45 (2022).
- (8) Alahmadi, F. Arabic YouTube comments by Khalaya. *Kaggle* <https://www.kaggle.com/datasets/farisalahmadi/arabic-youtube-comments-by-khalaya> (2023).
- (9) Cruickshank, I. J., Ng, L. H. X. & Soofi, A. DIVERSE: a dataset of YouTube video comment stances with a data programming model. in *Proc. IEEE International Conference on Big Data (IEEE Big Data 2024)* 2080–2089 (IEEE, 2024). <https://doi.org/10.1109/BIGDATA62323.2024.10825480>
- (10) Siersdorfer, S., Chelaru, J., Nejd, W., Pedro, S.: How useful are your comments? Analyzing and predicting YouTube comments and comment ratings. In: *Proceedings of the International Conference on World Wide Web (WWW)*, pp. 891–900 (2010).
- (11) Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. in *Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)* Vol. 1, 4171–4186 (Association for Computational Linguistics, Minneapolis, MN, 2019). <https://doi.org/10.18653/v1/N19-1423>
- (12) Usuga-Cadavid, J. P., Lamouri, S., Grabot, B. & Fortin, A. Using deep learning to value free-form text data for predictive maintenance. *Int. J. Prod. Res.* **60**, 4548–4575 (2022). <https://doi.org/10.1080/00207543.2021.1951868>
- (13) Sun, C., Qiu, X., Xu, Y. & Huang, X. How to fine-tune BERT for text classification? in *Chinese Computational Linguistics — 18th China National Conference, CCL 2019* (eds Sun, M., Huang, X., Ji, H., Liu, Z. & Liu, Y.) *Lect. Notes Comput. Sci.* Vol. 11856, 194–206 (Springer, Cham, 2019). https://doi.org/10.1007/978-3-030-32381-3_16
- (14) Nebius. How many epochs are enough to train your model. *Nebius Blog* <https://nebius.com/blog/posts/factors-influencing-epoch-count> (2025).
- (15) Schneider, J. & Smalley, I. CPU vs GPU for Machine Learning. *IBM Think*. <https://www.ibm.com/think/topics/cpu-vs-gpu-machine-learning> (2025).
- (16) Rainio, O., Teuhio, J. & Klén, R. Evaluation metrics and statistical tests for machine learning. *Sci. Rep.* **14**, 6086 (2024). <https://doi.org/10.1038/s41598-024-56706-x>
- (17) Mistry, R. Data Splitting (Train-Test-Validation) in Machine Learning. *Medium*. <https://medium.com/@rohanmistry231/data-splitting-train-test-validation-in-machine-learning-2d5d1927fa69> (2025).
- (18) Hugging Face. The Hugging Face LLM Course. *Hugging Face Learn*. <https://huggingface.co/learn/llm-course> (2025).

Tumor Classification on the Bone Tumor X-ray Radiograph Dataset: Cross-Validation, Contrast-Limited Adaptive Histogram Equalization Ablation, and Explainability Assessment

Saad Bin Waqas and Muhammad Shahid Jabbar*

Cite <https://doi.org/10.64589/juri/218960>

Submitted: November 27, 2025 Revised: January 13, 2026 Accepted: January 28, 2026

ABSTRACT

Radiograph-facilitated triage of primary bone tumors is critical for timely referral; however, it remains difficult in settings without subspecialty expertise. Building on the multicenter Bone Tumor X-ray Radiograph Dataset (BTXRD), in this study, we aimed to systematically benchmark modern deep learning architectures for classifying normal bone, benign tumors, and malignant tumors using a unified, reproducible pipeline. We evaluated a broad family of compound-scaled Convolutional Neural Networks (CNNs) and vision transformers using standardized preprocessing, class-imbalance-aware training via effective number weighting, and variance-aware k-fold cross-validation. Furthermore, we compared their performance with the BTXRD YOLOv8 Classification baseline. In addition to backbone evaluation, we conducted an architecture-aware analysis of contrast-limited adaptive histogram equalization (CLAHE), quantifying its effects on classification performance and learned representation geometry. To characterize representation shifts, we embedded penultimate-layer features using Uniform Manifold Approximation and Projection (UMAP) and determined how CLAHE alters neighborhood structure and class separability across backbones. We also integrated High Resolution Class Activation Mapping (HiResCAM) to assess explanation faithfulness and elucidated the degree of spatial overlap between model attention and radiologist-provided lesion masks. The experimental protocol incorporated a frozen held-out test set, calibration-sensitive metrics, and per-image prediction auditing to reflect deployment-oriented behavior. Across architectures, EfficientNet-B6 with CLAHE was the most reliable configuration, delivering high accuracy and macro-F1, while maintaining clinically meaningful error patterns and precise malignant predictions. This work provides novel insights into how preprocessing, architectures, feature representations, and explanation methods interact in bone tumor X-ray classification and establishes reproducible baselines to support future research on BTXRD.

Keywords: bone tumor, X-ray radiography, classification, deep learning, explainable AI, uniform manifold approximation and projection, class activation mapping

1. INTRODUCTION

Despite being uncommon, bone tumors are important radiographic findings and remain the most accessible and cost-effective imaging modality for musculoskeletal care. However, reliable differentiation among normal bone and benign and malignant tumors can be challenging, particularly in settings with limited subspecialty expertise. Recent advances in deep learning have demonstrated strong performance in medical image interpretation and may assist in radiographic assessment^{1,2}. A growing body of medical imaging research has proposed task-driven classification pipelines that combine architectural enhancements and hybrid feature representations with explicit structural cues and, in some cases, segmentation or localization components. Representative examples include glaucoma screening frameworks that integrate segmentation-derived anatomical measurements and gastric histopathology studies that explore hybrid deep feature

extraction coupled with Vision Transformer models^{3,4}. These studies address their respective tasks through tailored modeling but do not specifically examine backbone-level performance trends under a common, controlled evaluation framework.

To systematically assess the behavior of widely used CNN and transformer backbones in a radiographic setting under consistent training and reporting conditions, this study benchmarks a diverse set of architectures on the Bone Tumor X-ray Radiograph Dataset (BTXRD) using a unified protocol, emphasizing cross-validated evaluation, imbalance-aware reporting, and mask-referenced localization analysis.

Prior studies have explored the automatic detection and categorization of bone lesions on X-rays, although many have evaluated a narrow set of architectures or treated models as “black boxes,” offering limited insight into decision mechanisms⁵. The recently released BTXRD dataset provides a timely testbed by pairing radiographs with expert labels and radiologist-provided

segmentation masks, enabling both classification and spatial localization⁶. The original BTXRD study and subsequent follow-ups using YOLO-family models reported strong baselines for classification⁷. One follow-up study applied SMOTE-based over-sampling to mitigate class imbalance⁸, a technique that may over-estimate minority-class performance when evaluated on a single split and can reduce generalization reliability in medical imaging⁹. Broader issues, such as class-imbalance strategies, model calibration, and explainability analyses, have received less emphasis, and evaluations commonly rely on a single train/test split without a dedicated validation set for principled model selection.

In this study, we addressed these gaps by benchmarking a diverse suite of CNN and transformer backbones using BTXRD within a unified, reproducible PyTorch pipeline, evaluated convolutional and transformer-based models using the original training/test split for comparability, and introduced an internal validation protocol to enable robust tuning. To enhance interpretability and clinical trust, we integrated class-activation mapping methods and used expert masks to assess the spatial alignment between highlighted regions and tumor locations.

We also conducted targeted ablation of contrast-limited adaptive histogram equalization (CLAHE), which has been included in at least one BTXRD preprocessing pipeline without a specific assessment of its impact⁷. CLAHE is widely used in related musculoskeletal and chest X-ray classification tasks with mixed effects on performance^{10,11}. Given that its effect on bone tumor classification has not previously been elucidated, we explicitly quantified its impact in this study. Performance was evaluated using metrics appropriate for class imbalance and clinical decision support, including accuracy, macro-F1, and per-class metrics. Previous BTXRD studies primarily reported improvements in single-split accuracy or task-specific architectural modifications; this study explicitly investigated how preprocessing choices, architecture families, representation geometry, and explanatory behavior interact under variance-aware evaluation. By combining paired statistical testing, representation-shift analysis, and failure-mode localization assessment within a unified pipeline, this study contributes beyond leaderboard performance and provides insights into the conditions under which specific design choices are reliable. The primary contributions of this study are as follows:

- A reproducible benchmark for evaluating contemporary CNN and transformer backbones for three-class BTXRD classification.
- A stratified k-fold cross-validation framework to address class imbalance, supported by model calibration analyses.
- The first architecture-aware CLAHE ablation for BTXRD was complemented by a UMAP-based representation-space visualization and quantification of feature-space shifts.
- A HiResCAM-based localization assessment that measures the concordance between model attention and expert masks, informing the interpretability of the resulting classifiers.

2. METHODOLOGY

2.1. Dataset. We used the publicly available BTXRD⁶. The dataset comprises 3,746 primary bone tumor images, with 1,879 normal, 1,525 benign, and 342 malignant cases. To ensure

comparability with prior studies, we adopted a predefined training/test split provided in the official release. Classification is defined as a 3-class problem (no tumor, benign, or malignant).

2.2. Model Architectures. We benchmarked convolutional- and transformer-based models in Torchvision, including ResNet-152, EfficientNet (B5–B7, V2-M, V2-L), ConvNeXt-Large, MobileNetV3-Large, and vision transformers (Swin V2-B, MaxViT-T)^{12–17}. The final classification head was replaced with a randomly initialized 3-way linear layer, whereas the earlier layers were fine-tuned end-to-end.

2.3. Preprocessing and Augmentations. The radiographs were loaded in grayscale and expanded to three channels to be compatible with ImageNet backbones. We applied an Albumentations pipeline¹⁸ consisting of random horizontal flips ($p = 0.5$), affine transformations ($\pm 10^\circ$ rotation, $\pm 5\%$ scaling, and translation), Gaussian blur ($p = 0.3$), random gamma correction ($p = 0.5$), and elastic deformation with ($\alpha = 2$, $\sigma = 8$, $p = 0.2$). For CLAHE, we used an 8×8 tile grid and a moderate clip limit (2.5), consistent with common defaults and typical medical imaging practices, to enhance local contrast while limiting noise amplification and artifact formation. The models were trained at their native resolution with an additional nearest-32 variant for architectures whose native size was not divisible by 32. Better-performing settings were also retained.

2.4. Training Setup. The models were trained using the AdamW optimizer¹⁹, ReduceLROnPlateau, early stopping, and automatic mixed precision. The batch size was 16 and was reduced to 8 for EfficientNet-B7 owing to memory limitations. We aimed to improve reproducibility by fixing random seeds across PyTorch, NumPy, and Python, thereby enabling deterministic cuDNN behavior (deterministic = true, benchmark = false). The hyperparameters were selected in accordance with established best practices for ImageNet-pretrained medical imaging models and were held constant across architectures to ensure a controlled comparison; no exhaustive hyperparameter search was performed. This design choice trades the absolute peak performance for reproducibility and comparability across backbones. The exact training times were not logged for the individual runs during the main experiments. For the final configuration, training took approximately 7.9 h for a full 64-epoch run on an NVIDIA T4 GPU accelerator.

2.5. Phase A: Candidate-Model Screening on Authors' Split (Baseline). Phase A provides a capacity check, ensuring that any metric changes in Phase B can be attributed to the evaluation protocol rather than to model performance. We computed accuracy and per-class metrics across 5 seeds (mean $\pm 95\%$ Confidence Interval [CI]) and ranked models by validation accuracy to select candidates for Phase B. The YOLO baseline reflects the authors' published results rather than those of the retrained model.

2.6. Phase B: K-Fold Cross-Validation with CLAHE Ablation. Phase B used five-fold stratified cross-validation to establish a variance-aware estimate of model generalization and to quantify the effect of contrast-limited adaptive histogram equalization (CLAHE) under controlled conditions.

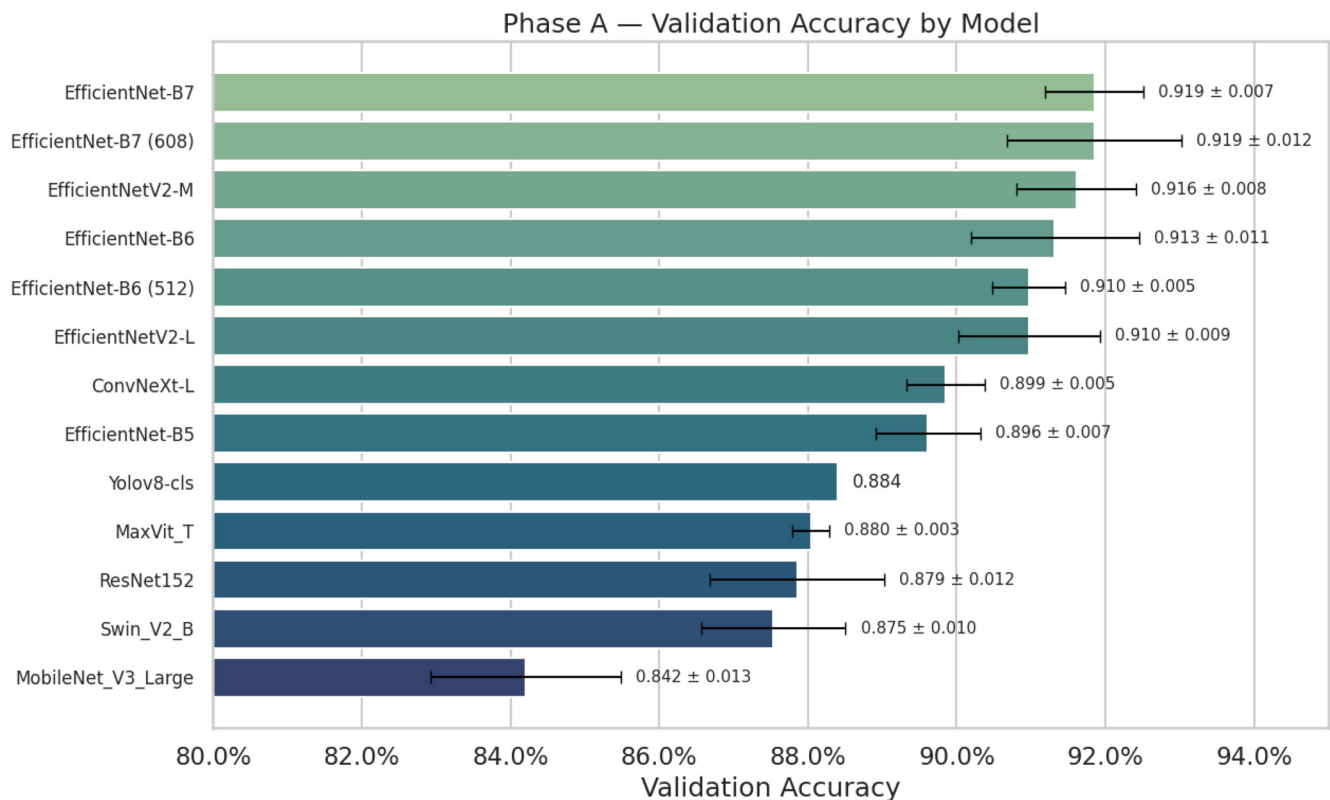


Figure 1. Phase-A comparative classification accuracies on the Bone Tumor X-ray Radiograph Dataset (BTXRD)

The training split was partitioned into five folds, and the original validation set was maintained as a held-out test set. Model checkpoints were selected by validating macro-F1 to prioritize balanced per-class performance.

Each architecture was trained twice per fold, once using the baseline pipeline and once using CLAHE, to produce fold-wise paired comparisons that sought to isolate the effect of contrast enhancement. The three best-performing architectures from Phase A were selected for Phase B, which belonged to the EfficientNet family, as discussed in Section 3. Each candidate model was retrained using three independent seeds to reduce the variance. To determine whether CLAHE's behavior generalized beyond the best models, ConvNeXt-Large and Swin V2-B were also included with just one seed per fold, as the goal was architectural comparison rather than hyperparameter tuning.

To mitigate class imbalance in the malignant distribution, we employed Effective-Number class-balanced cross-entropy²⁰ with β as 0.995, normalizing per-class weights to unit mean via $w_c \propto 1/(1 - \beta^{n_c})$, where n_c is the number of samples in class c and β controls the rate of diminishing returns as class size grows. For every run, we tracked the best-F1, best-accuracy, confusion matrices, and per-class CSVs with Wilson 95% CIs. Seed-level metrics were averaged within each fold, then summarized across folds (mean $\pm$ 95% t-distribution Confidence Interval [t-CI]). The CLAHE ablation is reported as paired deltas (+CLAHE – Baseline) in macro-F1 and accuracy.

2.7. Representation-Shift Visualization with UMAP.

To examine how CLAHE alters internal representations learned

by different architectures, we visualized the penultimate-layer feature embeddings of held-out test images using UMAP²¹.

UMAP constructs a fuzzy k-nearest neighbor graph in a high-dimensional feature space and optimizes the low-dimensional coordinates to preserve local structure. The UMAP was fitted jointly to the concatenated Baseline and CLAHE embeddings of the same test images ($n_neighbors = 15$). The baseline and CLAHE coordinates were then separated to compute the per-image displacement vectors. This joint fit strategy treats both domains symmetrically and reveals the relative shifts in the learned representation under CLAHE.

Activations were L2-normalized and embedded using the cosine distance. We visualized EfficientNet-B6, ConvNeXt-Large, and EfficientNet-B7 to assess the architecture-dependent effects. To quantify the observed changes, we computed per-class centroid displacements and $1-\sigma$ ellipse-area ratios (CLAHE/Base), capturing cluster movement and contraction or dispersion in the embedded space.

2.8. Full-Dev Training and Final Test Reporting.

After Phase B identified EfficientNet-B6 with CLAHE as the strongest configuration, we conducted the final model development stage using a training/calibration/testing protocol. The BTXRD dataset provides two predefined splits; we treated the authors' validation split as a fixed held-out test set and re-partitioned the original training split into a new training set (~90%) and a calibration set (~10%) via stratified shuffle splitting. Three independent models were trained using three seeds, and all model selections and checkpointing decisions were based on the calibration-set macro-F1. The final evaluation reports the overall accuracy and macro-F1; per-class precision, recall, and

accuracy; the precision-recall curve (more suitable than AUROC owing to class imbalance); the normalized confusion matrix; and per-class ECE values.

2.9. Class Activation Mapping . We computed CAMs using HiResCAM²² on the best-performing classifier selected by validation macro-F1 on the test set to assess the spatial alignment between model attention and annotated lesions. CAMs were extracted from the last convolutional layer, overlaid on grayscale radiographs, and displayed using expert annotation contours. In addition to correctly classifying tumor cases, we evaluated the CAM localization metrics for misclassified tumors, stratified by misclassification direction (malignant predicted as benign and benign predicted as malignant), to analyze failure modes. Because expert tumor masks are often irregular, and the CAMs generated by convolutional networks are spatially coarser than the input images²³, we dilated each polygon mask using a small, fixed amount of ~5% of the larger image dimension to compensate for minor boundary mismatches introduced by resizing and interpolation.

The localization quality was evaluated using the Quantus library. The Pointing Game (PG) metric captures whether the single most activated CAM pixel falls inside an expert lesion mask, and the Relevance Mass Accuracy (RMA)²⁴ quantifies the proportion of the total attribution mass located within the ground-truth mask. The metrics were reported overall and per class. Representative CAM panels were drawn from a balanced subset of tumor subtypes and various levels of model certainty.

3. RESULTS AND DISCUSSION

3.1. Phase A: Comparative Classification Performance. Table 1 shows Phase-A validation accuracy (mean $\pm$ 95% CI) and per-class precision and recall. Models are sorted by accuracy, and custom image-size models were excluded from Table 1 to improve readability, as these variants underperformed their default-resolution baselines. Their metrics are provided in Figure 1.

The EfficientNet family dominated the Phase A evaluation. EfficientNet-B7 achieves the highest mean validation accuracy (0.9185). The closest behind were EfficientNet-V2-M (0.9161) and EfficientNet-B6 (0.9132). The overlapping CIs indicated similar performance differences among the top EfficientNets. ConvNeXt-Large performed well for a non-EfficientNet CNN (0.8985), yet failed. The transformer-based backbones were consistently weaker: MaxViT-T and Swin-V2-B achieved accuracies of 0.8804 and 0.8754, respectively. Although both transformers were evaluated at their native ImageNet input resolutions, these resolutions were substantially lower than those used by the high-performance EfficientNet variants. Thus, it remains plausible that higher-resolution transformer architectures would better capture fine-grained radiographic details and may therefore close part of the observed performance gap.

Each row lists a model's mean validation accuracy ($\pm$ 95% CI) and per-class precision/recall. Models are sorted in ascending order by overall accuracy (best at the bottom). MobileNet is excluded for brevity.

Across models, recall for no tumor and benign is high (typically $\geq$ 0.93 and ~0.86–0.90, respectively, for the top models).

The malignant class is consistently more difficult, with recalls mostly in the 0.74–0.82 range. Therefore, while B7 topped the accuracy leaderboard, B6 led to malignant recall among CNNs, indicating that class-balanced training may be a valuable lever in Phase B. From a clinical perspective, reduced malignant recall corresponds to false-negative errors, which carry the highest risk owing to delayed diagnosis or missed malignancy. In contrast, false positives primarily require additional imaging or biopsy, which, although undesirable, are generally less harmful than missed diagnoses. The consistent difficulty of the malignant class motivated a later emphasis on macro-F1 and malignant class behavior rather than accuracy.

The baseline YOLO B7XRD authors reported an accuracy of 0.884, and posted the highest malignancy recall (0.839) at the cost of lower malignancy precision (0.734). Being recall-heavy, it trades more false positives for sensitivity. In contrast, EfficientNets achieved a much higher malignant precision while maintaining competitive recall, resulting in a more clinically balanced trade-off.

Figure 1 shows a clear separation between the EfficientNet models and other backbones, and the overlapping confidence intervals among the top-performing CNNs.

3.2. Cross-Validated Performance and CLAHE Ablation.

3.2.1. Model ranking results. Phase B introduces a more rigorous evaluation using five-fold stratified cross-validation with fixed seeds and model selection, by validating macro-F1, which provides a stricter measure of generalization and isolates the effect of CLAHE under identical training conditions. Macro-F1 was used as the primary performance metric to address class imbalance, with accuracy reported for completeness.

Table 2 ranks the five evaluated architectures by their cross-validated macro-F1 (mean $\pm$ 95% CI). EfficientNetV2-M achieved the highest macro-F1 (0.8863), followed closely by EfficientNet-B7 (0.8860) and EfficientNet-B6 (0.8750). ConvNeXt-Large and Swin V2-B performed reliably but remained below the EfficientNet models.

Table 3 reports the same metrics for models trained with CLAHE activated. Interestingly, EfficientNet-B6 becomes the top performer (0.8886), exceeding its baseline counterpart, and has the best overall Macro F1. EfficientNet-B7 also improves marginally (0.8875), whereas EfficientNetV2-M, ConvNeXt-Large, and Swin V2-B exhibit a decline in Macro F1.

Figure 2 presents the same validation macro-F1 values shown in Tables 2 and 3 as a horizontal bar chart for easier comparison. The CLAHE variants are treated as separate entries.

3.2.2. CLAHE ablation findings. Figure 3 plots macro-F1 across all five folds for each architecture. The following are the key observations from the fold variability plots:

- EfficientNet-B6 displays the most consistent positive shift under CLAHE
- EfficientNetV2-M and EfficientNet-B7 show mixed responses.
- Swin V2-B shows unstable behavior across folds
- ConvNeXt-Large shows mild degradation under CLAHE in most folds

Table 1. Phase-A comparative classification results on BTXRD. Results represent validation performance (mean $\pm$ 95% Confidence Interval [CI]) for models. Abbreviations: BTXRD, Bone Tumor X-ray Radiograph Dataset; CI, Confidence Interval

Model	Overall accuracy	Class	Precision	Recall
Swin_V2_B	0.8754 $\pm$ 0.0097	No Tumor	0.8795 $\pm$ 0.0108	0.9210 $\pm$ 0.0108
		B-tumor	0.8689 $\pm$ 0.0204	0.8484 $\pm$ 0.0179
		M-tumor	0.8862 $\pm$ 0.0532	0.7375 $\pm$ 0.0756
ResNet152	0.8786 $\pm$ 0.0117	No Tumor	0.8889 $\pm$ 0.0126	0.9156 $\pm$ 0.0173
		B-tumor	0.8606 $\pm$ 0.0217	0.8658 $\pm$ 0.0193
		M-tumor	0.9180 $\pm$ 0.0694	0.7219 $\pm$ 0.0538
MaxVit-T	0.8804 $\pm$ 0.0025	No Tumor	0.8815 $\pm$ 0.0137	0.9332 $\pm$ 0.0146
		B-tumor	0.8770 $\pm$ 0.0128	0.8452 $\pm$ 0.0188
		M-tumor	0.8971 $\pm$ 0.0572	0.7406 $\pm$ 0.0402
Yolov8-cls	0.884	No Tumor	0.913	0.898
		B-tumor	0.881	0.875
		M-tumor	0.734	0.839
EfficientNet_B5	0.8961 $\pm$ 0.0070	No Tumor	0.8961 $\pm$ 0.0147	0.9401 $\pm$ 0.0135
		B-tumor	0.8967 $\pm$ 0.0122	0.8658 $\pm$ 0.0241
		M-tumor	0.8982 $\pm$ 0.0657	0.7844 $\pm$ 0.0319
ConvNeXt Large	0.8985 $\pm$ 0.0053	No Tumor	0.9071 $\pm$ 0.0142	0.9347 $\pm$ 0.0154
		B-tumor	0.8900 $\pm$ 0.0179	0.8787 $\pm$ 0.0233
		M-tumor	0.8919 $\pm$ 0.0464	0.7812 $\pm$ 0.0475
EfficientNetV2-L	0.9097 $\pm$ 0.0095	No Tumor	0.9110 $\pm$ 0.0118	0.9480 $\pm$ 0.0236
		B-tumor	0.9120 $\pm$ 0.0214	0.8858 $\pm$ 0.0154
		M-tumor	0.8998 $\pm$ 0.0894	0.8000 $\pm$ 0.0588
EfficientNet-B6	0.9132 $\pm$ 0.0113	No Tumor	0.9140 $\pm$ 0.0160	0.9448 $\pm$ 0.0085
		B-tumor	0.9150 $\pm$ 0.0179	0.8942 $\pm$ 0.0260
		M-tumor	0.9025 $\pm$ 0.0449	0.8187 $\pm$ 0.0506
EfficientNetV2-M	0.9161 $\pm$ 0.0080	No Tumor	0.9101 $\pm$ 0.0114	0.9538 $\pm$ 0.0169
		B-tumor	0.9175 $\pm$ 0.0216	0.8935 $\pm$ 0.0165
		M-tumor	0.9566 $\pm$ 0.0416	0.8031 $\pm$ 0.0378
EfficientNet-B7	0.9185 $\pm$ 0.0066	No Tumor	0.9107 $\pm$ 0.0102	0.9676 $\pm$ 0.0103
		B-tumor	0.9348 $\pm$ 0.0171	0.8839 $\pm$ 0.0258
		M-tumor	0.8935 $\pm$ 0.0487	0.7969 $\pm$ 0.0274

To isolate the effect of CLAHE, Figure 4 shows the paired deltas (CLAHE – baseline) for macro-F1 and the accuracy. To determine whether CLAHE produced a consistent performance change beyond fold-to-fold variability, we computed paired per-fold differences (Δ = CLAHE – baseline) for each architecture (seeds averaged within fold) and applied the exact two-sided Wilcoxon signed-rank test, which tests whether the median paired difference is zero without assuming normality.

EfficientNet-B6 showed the clearest improvement: macro-F1 went from 0.875 to 0.889 (Δ = +0.0136; bootstrap 95% t-CI for mean Δ [0.00850, 0.0190], p = 0.0625). Other models showed small or negative macro-F1 shifts with CIs spanning zero: ConvNeXt-Large Δ macro-F1 = -0.00925 (p = 0.187), EfficientNet-B7 Δ = +0.00151 (p = 0.625), EfficientNetV2-M Δ = -0.00242 (p = 0.812), and Swin V2-B Δ = -0.00676 (p = 0.625). Their accuracy deltas were also small (ConvNeXt

Table 2. Performance of models using baseline preprocessing in Phase B. Architectures are ranked by their cross-validated Macro-F1 score to prioritize performance on imbalanced classes. Abbreviations: Phase B, secondary evaluation phase using stratified cross-validation

Rank	Model	Macro F1	Overall Accuracy
5	Swin V2-B	0.8471 $\pm$ 0.0137	0.8548 $\pm$ 0.0129
4	ConvNeXt Large	0.8744 $\pm$ 0.0202	0.8791 $\pm$ 0.0144
3	EfficientNet-B6	0.8750 $\pm$ 0.0065	0.8875 $\pm$ 0.0044
2	EfficientNet-B7	0.8860 $\pm$ 0.0131	0.8937 $\pm$ 0.0148
1	EfficientNetV2-M	0.8863 $\pm$ 0.0077	0.8914 $\pm$ 0.0057

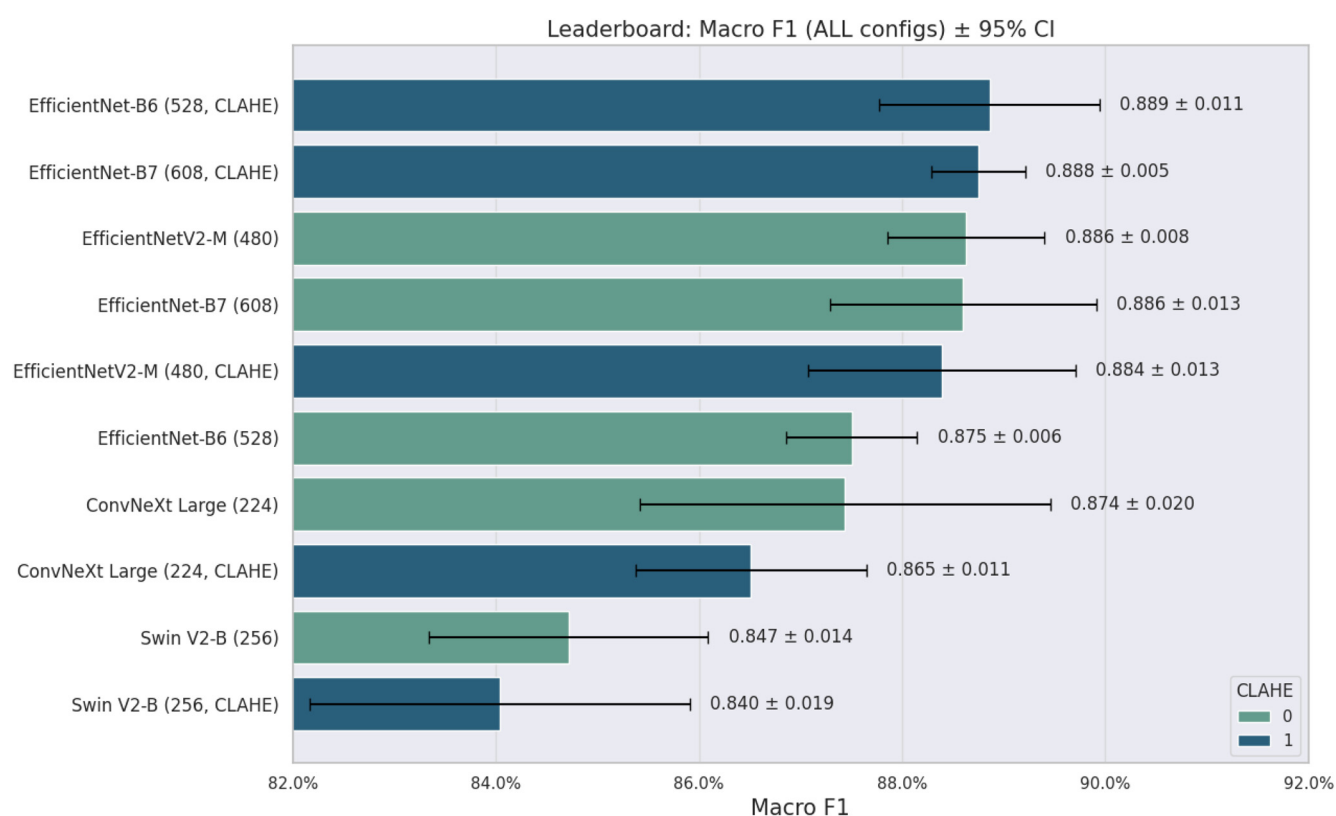


Figure 2. Model performance (Macro-F1) across baseline and CLAHE preprocessing. Green represents baseline (no CLAHE) and blue represents CLAHE images. Results are ranked by baseline performance, with error bars indicating the standard deviation across cross-validation folds. Abbreviations: CLAHE, Contrast Limited Adaptive Histogram Equalization

–0.000223; B7 +0.00275; V2-M +0.00319; Swin +0.00156, all $p \geq 0.312$. Across architectures, overall accuracy tended to increase modestly under CLAHE, even when macro-F1 declined, indicating that accuracy gains were primarily driven by majority-class behavior rather than by improved balance across tumor classes.

The results show that CLAHE is not uniformly beneficial; its impact is strongly dependent on the architecture. This implies that CLAHE should not be considered beneficial for radiographic tumor classification pipelines without model-specific evaluation. The observed divergence between the accuracy and macro-F1 under CLAHE has direct clinical implications. Improvements

Table 3. Performance of models using CLAHE preprocessing in Phase B. Architectures are ranked by their cross-validated Macro-F1 score to prioritize performance on imbalanced classes. Abbreviations: CLAHE, Contrast Limited Adaptive Histogram Equalization; Phase B, secondary evaluation phase using stratified cross-validation

Rank	Model	Macro F1	Overall Accuracy
5	Swin V2-B	0.8404 $\pm$ 0.0187	0.8571 $\pm$ 0.0136
4	ConvNeXt Large	0.8651 $\pm$ 0.0113	0.8788 $\pm$ 0.0121
3	EfficientNetV2-M	0.8839 $\pm$ 0.0132	0.8950 $\pm$ 0.0107
2	EfficientNet-B7	0.8875 $\pm$ 0.0046	0.8978 $\pm$ 0.0046
1	EfficientNet-B6	0.8886 $\pm$ 0.0108	0.8968 $\pm$ 0.0080

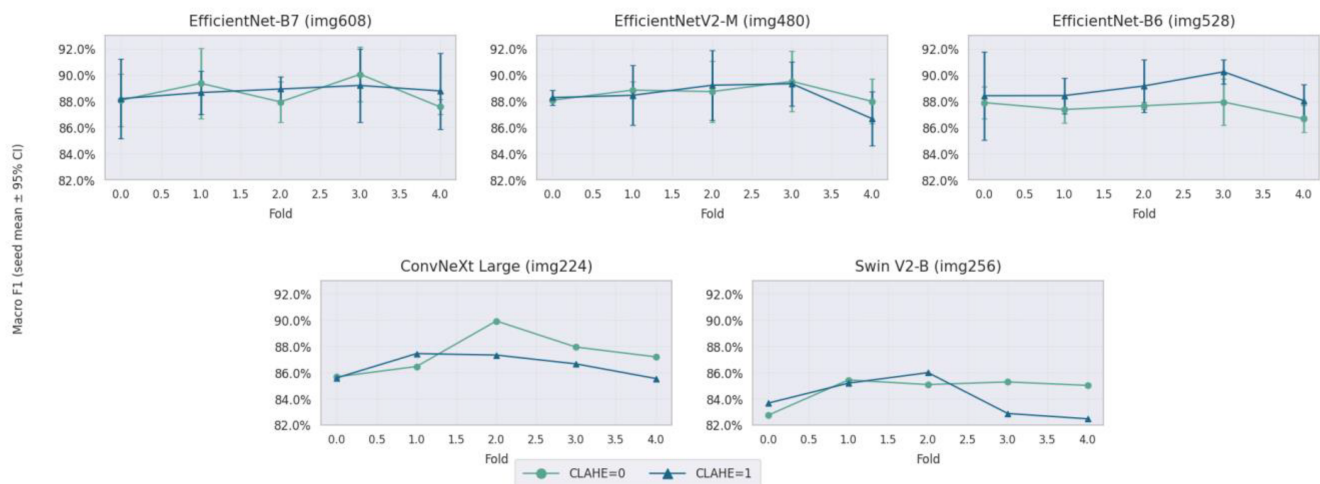


Figure 3. Evaluation of model stability across cross-validation folds using baseline and CLAHE preprocessing. Error bars indicate the 95% CI where applicable. Abbreviations: CLAHE, Contrast Limited Adaptive Histogram Equalization; CI, Confidence Interval

Table 4. Class-cluster stability in UMAP space under baseline and CLAHE preprocessing. Centroid shift denotes the Euclidean distance between class means. The 1- σ ellipse area ratio ($\text{Area}_{\text{CLAHE}} / \text{Area}_{\text{Base}}$) reflects changes in cluster dispersion. Abbreviations: UMAP, Uniform Manifold Approximation and Projection; CLAHE, Contrast Limited Adaptive Histogram Equalization; 1- σ ellipse, an elliptical boundary encompassing 68% of class samples

Model	Centroid Shift (No Tumor)	Centroid shift (Benign)	Centroid shift (Malignant)	Ellipse Area Ratio (CLAHE/Base) No Tumor	Ellipse Area Ratio (CLAHE/Base) Benign	Ellipse Area Ratio (CLAHE/Base) Malignant
ConvNeXt Large	3.26	1.85	1.22	0.77	1.30	1.79
EfficientNet-B6	7.26	7.53	2.49	0.47	1.18	0.34
EfficientNet-B7	23.67	17.91	3.47	0.73	1.31	0.89

in overall accuracy may obscure the degraded performance in minority or high-risk classes, particularly malignant tumors. Therefore, preprocessing choices, such as CLAHE, must be evaluated using class-sensitive metrics to avoid inadvertently increasing clinically consequential errors.

3.3. UMAP Representation-Shift Analysis. For the UMAP analysis, we used the most representative models to illustrate the full range of CLAHE effects observed during cross-validation: ConvNeXt (CLAHE-hurt), EfficientNet-B6 (CLAHE-helped), and EfficientNet-B7 (relatively neutral). Each representative model was selected from the fold-seed combination whose CLAHE-induced change in macro-F1 was closest to that model's overall mean macro-F1, ensuring typical rather than outlier behavior.

Table 4 summarizes the per-class centroid shifts and elliptical area ratios. Figure 5 shows the corresponding embeddings. EfficientNet-B6 exhibited the most pronounced benefit: CLAHE contracted the malignant cluster and improved its separation from benign features. ConvNeXt-Large displayed the opposite trend, with malignant features becoming more dispersed and overlapping more heavily with benign representations, reflecting a reduced macro-F1 under CLAHE. EfficientNet-B7 exhibited moderate centroid movement with comparatively small changes in cluster compactness.

Quantitatively, EfficientNetB6 showed the most favorable CLAHE-induced representational change, with its malignant cluster area reduced to 0.34 of baseline (66% contraction) and a modest centroid shift of 2.49. Conversely, ConvNeXtLarge exhibited increased malignant dispersion (elliptical area ratio

Table 5. Final test performance metrics for the EfficientNet-B6 architecture. Results reflect the model's performance on the independent test set following the Train/Calibration/Test protocol. Abbreviations: B-tumor, Benign tumor; M-tumor, Malignant tumor

Model	Overall accuracy	Macro-F1	Class	Precision	Recall	Accuracy
EfficientNet-B6	0.8908	0.8810	No Tumor	0.8923	0.9231	0.9055
			B-tumor	0.8746	0.8774	0.8975
			M-tumor	0.9800	0.7656	0.9787

Table 6. Expected Calibration Error (ECE) for diagnostic classes. ECE values measure the correspondence between predicted probabilities and actual accuracy

Model	No Tumor ECE	Benign ECE	Malignant ECE
EfficientNet-B6	0.079	0.087	0.022

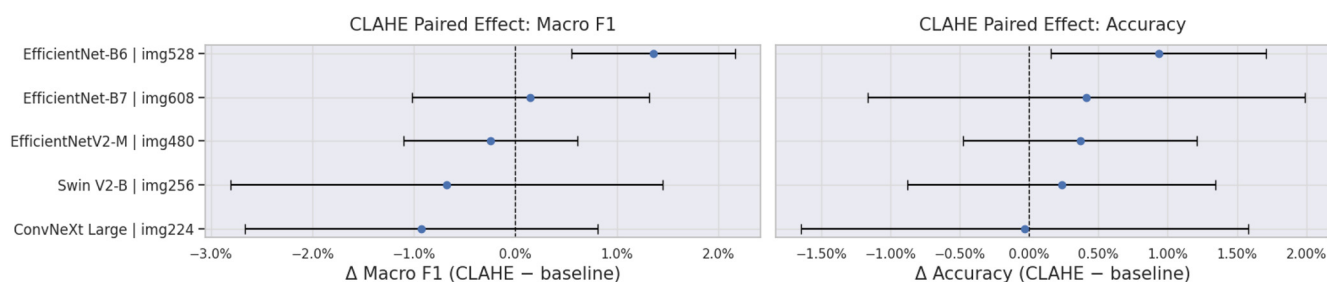


Figure 4. Paired effect of CLAHE preprocessing on model performance metrics. Bars represent the paired difference (CLAHE - baseline) for Macro-F1 and Overall Accuracy. Error bars represent the bootstrap 95% CI for the mean difference. Abbreviations: CLAHE, Contrast Limited Adaptive Histogram Equalization; CI, Confidence Interval

1.79) despite minimal centroid movement (1.22), consistent with its reduced macroF1. EfficientNetB7 displays large centroid shifts in the no-tumor and benign classes (23.67 and 17.91%, respectively), while the malignant geometry remains relatively stable (centroid 3.47, area 0.89), aligning with its near-neutral performance change. Where the Ellipse area ratio refers to the ratio of the $1-\sigma$ covariance ellipse area under CLAHE relative to baseline, with values below 1 indicating cluster contraction. Overall, the UMAP analysis reinforces that CLAHE's representational impact is architecture-dependent: it enhances feature

geometry only in EfficientNet-B6, whereas it alters or disrupts it in other models.

3.4. Final Model Training and Results. Table 5 shows the final test results for EfficientNet-B6, selected from Phase B (best F1) and retrained using the final train-calibration-test protocol. The model achieved strong overall test performance (accuracy = 0.8908, Macro-F1 = 0.8810). Normal and benign radiographs were handled well, whereas the malignant class exhibited the largest disparity between the precision (0.98) and

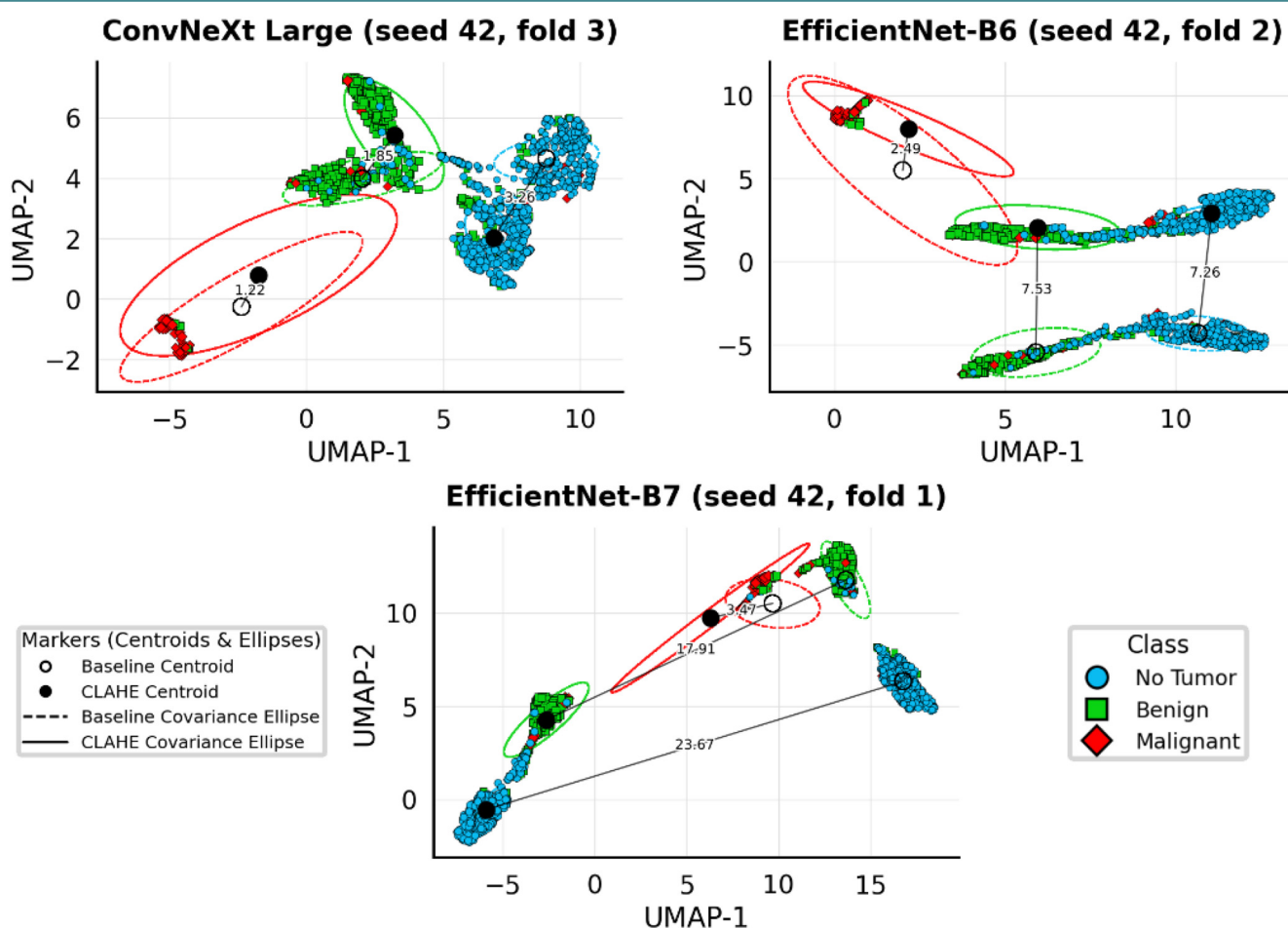


Figure 5. Joint UMAP visualizations of model embeddings under baseline and CLAHE preprocessing. Plots show the distribution of tumor classes in embedding space. Abbreviations: UMAP, Uniform Manifold Approximation and Projection; CLAHE, Contrast Limited Adaptive Histogram Equalization

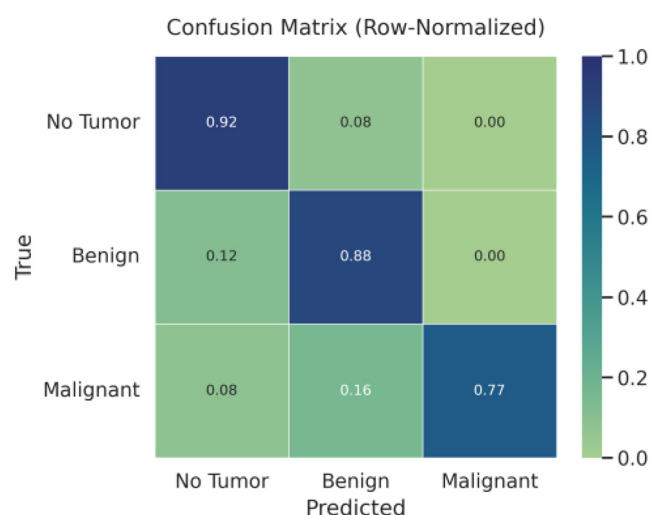


Figure 6. Normalized confusion matrix for EfficientNet-B6 Rows represent the ground-truth physician annotations (True Class), while columns represent the model's predictions (Predicted Class)

recall (0.7656), indicating that the model was highly selective for predicting malignancy. This operating point reflects a clinically conservative strategy that prioritizes high-confidence malignancy predictions at the expense of sensitivity. Such behavior is suitable for decision support or use-case prioritization, whereas screening applications would require threshold tuning to increase malignant recall.

The row-normalized confusion matrix shown in Figure 6 supports this interpretation. Most misclassifications were concentrated between non-tumor–benign and benign–malignant boundaries. Malignant errors predominantly shifted toward the benign class, which is clinically preferable. This pattern highlights the inherent radiographic overlap between benign and malignant lesions, making it one of the most clinically challenging distinctions. It emphasizes the importance of cautious interpretation near this boundary. The predicted malignancy column shows 0 in the middle row because of rounding.

The precision–recall analysis in Figure 7 provides deeper insight into the model. No tumors remained highly separable across most thresholds. The benign and macro-curves declined faster, whereas the malignant curve deteriorated the earliest. Malignant precision remains above 90% until ~0.75 recall, then steeply drops, reflecting the increased false-positive burden typical of imbalanced datasets.

The calibration analysis in Table 6 shows moderate miscalibration for nontumor (ECE 0.079) and benign (0.087) tumors, with the malignant class showing the lowest error (0.022), consistent with the model's conservative probability outputs for malignancy.

Table 7. Performance of HiResCAM for correctly classified tumor cases. Results are evaluated using Pointing Game (PG) accuracy and Relevance Mass Accuracy (RMA). Abbreviations: HiResCAM, High-Resolution Class Activation Mapping; B-tumor, Benign tumor; M-tumor, Malignant tumor

Model	Method	Class	Number of Images	Pointing Game Accuracy	Relevance Mass Accuracy
EfficientNet-B6	HiResCam	Overall	30	0.5	0.392
		B-tumor	20	0.35	0.322
		M-tumor	10	0.8	0.532

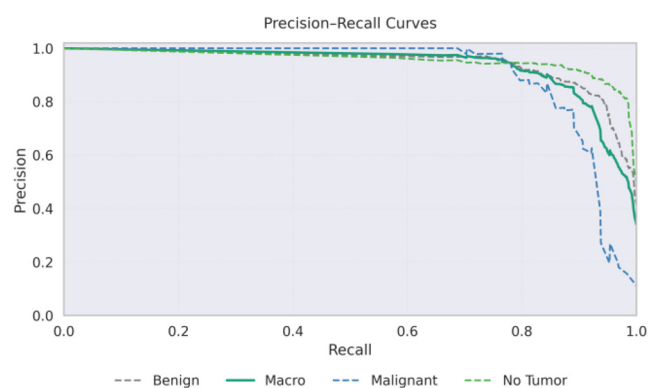


Figure 7. Precision-Recall (PR) curves for EfficientNet-B6. Curves illustrate the diagnostic trade-off between precision and recall for each tumor class. The steep decline in the malignant curve highlights the threshold where false positives increase

Overall, the model demonstrated strong discrimination across all classes; however, malignancy detection remained the primary source of performance variability. The PR and calibration analyses together indicate that threshold tuning or post hoc calibration may improve operational reliability.

3.5. Class Activation Mapping Analysis. Table 7 summarizes the localization performance of the CAM analysis on correctly classified tumor cases. Across 30 images, EfficientNet-B6 achieved a Pointing Game (PG) accuracy of 0.50, and Relevance Mass Accuracy (RMA) of 0.392, indicating that roughly half of the peak activations and less than half of the attribution mass fell within the annotated lesion regions. Localization differed by class; benign cases showed lower alignment, whereas malignant cases achieved substantially higher PG and RMA scores.

To better understand the model failure modes, we extended the CAM evaluation to misclassified tumor cases (Table 8). Compared to correctly classified images, the localization performance deteriorated markedly, with a Pointing Game accuracy of 0.27 and a mean Relevance Mass Accuracy (RMA) of 0.17, indicating an association between erroneous predictions and diffuse or off-lesion attention.

Failure mode-specific analysis revealed distinct patterns. In malignant-to-benign errors (false negatives), the Pointing Game accuracy remained moderate (0.60), with a partial attribution mass inside the lesion (Mean RMA = 0.31), suggesting that the model often attends to the tumor but fails to extract sufficiently discriminative malignant cues.

By contrast, only a single benign-to-malignant error was observed, and this case exhibited lesion-centered attention (PG = 1.00, RMA = 0.59), illustrating that spatially aligned attention does not necessarily guarantee correct classification.

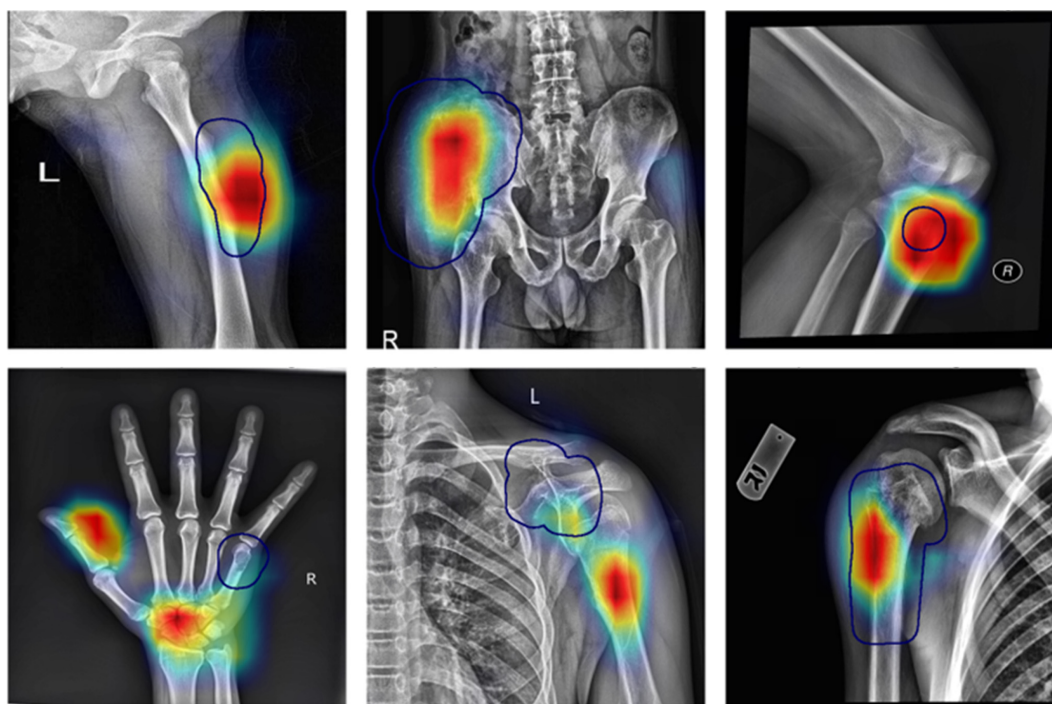


Figure 8. Representative examples from the test split showing images with corresponding HiResCAM overlays, highlighting discriminative regions for interpretability; not intended as precise clinical segmentations. Abbreviation: HiResCAM, High-Resolution Class Activation Mapping

Figure 8 shows representative CAM overlays for both well-aligned and less-aligned cases. The panel includes examples across different confidence levels, illustrating the variety of localization behaviors seen in the quantitative metrics. Importantly, CAMs should be viewed as approximate visual rationales, highlighting where the model identified discriminative evidence rather than precise explanations segmentation.

Collectively, CAM explanations capture only part of the annotated tumor regions, with less than half of the attributed mass typically falling inside the ground-truth mask. PG and RMA indicate that CAMs are helpful for understanding the model's general focus, although they offer limited spatial precision. As a result, they can support clinical transparency, but should be interpreted cautiously, not as definitive evidence of lesion boundaries. These localization results indicate that CAMs should be interpreted as coarse visual explanations rather than precise delineations of tumor extent. Analysis of misclassified cases showed that lesion-centered attention may persist in benign cases predicted as malignant errors, whereas malignant cases predicted as benign errors were more often associated with reduced or partial localization within the lesion, reflecting a weaker attribution alignment. This indicates that spatially plausible attention does not necessarily guarantee an accurate diagnosis. Although they may enhance transparency and clinician trust by

highlighting regions of model attention, they are not sufficiently accurate for tasks that require precise lesion boundaries, such as measurements or surgical planning.

4. CONCLUSIONS

This study provides a unified, reproducible benchmark for contemporary deep learning architectures for bone tumor radiograph classification, using the multicenter BTXRD dataset. Across both the canonical split and a more rigorous k-fold cross-validated setting with a frozen test set, compound-scaled CNNs, particularly those of the EfficientNet family, consistently outperformed the transformer backbones and reported YOLO-based BTXRD baselines. CLAHE ablation revealed that contrast enhancement exerted a strong architecture-dependent effect; although performance was unstable across several models, EfficientNet-B6 demonstrated a clear and reproducible improvement. The UMAP-based analysis aligned with these findings, showing that CLAHE tightened the malignant feature clusters for EfficientNet-B6 but disrupted separability in weaker backbones. The final full-development model, EfficientNet-B6 with CLAHE, achieved 0.8908 accuracy and 0.8810 macro-F1 on the held-out test set, with errors largely restricted to clinically adjacent classes and malignant predictions, and exhibited high precision

Table 8. Performance of HiResCAM saliency maps for misclassified tumor cases. Abbreviations: HiResCAM, High-Resolution Class Activation Mapping; PG, Pointing Game; RMA, Relevance Mass Accuracy

Misclassification	Number of Images	Pointing Game Accuracy	Relevance Mass Accuracy
All	30	0.27	0.172
Benign as Malignant	1	1.00	0.585
Malignant as Benign	10	0.60	0.309

but conservative recall. HiResCAM explanations indicated partial alignment with tumor regions, especially in malignant cases; however, the overall localization remained limited.

The key limitations of this study include the lack of external validation, advanced calibration strategies, and validation across larger datasets. To partially mitigate the limitations of external validation, we employed stratified multi-seed k-fold cross-validation, a frozen held-out test set, paired statistical testing, and representation-level analyses to reduce split-dependent bias and provide variance-aware estimates of generalization. Further research is warranted to validate findings across institutions; imaging systems are also an important direction for future research. This study established a reproducible, well-regularized pipeline and identified EfficientNet-B6+CLAHE as a strong baseline for future research on BTXRD classification, with supporting explainability analyses.

5. CODE AVAILABILITY

The code supporting the findings of this study is publicly available at: <https://github.com/sbw200/btxrd-benchmark>.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Saad Bin Waqas – Information & Computer Science Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; s202337730@kfupm.edu.sa

Corresponding Author

Muhammad Shahid Jabbar – Research Mentor, Postdoctoral Fellow, SDAIA-KFUPM JRC for Artificial Intelligence, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; muhammad.jabbar@kfupm.edu.sa

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the SDAIA-KFUPM JRC for Artificial Intelligence for providing computational resources and research infrastructure.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

- (1) Saba, L. *et al.* The present and future of deep learning in radiology. *Eur. J. Radiol.* **114**, 14–24 (2019).
- (2) He, Y. *et al.* Deep learning-based classification of primary bone tumors on radiographs: A preliminary study. *eBioMedicine* **62**, 103121 (2020).
- (3) Pannu, R. *et al.* Enhanced glaucoma classification through advanced segmentation by integrating cup-to-disc ratio and neuro-retinal rim features. *Comput. Med. Imaging Graph.* **123**, 102559 (2025).
- (4) Zubair, M. *et al.* An interpretable framework for gastric cancer classification using multi-channel attention mechanisms and transfer learning approach on histopathology images. *Sci. Rep.* **15**, 13087 (2025).
- (5) Von Schacky, C. E. *et al.* Multitask Deep Learning for Segmentation and Classification of Primary Bone Tumors on Radiographs. *Radiology* **301**, 398–406 (2021).
- (6) Yao, S. *et al.* A Radiograph Dataset for the Classification, Localization, and Segmentation of Primary Bone Tumors. *Sci. Data* **12**, 88 (2025).
- (7) Chen, S., Peng, Y., Liu, Y., Wang, P. & Liu, T. Enhancing YOLOv11 with Large Kernel Attention and Multi-Scale Fusion for Accurate Small and Multi-Lesion Bone Tumor Detection in Radiographs. *Diagnostics* **15**, 1988 (2025).
- (8) Bokhari, S. M. & Shafi, M. Deep learning driven radiographic classification of primary bone tumors using attention augmented hybrid models. *Biomed. Signal Process. Control* **113**, 108888 (2026).
- (9) Widiyaningtyas, T., Hairani, H., Prasetya, D. D., Pujiyanto, U. & Caesarendra, W. A Modified SMOTE with Noise Filtering and Manhattan Distance Metric Approach to Address Imbalanced Health Datasets. *Eng. Technol. Appl. Sci. Res.* **15**, 25452–25459 (2025).
- (10) Muttaqin, F. *et al.* A Combination Method of ROI, CLAHE, and DenseNet-169 for Hip Osteoarthritis Detection. *Eng. Technol. Appl. Sci. Res.* **15**, 22690–22697 (2025).
- (11) Gupta, S. & Sharma, D. Bone Fracture Classification using Transfer Learning. Preprint at <https://doi.org/10.48550/ARXIV.2406.15958> (2024).
- (12) He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770–778).
- (13) Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105–6114). PMLR.
- (14) Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11976–11986).
- (15) Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H. (2019). Searching for mobilenetv3. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1314–1324).
- (16) Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., ... & Guo, B. (2022). Swin transformer v2: Scaling up capacity and resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12009–12019).
- (17) Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., & Li, Y. (2022, October). Maxvit: Multi-axis vision transformer. In European conference on computer vision (pp. 459–479). Cham: Springer Nature Switzerland.
- (18) Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: fast and flexible image augmentations. *Information*, **11**(2), 125.
- (19) Loshchilov, I. & Hutter, F. Decoupled Weight Decay Regularization. Preprint at <https://doi.org/10.48550/ARXIV.1711.05101> (2017).
- (20) Cui, Y., Jia, M., Lin, T. Y., Song, Y., & Belongie, S. (2019). Class-balanced loss based on effective number of samples. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9268–9277).
- (21) McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, **3**(29).
- (22) Draelos, R. L. & Carin, L. Use HiResCAM instead of Grad-CAM for faithful explanations of convolutional neural networks. Preprint at <https://doi.org/10.48550/arXiv.2011.08891> (2021).
- (23) Saporta, A. *et al.* Benchmarking saliency methods for chest X-ray interpretation. *Nat. Mach. Intell.* **4**, 867–878 (2022).
- (24) Arras, L., Osman, A. & Samek, W. Ground Truth Evaluation of Neural Network Explanations with CLEVR-XAI. *Inf. Fusion* **81**, 14–40 (2022).

Fuzzy-Based Sliding Mode Control for Quadrotor Trajectory Tracking

Mansour Abdullah Nasrullah and Syed Muhammad Amri*

Cite <https://doi.org/10.64589/juri/219334>

Submitted: November 26, 2025 Revised: January 13, 2026 Accepted: February 16, 2026

ABSTRACT

With the rapid advancement of low-cost electronics and control technologies, unmanned aerial vehicles (UAVs) are now widely used in many practical applications. However, controlling UAVs remains challenging because of their highly nonlinear dynamics, as well as the presence of external disturbances and model uncertainties. To handle these difficulties, sliding mode control (SMC) has become a popular choice for quadrotor UAVs due to its strong robustness against uncertainties and disturbances. But, conventional SMC often suffers from chattering, which can degrade control smoothness and affect actuator performance. To address this limitation, this paper proposes a fuzzy-based sliding mode control (FSMC) scheme for robust quadrotor trajectory tracking under model uncertainties and time-varying disturbances. The controller employs linear sliding surfaces for position and attitude dynamics. A fuzzy inference mechanism adaptively adjusts switching gains based on sliding variables and their derivatives. This adaptive tuning reduces switching intensity near desired trajectories and increases control action when tracking errors grow, thereby mitigating chattering while preserving robustness. Membership functions and rule base are designed to balance tracking accuracy and control effort. Lyapunov stability analysis proves global convergence of sliding surfaces and tracking errors. Numerical simulations in a MATLAB/Simulink environment, including parameter variations and disturbance scenarios, validate controller effectiveness. Compared with conventional PID control and feedback linearization, the proposed scheme achieves smoother control inputs, improved tracking performance, and total error reduction of up to 60%. These results indicate that adaptive fuzzy tuning enhances sliding mode performance while maintaining robustness, making the proposed approach suitable for practical UAV applications that require reliable operation under uncertain and dynamic conditions.

Keywords: unmanned aerial vehicle, trajectory tracking, fuzzy-based sliding mode control, disturbance rejection, Lyapunov theory

1. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) are increasingly used in a wide range of civilian and industrial fields, such as infrastructure inspection, delivery services, surveillance, mining, telecommunications, and precision agriculture¹. As these aircraft are now commonly deployed for both routine operations and safety-critical tasks, there is a growing need for control systems that are reliable and precise. Among the different types of UAVs, quadrotors are especially popular because of their simple mechanical structure, high maneuverability, and capability to function effectively in confined or complex environments.

Conventional linear controllers, such as proportional–integral–derivative (PID) and linear–quadratic regulator (LQR) controllers, are widely used owing to their simple structure and satisfactory nominal performance^{2,3}. However, under nonlinear and uncertain operating conditions, their performance degrades significantly, particularly in the presence of external disturbances and parameter variations. These limitations have motivated extensive research on nonlinear and robust control techniques, among which sliding mode control (SMC) has received considerable attention for quadrotor UAV applications⁴.

SMC is widely recognized for its strong robustness against modeling uncertainties and external disturbances. Lyapunov-based stability analyses have shown that SMC can guarantee the finite-time or exponential convergence of tracking errors in nonlinear systems, even in the presence of uncertainties and perturbations^{5,6}. Despite these advantages, classical SMC suffers from chattering due to the discontinuous switching action of the sign function. Chattering causes high-frequency oscillations in the control input, which may excite unmodeled dynamics and cause actuator performance degradation, thereby limiting practical SMC implementation⁷.

Various methods have been introduced to reduce the chattering effect in SMC while maintaining its strong robustness. One widely used approach is to replace the discontinuous sign function with continuous alternatives, such as saturation or hyperbolic tangent functions⁸. Although this method can successfully decrease chattering, it creates a boundary layer near the sliding surface, which may lead to steady-state tracking errors and reduced robustness. Another approach is High-Order Sliding Mode Control (HOSMC), which produces smoother control signals by moving the discontinuity to higher-order derivatives of the sliding variable^{9,10}. While HOSMC is more effective in

minimizing chattering, it often increases computational demands and depends on accurate system models, making real-time application more challenging. To address these challenges, adaptive gain methods, especially fuzzy-based Sliding Mode Control (FSMC), have been proposed. FSMC uses fuzzy logic to adjust the switching gain in real time according to changes in system behavior. This approach helps to reduce chattering while preserving robustness and improving tracking performance^{11–13}. Fuzzy logic has also been applied to improve traditional PID controllers. Although fuzzy PID controllers are generally more robust than fixed-gain PID controllers, they often do not provide strong stability guarantees when dealing with severe nonlinearities and matched disturbances. This limitation reduces their suitability for highly uncertain quadrotor UAV systems^{14,15}. Therefore, FSMC-based methods offer a more reliable and effective solution for trajectory tracking control of quadrotor UAVs. Although there has been a significant increase in research on FSMC, most existing studies mainly concentrate on fuzzy tuning methods and validation through simulations. In many cases, clear guarantees of global exponential or practical stability for the complete position and attitude dynamics are not clearly demonstrated. In addition, explicit limits on the adaptive switching gains needed to maintain robustness against unknown disturbances are seldom discussed in detail. Based on these observations, this paper presents an FSMC framework for trajectory tracking of quadrotor position and attitude in the presence of model uncertainties and external disturbances. The proposed controller uses a two-input fuzzy inference system (FIS) to adjust the switching gain in real time, which helps to reduce chattering while maintaining the inherent robustness of SMC. In addition, a Lyapunov-based analysis is carried out to prove the global exponential stability of the closed-loop nonlinear system. The main contributions of this study can be summarized as follows:

- (i) An FSMC scheme is proposed for quadrotor position and attitude tracking under uncertainties and external disturbances without requiring prior knowledge regarding the upper bounds.
- (ii) A fuzzy inference mechanism is introduced to adaptively tune the switching gains and effectively mitigate the chattering effects associated with classical SMC.
- (iii) A Lyapunov-based stability analysis is presented to guarantee the global exponential stability of a multi-variable nonlinear closed-loop system.
- (iv) Simulation results demonstrate improved tracking accuracy and smoother control effort using the proposed scheme compared to a conventional PID controller.

The remainder of this paper is organized as follows. Section 2 defines the problem, introduces quadrotor dynamics, and outlines the necessary assumptions and preliminaries. In Section 3, the proposed FSMC approach is presented, detailing the two-input FIS (including the membership functions and rule base) and presenting the Lyapunov-based stability analyses—global exponential stability for the discontinuous control law and practical stability under saturation. Section 4 presents the MATLAB/Simulink simulation results incorporating parameter variations and time-varying disturbances and provides a performance comparison between the proposed FSMC and a PID controller. Finally, Section 5 concludes the paper.

2. PROBLEM FORMULATION

A quadrotor UAV is a six-degree-of-freedom (6-DOF) system that exhibits three translational and three rotational motions with respect to its reference axes. By adjusting the rotational speeds of the four motors, the quadrotor can generate the forces and moments required to move in any desired direction. Its altitude is mainly controlled by the total thrust produced by the propellers, whereas its orientation, defined by roll, pitch, and yaw, is regulated by the differential rotor speeds that produce the necessary torques. The UAV's position in space is represented by coordinates (x, y, z) , and its orientation is described by Euler angles (ϕ, θ, ψ) , corresponding to roll, pitch, and yaw rotations, respectively.

2.1. Quadrotor Dynamic Model. In the context of modeling and control design, the quadrotor is commonly decomposed into two primary subsystems: the position and attitude subsystems, which govern translational and rotational dynamics, respectively.

Position Subsystem: The equations describing the UAV translational motion are as follows¹⁶:

$$\begin{cases} \ddot{x} = \frac{U_1}{m} U_x + d_x, \\ \ddot{y} = \frac{U_1}{m} U_y + d_y, \\ \ddot{z} = \frac{U_1}{m} \cos(\phi) \cos(\theta) - g + d_z, \end{cases} \quad (1)$$

where U_1 represents the control thrust applied along the z -axis. $U_x = \sin(\phi) \sin(\psi) + \cos(\phi) \sin(\theta) \cos(\psi)$ and $U_y = -\sin(\phi) \cos(\psi) + \cos(\phi) \sin(\theta) \sin(\psi)$ denote the virtual control inputs corresponding to the x - and y -axes, respectively. m denotes the total quadrotor mass, g is the gravitational acceleration, and d_z , d_x , and d_y represent the external disturbances acting along the z -, x -, and y -axes, respectively.

Attitude Subsystem: The quadrotor rotational dynamics can be described by the following model¹⁵:

$$\begin{cases} \ddot{\phi} = a_1 \dot{\phi} \dot{\psi} + b_1 U_2 + d_\phi, \\ \ddot{\theta} = a_2 \dot{\phi} \dot{\psi} + b_2 U_3 + d_\theta, \\ \ddot{\psi} = a_3 \dot{\phi} \dot{\theta} + b_3 U_4 + d_\psi. \end{cases} \quad (2)$$

The parameters in Eq. 2 are defined as:

$$\begin{cases} a_1 = \frac{J_y - J_z}{J_x}, & a_2 = \frac{J_z - J_x}{J_y}, & a_3 = \frac{J_x - J_y}{J_z}, \\ b_1 = \frac{1}{J_x}, & b_2 = \frac{1}{J_y}, & b_3 = \frac{1}{J_z}, \end{cases} \quad (3)$$

where J_x , J_y , and J_z represent the principal moments of inertia around the respective body axes; d_ϕ , d_θ , and d_ψ represent the external disturbances affecting the roll, pitch, and yaw motions, respectively.

The position and attitude subsystems described in Eqs. 1 and 2 can be expressed in compact state-space representation as follows:

$$\dot{\mathbf{P}} = \mathbf{F}_P + \mathbf{G}_P \mathbf{U}_P + \mathbf{D}_P, \quad (4)$$

$$\dot{\mathbf{R}} = \mathbf{F}_R + \mathbf{G}_R \mathbf{U}_R + \mathbf{D}_R, \quad (5)$$

where $\mathbf{P} = [z, x, y]^T \in \mathbb{R}^3$, $\mathbf{R} = [\phi, \theta, \psi]^T \in \mathbb{R}^3$, $\mathbf{U}_P = [U_1, U_x, U_y]^T \in \mathbb{R}^3$, $\mathbf{U}_R = [U_2, U_3, U_4]^T \in \mathbb{R}^3$, $\mathbf{D}_P = [d_z, d_x, d_y]^T \in \mathbb{R}^3$, and $\mathbf{D}_R = [d_\phi, d_\theta, d_\psi]^T \in \mathbb{R}^3$. The force vectors are represented as $\mathbf{F}_P = [-g, 0, 0]^T \in \mathbb{R}^3$ and

$\mathbf{F}_R = [a_1\dot{\theta}\dot{\psi}, a_2\dot{\theta}\dot{\psi}, a_3\dot{\theta}\dot{\psi}]^T \in \mathbb{R}^3$, and the input matrices are defined by $\mathbf{G}_P = \text{diag}\left(\frac{\cos(\phi)\cos(\theta)}{m}, \frac{1}{m}, \frac{1}{m}\right) \in \mathbb{R}^{3 \times 3}$ and $\mathbf{G}_R = \text{diag}(b_1, b_2, b_3) \in \mathbb{R}^{3 \times 3}$.

Assumption 1. External disturbances $\mathbf{D}_P$ and $\mathbf{D}_R$ are assumed to be bounded such that $\|\mathbf{D}_P\| \leq \bar{d}_P$ and $\|\mathbf{D}_R\| \leq \bar{d}_R$, where $\bar{d}_P > 0$ and $\bar{d}_R > 0$ are unknown constants.

2.2. Error Dynamics. The main objective of a quadrotor control system is to guarantee that the UAV tracks the desired trajectory with exponential accuracy. This means that the tracking errors $\mathbf{E}_P = \mathbf{P} - \mathbf{P}_d$ and $\mathbf{E}_R = \mathbf{R} - \mathbf{R}_d$ converge to zero as time t reaches infinity. The desired trajectories are given by

$$\mathbf{P}_d = [z_d, x_d, y_d]^T, \mathbf{R}_d = [\phi_d, \theta_d, \psi_d]^T.$$

The desired time-varying reference values z_d, x_d, y_d , and ψ_d can be obtained directly through trajectory planning tools or sensor feedback. However, the desired roll and pitch angles, ϕ_d and θ_d , respectively, need to be computed for control implementation. As described by Chen et al.¹⁷, these are calculated as

$$\begin{cases} U'_1 = \frac{U_1}{m} \cos(\phi) \cos(\theta) - g \\ U_z = m \sqrt{U_x^2 + U_y^2 + (U'_1 + g)^2} \\ \phi_d = \arcsin\left(m \frac{U_x \sin \psi_d - U_y \cos \psi_d}{U'_1}\right) \\ \theta_d = \arctan\left(\frac{U_x \cos \psi_d + U_y \sin \psi_d}{U_z}\right). \end{cases} \quad (6)$$

The second-order error dynamics for the system in Eqs. 4 and 5 can be expressed as

$$\ddot{\mathbf{E}}_P = \mathbf{F}_P + \mathbf{G}_P \mathbf{U}_P + \mathbf{D}_P - \ddot{\mathbf{P}}_d, \quad (7)$$

$$\ddot{\mathbf{E}}_R = \mathbf{F}_R + \mathbf{G}_R \mathbf{U}_R + \mathbf{D}_R - \ddot{\mathbf{R}}_d. \quad (8)$$

The error dynamics in Eqs. 7 and 8 can be rewritten as a system of first-order differential equations by introducing transformation variables as follows:

$$\delta_{P1} = \mathbf{E}_P, \delta_{P2} = \dot{\mathbf{E}}_P, \delta_{R1} = \mathbf{E}_R, \delta_{R2} = \dot{\mathbf{E}}_R,$$

Using these new variables, Eqs. 7 and 8 are reformulated as

$$\begin{cases} \dot{\delta}_{P1} = \delta_{P2}, \\ \dot{\delta}_{P2} = \mathbf{F}_P + \mathbf{G}_P \mathbf{U}_P + \mathbf{D}_P - \ddot{\mathbf{P}}_d, \end{cases} \quad (9)$$

$$\begin{cases} \dot{\delta}_{R1} = \delta_{R2}, \\ \dot{\delta}_{R2} = \mathbf{F}_R + \mathbf{G}_R \mathbf{U}_R + \mathbf{D}_R - \ddot{\mathbf{R}}_d. \end{cases} \quad (10)$$

Assumption 2. The UAV state variables are assumed to be fully measurable, and the reference trajectories are known in advance and at least twice continuously differentiable.

2.3. Problem Statement. The main objective of this study was to develop an SMC with a fuzzy-tuned switching gain for UAV trajectory tracking under multiple uncertainties. The proposed strategy ensures robust exponential convergence of tracking errors while adaptively tuning the control parameters. Mathematically, this can be expressed as

$$\lim_{t \rightarrow \infty} \delta_{P1} = 0, \quad \lim_{t \rightarrow \infty} \delta_{R1} = 0. \quad (11)$$

3. PROPOSED CONTROLLER

The proposed control approach is based on a linear SMC framework. Initially, linear sliding surfaces are defined for both translational and rotational dynamics. Based on these surfaces, a discontinuous SMC law is developed to drive the sliding variables to zero, effectively handling matched uncertainties and bounded disturbances and guaranteeing robust and rapid convergence. Two sliding surface vectors are constructed to govern the desired position and attitude dynamics, as follows:

$$\begin{cases} \mathbf{S}_P = \delta_{P2} + C_P \delta_{P1}, \\ \mathbf{S}_R = \delta_{R2} + C_R \delta_{R1}, \end{cases} \quad (12)$$

where $C_P \in \mathbb{R}^{3 \times 3}$ and $C_R \in \mathbb{R}^{3 \times 3}$ are the positive definite diagonal matrices. Taking the derivative of Eq. 12 and using Eqs. 9 and 10 yields

$$\begin{aligned} \dot{\mathbf{S}}_P &= \dot{\delta}_{P2} + C_P \dot{\delta}_{P1} \\ &= \mathbf{F}_P + \mathbf{G}_P \mathbf{U}_P + \mathbf{D}_P - \ddot{\mathbf{P}}_d + C_P \dot{\delta}_{P1} \end{aligned} \quad (13)$$

$$\begin{aligned} \dot{\mathbf{S}}_R &= \dot{\delta}_{R2} + C_R \dot{\delta}_{R1} \\ &= \mathbf{F}_R + \mathbf{G}_R \mathbf{U}_R - \ddot{\mathbf{R}}_d + C_R \dot{\delta}_{R1} \end{aligned} \quad (14)$$

Using the inverse kinematic approach for Eqs. 13 and 14, the proposed FSMC is designed as follows:

$$\mathbf{U}_P = \mathbf{G}_P^{-1} \{ \ddot{\mathbf{P}}_d - \mathbf{F}_P - \mathbf{K}_1 \mathbf{S}_P - (\hat{c}_P + k_{d1}) \text{sign}(\mathbf{S}_P) \}, \quad (15)$$

$$\mathbf{U}_R = \mathbf{G}_R^{-1} \{ \ddot{\mathbf{R}}_d - \mathbf{F}_R - \mathbf{K}_2 \mathbf{S}_R - (\hat{c}_R + k_{d2}) \text{sign}(\mathbf{S}_R) \}, \quad (16)$$

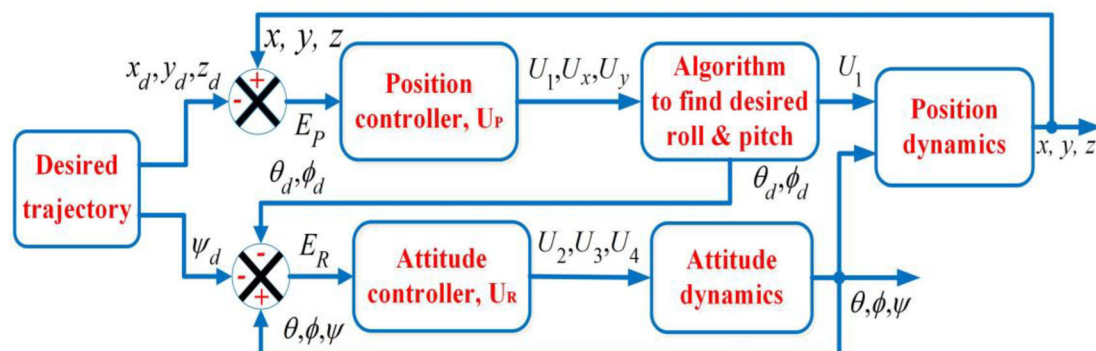


Figure 1. Block diagram for the proposed control scheme

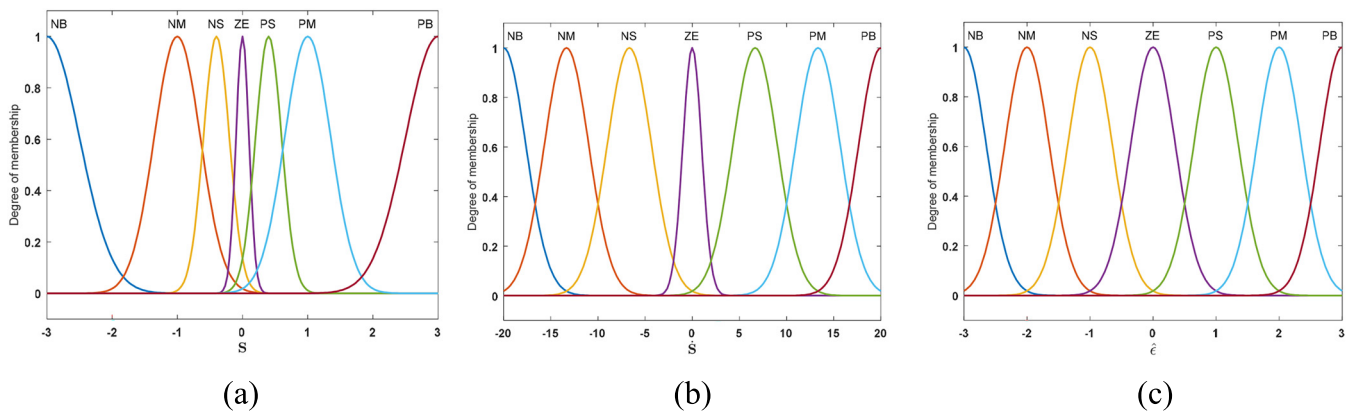


Figure 2. Membership functions for inputs (a) S , (b) $\dot{S}$, and (c) output $\hat{\epsilon}$

where $\mathbf{K}_1 \in \mathbb{R}^{3 \times 3}$ and $\mathbf{K}_2 \in \mathbb{R}^{3 \times 3}$ are positive diagonal gain matrices, $k_{d1} > 0$ and $k_{d2} > 0$ are significantly small switching gains used to ensure robustness during the steady-state period. The schematic of the proposed control scheme is shown in Figure 1.

3.1. Fuzzy System Design. This section describes the fuzzy mechanism used to enhance the SMC switching gain. The fuzzy system maps the selected inputs to the switching gain output, which is specified by its membership functions and rule base. Figure 2 illustrates the membership functions.

The fuzzy system has two inputs (S and $\dot{S}$) and one output ($\hat{\epsilon}$). Output $\hat{\epsilon} = \text{Fuzzy}(S, \dot{S})$ is used as the switching gain in Eqs. 15–16. The membership functions for S , $\dot{S}$, and $\hat{\epsilon}$ are shown in Figure 2. Additionally, the FIS input–output surface $\hat{\epsilon} = \text{Fuzzy}(S, \dot{S})$ is shown in Figure 3, illustrating how the tuned switching gain varies over the $(S, \dot{S})$ plane. The fuzzy rules mapping $(S, \dot{S})$ to $\hat{\epsilon}$ are summarized in Table 1.

The following lemma supports the subsequent stability analysis.

Lemma 1. Consider the continuous-time system¹⁷

$$\dot{\vartheta}(t) = f(\vartheta, v, t), \quad (17)$$

where $\vartheta \in \mathbb{R}^n$ is the state vector, $v \in \mathbb{R}^m$ is the input, and function $f \in \mathbb{R}^n$ is locally Lipschitz-continuous in (ϑ, v) and continuous in t , with $f(0, 0, t) = 0$ for all $t \geq 0$.

Assume there exists a Lyapunov function $\mathcal{V}(\vartheta)$ such that:

- $\mathcal{V}(0) = 0$,
- $\mathcal{V}(\vartheta) > 0$ for all $\vartheta \neq 0$,
- $\mathcal{V}(\vartheta) \rightarrow \infty$ as $\|\vartheta\| \rightarrow \infty$, and
- $\dot{\mathcal{V}}(\vartheta) \leq -\alpha \mathcal{V}(\vartheta)$ for all $\vartheta \neq 0$.

Then, the system in Eq. 17 is globally exponentially stable.

3.2. Stability Analysis. This section presents a stability analysis of the quadrotor system using the proposed FSMC approach.

Theorem 1. Consider the position and attitude sliding dynamics in Eqs. 13 and 14 under Assumptions 1 and 2. The proposed SMC scheme with fuzzy-tuned switching gains in Eqs. 15 and 16

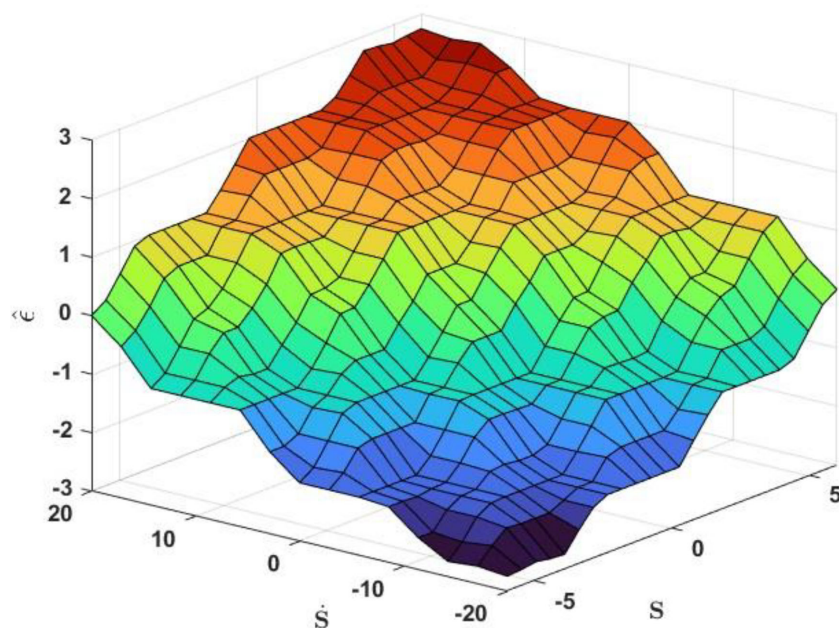


Figure 3. FIS surface $\hat{\epsilon}_p = \text{Fuzzy}(S, \dot{S})$ over inputs S and $\dot{S}$

Table 1. Fuzzy rule base (reordered from PB to NB)

$\hat{e}$	$\hat{s}$						
S	PB	PM	PS	ZE	NS	NM	NB
PB	PB	PB	PM	PM	PS	PS	ZE
PM	PB	PM	PM	PS	PS	ZE	NS
PS	PM	PM	PS	PS	ZE	NS	NS
NS	PS	PS	ZE	NS	NS	NM	NM
ZE	PM	PS	PS	ZE	NS	NS	NM
NM	PS	ZE	NS	NS	NM	NM	NB
NB	ZE	NS	NS	NM	NM	NB	NB

guarantees that all sliding surfaces and corresponding error trajectories converge to zero exponentially.

Proof: This proof proceeds in two stages. The position dynamics stability is first analyzed, followed by the attitude dynamics stability analysis.

Part 1: Consider a Lyapunov function V_p as

$$V_p = \frac{1}{2} \mathbf{S}_p^T \mathbf{S}_p. \quad (18)$$

Taking the time derivative of Eq. 18 and substituting Eqs. 13 and 15 yields

$$\begin{aligned} \dot{V}_p &= \mathbf{S}_p^T \dot{\mathbf{S}}_p = \mathbf{S}_p^T (-\mathbf{K}_1 \mathbf{S}_p + \mathbf{D}_p - (\hat{e}_p + k_{d1}) \text{sign}(\mathbf{S}_p)), \\ &\leq -\lambda_{\min}(\mathbf{K}_1) \|\mathbf{S}_p\|^2 - ((\hat{e}_p + k_{d1}) - \|\mathbf{D}_p\|) \|\mathbf{S}_p\|, \\ &\leq -\lambda_{\min}(\mathbf{K}_1) \|\mathbf{S}_p\|^2 - ((\hat{e}_p + k_{d1}) - \bar{d}_p) \|\mathbf{S}_p\|, \\ &= -\lambda_{\min}(\mathbf{K}_1) \|\mathbf{S}_p\|^2 - \bar{k}_{d1} \|\mathbf{S}_p\|, \\ &\leq -2\lambda_{\min}(\mathbf{K}_1) \left(\frac{\|\mathbf{S}_p\|^2}{2} \right), \\ &\leq -2\lambda_{\min}(\mathbf{K}_1) V_p \leq -\eta_p V_p, \end{aligned} \quad (19)$$

where $\eta_p = 2\lambda_{\min}(\mathbf{K}_1) > 0$ and operator $\lambda_{\min}(\cdot)$ gives the minimum eigenvalue of the associated matrix. From Lemma 1, the above inequality ensures global exponential stability. As a result, the sliding surface $\mathbf{S}_p$ approaches zero as $t \rightarrow \infty$.

Once $\mathbf{S}_p$ converges to zero, Eq. 12 can be expressed as

$$\begin{cases} \delta_{p2} = -C_p \delta_{p1} \Rightarrow \dot{\delta}_{p1} = -C_p \delta_{p1}, \\ \delta_{R2} = -C_R \delta_{R1} \Rightarrow \dot{\delta}_{R1} = -C_R \delta_{R1}. \end{cases} \quad (20)$$

Consider another Lyapunov function $V_{\delta p} = (1/2) \delta_{p1}^T \delta_{p1}$. Taking the time derivative of $V_{\delta p}$ and substituting Eq. 20 yields

$$\begin{aligned} \dot{V}_{\delta p} &= \delta_{p1}^T \dot{\delta}_{p1} = \delta_{p1}^T (-C_p \delta_{p1}), \\ &= -\lambda_{\min}(C_p) \|\delta_{p1}\|^2, \\ &= -2\lambda_{\min}(C_p) \left(\frac{\|\delta_{p1}\|^2}{2} \right) \leq -\mu_1 V_{\delta p}, \end{aligned}$$

where $\mu_1 = 2\lambda_{\min}(C_p) > 0$. From Lemma 1, this inequality ensures asymptotic stability. Therefore, δ_{p1} converges to zero as $t \rightarrow \infty$, which implies that $\mathbf{E}_p \rightarrow 0$.

Part 2: Similarly, the attitude dynamics stability can be established using the same procedure. Accordingly, we consider a

Table 2. System parameters

Parameter	Value	Parameter	Value
m	0.65 kg	J_y	$7.5 \times 10^{-3} \text{ N} \cdot \text{m} \cdot \text{s}^2/\text{rad}$
L	0.23 m	J_x	$7.5 \times 10^{-3} \text{ N} \cdot \text{m} \cdot \text{s}^2/\text{rad}$
g	9.81 m/s^2	J_z	$1.3 \times 10^{-2} \text{ N} \cdot \text{m} \cdot \text{s}^2/\text{rad}$

Lyapunov candidate V_R as

$$V_R = \frac{1}{2} \mathbf{S}_R^T \mathbf{S}_R. \quad (21)$$

Expressing $\dot{V}_R$ using Eqs. 14 and 16 gives

$$\begin{aligned} \dot{V}_R &= \mathbf{S}_R^T \dot{\mathbf{S}}_R = \mathbf{S}_R^T (-\mathbf{K}_2 \mathbf{S}_R + \mathbf{D}_R - (\hat{e}_R + k_{d2}) \text{sign}(\mathbf{S}_R)), \\ &\leq -\lambda_{\min}(\mathbf{K}_2) \|\mathbf{S}_R\|^2 - ((\hat{e}_R + k_{d2}) - \|\mathbf{D}_R\|) \|\mathbf{S}_R\|, \\ &\leq -\lambda_{\min}(\mathbf{K}_2) \|\mathbf{S}_R\|^2 - ((\hat{e}_R + k_{d2}) - \bar{d}_R) \|\mathbf{S}_R\|, \\ &= -\lambda_{\min}(\mathbf{K}_2) \|\mathbf{S}_R\|^2 - \bar{k}_{d2} \|\mathbf{S}_R\|, \\ &\leq -2\lambda_{\min}(\mathbf{K}_2) \left(\frac{\|\mathbf{S}_R\|^2}{2} \right), \\ &\leq -2\lambda_{\min}(\mathbf{K}_2) V_R \leq -\eta_R V_R, \end{aligned} \quad (22)$$

where $\eta_R = 2\lambda_{\min}(\mathbf{K}_2) > 0$. From Lemma 1, this inequality ensures global exponential stability. As a result, the sliding surface $\mathbf{S}_R$ approaches zero as $t \rightarrow \infty$.

Similarly, it can also be established that the attitude error δ_{R1} converges to zero exponentially by selecting a Lyapunov candidate $V_{\delta r} = (1/2) \delta_{R1}^T \delta_{R1}$. The derivative of $V_{\delta r}$ using Eq. 20 can be expressed as

$$\begin{aligned} \dot{V}_{\delta r} &= \delta_{R1}^T \dot{\delta}_{R1} = \delta_{R1}^T (-C_R \delta_{R1}), \\ &= -\lambda_{\min}(C_R) \|\delta_{R1}\|^2, \\ &= -2\lambda_{\min}(C_R) \left(\frac{\|\delta_{R1}\|^2}{2} \right) \leq -\mu_2 V_{\delta r}, \end{aligned}$$

where $\mu_2 = 2\lambda_{\min}(C_R) > 0$. From Lemma 1, this inequality ensures global exponential stability. Therefore, δ_{R1} converges to zero as $t \rightarrow \infty$.

3.3. Statistical Validation Procedure. This study is analytical and simulation-based; therefore, conventional statistical hypothesis testing is not applicable. The performance of the proposed fuzzy sliding mode controller is confirmed through deterministic numerical simulations, complemented by a stability analysis based on Lyapunov theory. The assessment is carried out using conventional control performance criteria, such as tracking error convergence, steady-state accuracy, resilience to bounded disturbances, and required control effort. All comparative simulations are conducted using the same initial conditions and disturbance scenarios to guarantee a fair and consistent evaluation of the controller's performance.

4. SIMULATION RESULTS

The effectiveness of the proposed control method was assessed through numerical simulations carried out in the MATLAB/Simulink environment. The physical parameters of the quadrotor applied in these simulations are presented in Table 2.

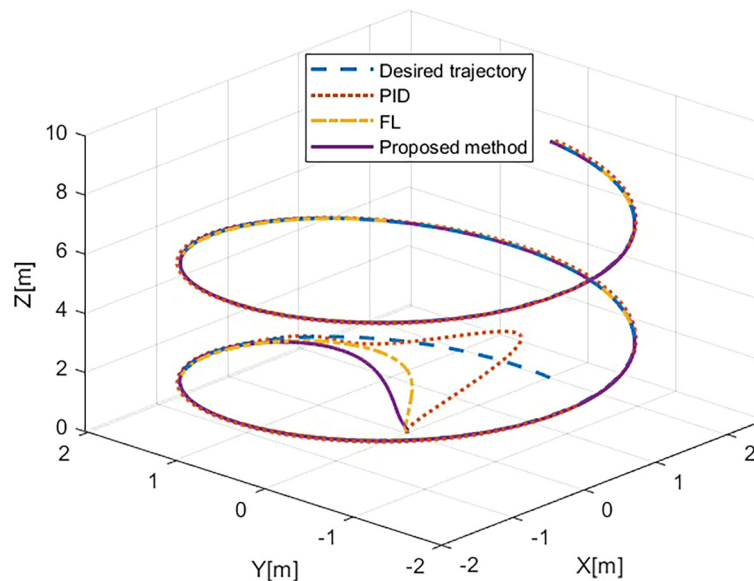


Figure 4. 3D trajectory tracking response under different approaches

For system robustness evaluation, the quadrotor was tested under a 20% mass variation and exposed to time-varying external disturbances following the model in Khalil and Grizzle¹⁸. The external disturbances are considered as $\mathbf{D}_P = [0.1\cos(0.1\pi t) \ 0.1\sin(0.1\pi t) \ 0.1\cos(0.1\pi t)]^T$ and $\mathbf{D}_R = [0.1\sin(0.1\pi t) \ 0.1\cos(0.1\pi t) \ 0.1\sin(0.1\pi t)]^T$.

The selected nominal controller parameters are $\mathbf{K}_1 = \text{diag}(3, 1, 1)$, $\mathbf{K}_2 = \text{diag}(0.6, 0.8, 0.7)$, $k_{d1} = k_{d2} = 0.05$, $C_P = \text{diag}(1, 1, 1)$, and $C_R = \text{diag}(1, 1, 1)$. The reference trajectory used to assess tracking performance is defined as: $z_d = \frac{t}{10}$, $x_d = 2\cos(2\pi\omega t)$, $y_d = 2\sin(2\pi\omega t)$, $\psi_d = 0.0872\sin(2\pi\omega t)$, and $\omega = \frac{2\pi}{40}$.

4.1. Proposed FSMC Performance. This case study demonstrates the overall tracking performance of the proposed FSMC. As illustrated in Figure 4, the quadrotor accurately follows the three-dimensional (3D) reference trajectory, gradually converges to the desired helical path, and maintains stable motion despite the presence of disturbances and uncertainties. Figure 5 shows the position and attitude error responses, providing further evidence of the proposed controller. All tracking errors rapidly converged to zero within approximately 10 s, demonstrating the strong stabilization and ability of the proposed FSMC scheme to quickly achieve the desired position and attitude, with a smooth error graph in both the position and attitude dynamics.

The sliding surface plots in Figure 6 show how the proposed FSMC drives both the position and attitude sliding variables to zero. All sliding surface parameters decreased quickly within the first 10 s and then remained close to zero for the remainder of the motion. This implies that the system reaches the sliding surface quickly and remains there, confirming that the controller keeps the quadrotor stable and on track throughout the desired trajectory.

The control input profiles in Figure 7 illustrate how the proposed FSMC controller regulates thrust force U_1 and torque inputs U_2 , U_3 , and U_4 throughout the trajectory. The main thrust input U_1 is smooth and shows only a slight overshoot during the

initial transient phase. The torque inputs remain relatively constant and low compared to the actuator limits. Moreover, the lack of rapid oscillations in all control channels indicates that chattering is effectively reduced. As a result, the proposed FSMC ensures stable motion without requiring excessive control effort.

Figure 8 shows the fuzzy-tuned switching gains for position and attitude. In the position channel, $\hat{e}$, $\hat{\dot{e}}$, and $\hat{\ddot{e}}$ increase at the onset of the maneuver, and then decrease as the sliding variables approach zero. In the attitude channel, $\hat{e}$, $\hat{\dot{e}}$, and $\hat{\ddot{e}}$ exhibit brief transients before converging near zero once the attitude errors diminish. Therefore, the gains remain high only during large-error phases and decrease in the steady state, concentrating the switching action on the transients.

To further evaluate the effectiveness of the controller, several standard error indices were calculated. The integral of squared error (ISE) was obtained as 4.201, the integral of absolute error (IAE) as 7.413, and the integral of time-squared error (ITSE) as 6.433. The integral of time-absolute error (ITAE) was determined to be 6.5×10^{-4} . As summarized in Table 3, these values highlight the high tracking accuracy and overall control performance of the proposed scheme.

Collectively, the simulations demonstrate that the proposed FSMC maintains robust performance under parameter variations and time-varying disturbances. The controller achieved fast transient, small steady-state errors, and strong disturbance rejection. Integrating the fuzzy tuner into SMC concentrates the switching

Table 3. Error performance measures under two approaches

Approach	Error Performance Index			
	ISE	IAE	ITSE	ITAE
Types of Error				
PID control	9.521	9.258	15.804	1.8×10^{-3}
FL	6.354	8.601	9.560	9.8×10^{-4}
Proposed FSMC	4.201	7.413	6.433	6.5×10^{-4}
Reduction with PID control	55.9%	19.9%	59.3%	63.9%
Reduction with FL	33.9%	13.8%	32.7%	4.2%

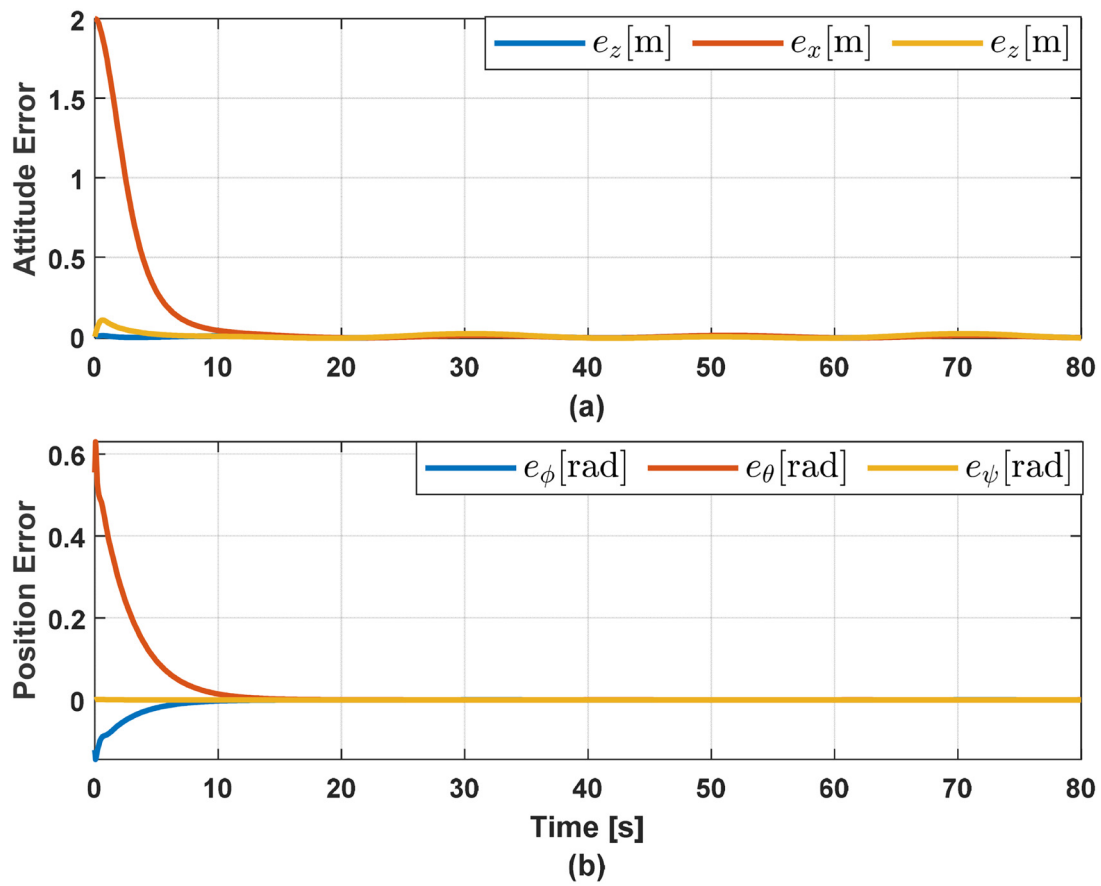


Figure 5. Position and attitude error response under the proposed scheme

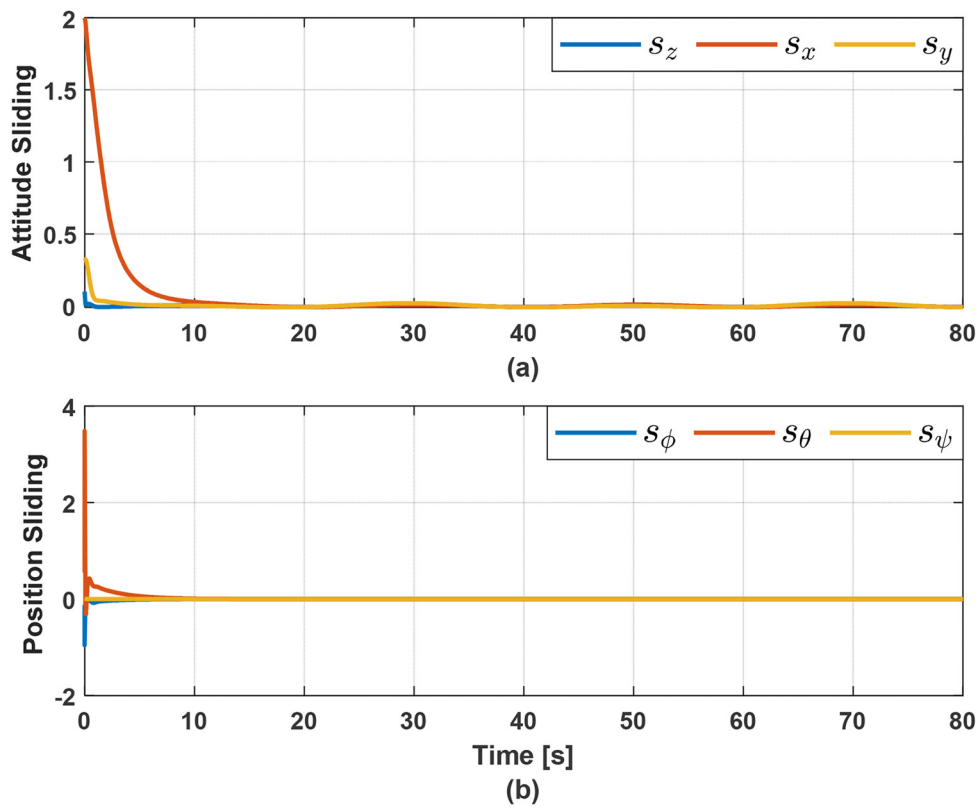


Figure 6. Sliding surface response

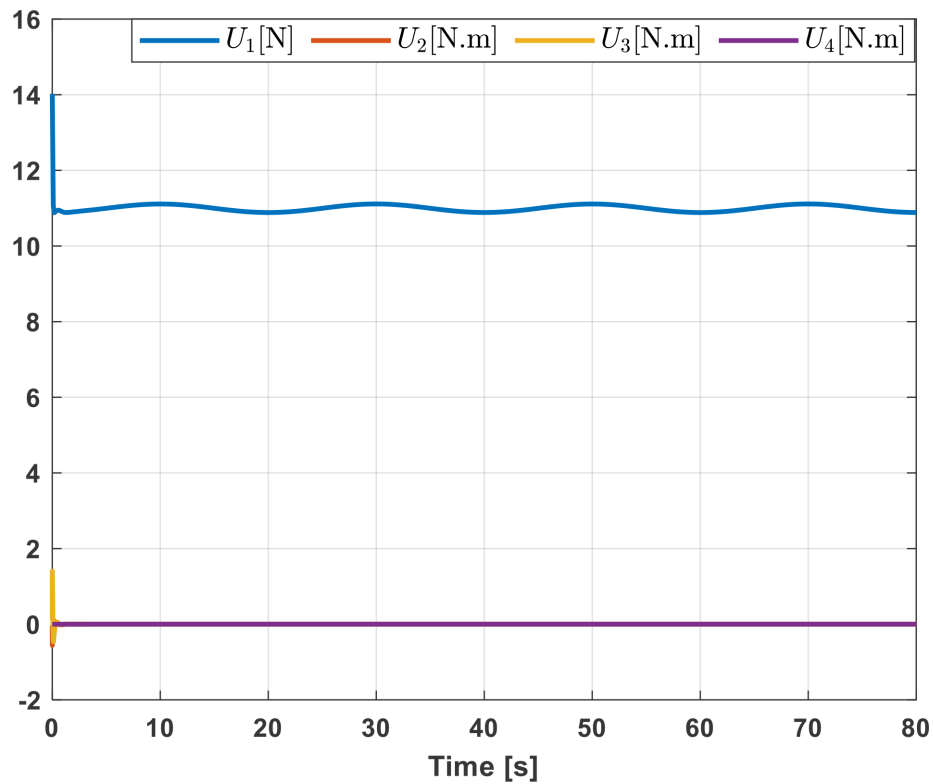


Figure 7. FSMC input response

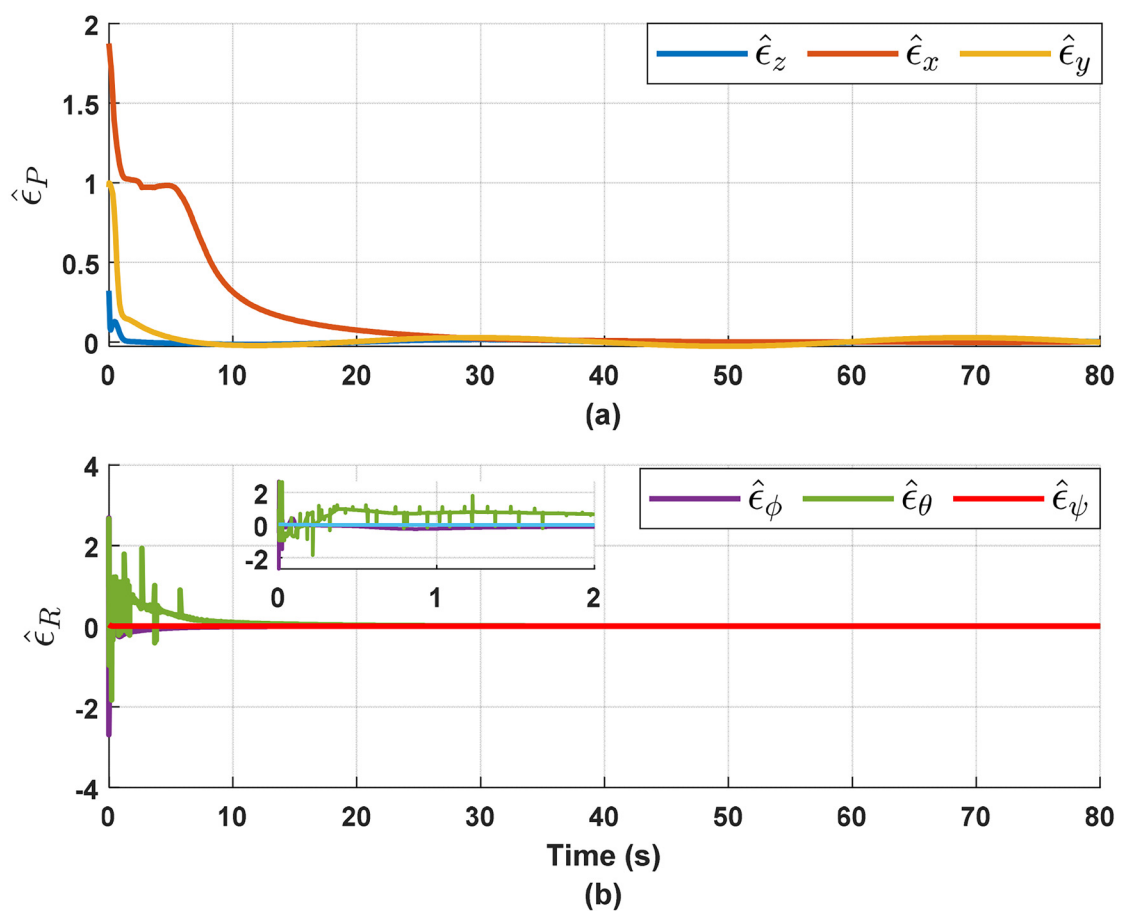


Figure 8. Fuzzy parameters

action during large-error phases, reducing chattering and control effort while preserving tracking accuracy and reliable trajectory tracking.

4.2. Comparative Performance. For the tracking capability evaluation, the proposed FSMC was compared with both a conventional PID controller and a feedback linearization (FL) scheme. As shown by the 3D trajectories in Figure 4, all controllers can follow the reference path; however, the PID controller's response presents a larger initial deviation (overshoot-like transient), slower convergence, and residual steady-state tracking error. The FL approach improves tracking relative to the PID approach; however, it still exhibits a noticeable transient excursion before settling. By contrast, the proposed FSMC yielded a smoother transient response with tighter trajectory adherence throughout the motion.

To further assess the performance, tracking errors were quantified using standard error performance indices, as reported in Table 3. The proposed FSMC achieves lower values across all indices than both the PID controller and FL approach, indicating improved tracking accuracy. The percentage reductions in these indices, also summarized in Table 3, highlight the considerable performance advantage of the proposed method over the PID and FL approaches.

4.3. Study Limitations. Although the proposed FSMC shows better tracking performance in the simulation results, several limitations should be recognized. First, the validation of the controller is based only on numerical simulations using a mathematical model of the quadrotor. Even though bounded disturbances and uncertainties are included, real-world conditions may introduce additional challenges. Factors such as unmodeled aerodynamic effects, actuator limitations, sensor noise, time delays, and hardware nonlinearities could influence the actual performance of the controller. Second, the fuzzy membership functions and rule base were selected and tuned based on expert knowledge and simulation trials. While this approach provides satisfactory results, a more systematic optimization or adaptive tuning method was not investigated in this study. Third, the comparative study is limited to PID and feedback linearization controllers. Although these methods are commonly used for comparison, including other advanced nonlinear or adaptive control strategies could provide a more comprehensive evaluation. Finally, experimental validation on a real quadrotor platform has not been conducted. Future work will focus on hardware implementation to further verify the robustness and practical applicability of the proposed control scheme.

5. CONCLUSIONS

This paper proposes a fuzzy sliding mode control (FSMC) strategy to achieve robust position and attitude tracking for a quadrotor under external disturbances. The controller combines sliding mode control with a fuzzy logic system that adjusts the switching gain in a continuous manner. This approach helps to reduce chattering while maintaining strong robustness against uncertainties. The stability of the closed-loop system is demonstrated through Lyapunov theory, verifying the global exponential convergence of both the sliding surfaces and the tracking errors. The simulation results demonstrate that the proposed FSMC achieves smoother

control signals, better disturbance rejection, and higher tracking accuracy compared with the conventional PID and feedback linearization methods. The results of this study suggest that adaptive adjustment of the switching gain through fuzzy inference can significantly enhance the performance of nonlinear control systems. In particular, integrating intelligent tuning mechanisms into sliding mode control offers a practical solution to achieve a balance between robustness and smooth control action. This method can also be applied to other underactuated and nonlinear systems operating in uncertain conditions. However, the current study is limited to numerical simulations based on a mathematical model of the quadrotor. Future work should focus on implementing the proposed controller on a real quadrotor platform to examine its performance in practical environments. In addition, systematic optimization techniques could be used to improve the design of the fuzzy membership functions and rule base. Further comparisons with other advanced nonlinear and adaptive control strategies would provide a more comprehensive assessment. These developments would enhance the practical applicability and reliability of the proposed control approach.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Mansour Abdullah Nasrullah – Department of Control & Instrumentation Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0009-0000-4172-5348
Email: s202278820@kfupm.edu.sa

Corresponding Author

Syed Muhammad Amrr Research Mentor, Assistant Professor, Department of Control & Instrumentation Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0000-0001-5084-5472
Email: syed.amrr@kfupm.edu.sa

ACKNOWLEDGMENTS

The authors would like to thank Mr. Abdullah Abushokor, a Ph.D. student at CIE Department, KFUPM, for his valuable assistance in simulations, technical discussions, and valuable suggestions that contributed to the development of this work.

REFERENCES

- (1) Panigrahy S, Karmakar S. Real-time condition monitoring of transmission line insulators using the YOLO object detection model with a UAV. *IEEE Trans Instrum Meas.* 2024;73:1–9.
- (2) Wang Y. Optimizing PID and sliding mode control for quadcopter UAV stabilization. In: *Proc. 2nd Int Conf Image Algorithms Artificial Intelligence (ICIAAI)*. Atlantis Press; 2024. p. 481–493.
- (3) Yang H, Zhang K, Zheng K, Qian Y. Leveraging linear quadratic regulator cost and energy consumption for ultrareliable and low-latency IoT control systems. *IEEE Internet Things J.* 2020;7(9):8356–8371.
- (4) Yesmin A, et al. Sliding mode controller for quadcopter UAVs. *Drones.* 2025;9(9):625.

- (5) Utkin VI. Sliding mode control design principles and applications. *IEEE Trans Ind Electron.* 1993;40(1):23–36.
- (6) Edwards C, Spurgeon SK. Sliding mode control: Theory and applications. London: Taylor & Francis; 1998.
- (7) Slotine JJ, Li W. Applied nonlinear control. Englewood Cliffs (NJ): Prentice-Hall; 1991.
- (8) Young KD, Utkin VI, Ozguner U. A control engineer's guide to sliding mode control. *IEEE Trans Control Syst Technol.* 1999;7(3):328–342.
- (9) Levant A. Higher-order sliding modes, differentiation and output-feedback control. *Int J Control.* 2003;76(9–10):924–941.
- (10) Ding S, Li S, Li Q. Second-order sliding mode control design subject to mismatched disturbances. *Automatica.* 2010;46(6):1018–1025.
- (11) Wang L. A course in fuzzy systems and control. Upper Saddle River (NJ): Prentice-Hall; 1997.
- (12) Sira-Ramírez H, Spurgeon SK. Fuzzy sliding mode control. *Control Eng Pract.* 1999;7(9):1151–1158.
- (13) Li Y, Zheng Q. Adaptive fuzzy sliding mode control for uncertain nonlinear systems. *IEEE Trans Fuzzy Syst.* 2013;21(6):1073–1085.
- (14) Jia K, Lin S, Du Y, Zou C, Lu M. Research on route tracking controller of quadrotor UAV based on fuzzy logic and RBF neural network. *IEEE Access.* 2023;11:111433–111446.
- (15) Abushokor A, Amrr SM. Variable exponent fixed-time optimized sliding mode for quadrotor attitude control. In: *Proc. IEEE IAS Annu Meet (IAS-AM).* Taiwan; 2025. p. 1–6.
- (16) Abushokor A, Amrr SM, Ferik SE, Al-Dhaifallah M. Optimized adaptive variable exponent nonsingular terminal SMC for trajectory tracking of a quadrotor. *J Vib Control.* 2026;Early Access:1–17.
- (17) Chen X, Liu J, Meng L. Fuzzy sliding mode control for trajectory tracking of quadrotor aircraft based on adaptive event-triggered mechanism. *J Franklin Inst.* 2025;362:107718.
- (18) Khalil HK, Grizzle JW. Nonlinear systems. Vol. 3. Upper Saddle River (NJ): Prentice Hall; 2002.
- (19) Kuang J, Chen M. Adaptive sliding mode control for trajectory tracking of quadrotor unmanned aerial vehicles under input saturation and disturbances. *Drones.* 2024;8(11):614.

Resource-Constrained Event Scheduling: An Optimization Approach

Shimaa Alwadai and Yasser Almoghathawi*

Cite <https://doi.org/10.64589/juri/218988>

Submitted: November 27, 2025 Revised: January 07, 2026 Accepted: January 28, 2026

ABSTRACT

Managing large-scale event schedules poses several challenges for decision makers, including limitations and considerations such as available resources, venues, and audiences. To address this complexity, we developed a multi-objective mixed-integer programming optimization model for resource-constrained event scheduling over a multi-day planning horizon. The proposed model aims to determine the optimal location assignment for all events and the starting day for each event within the event season. Additionally, the model accounts for class-based event structures, venue utilization, and seasonal continuity. A lexicographic multi-objective approach is adopted, where the primary objective is to maximize overall profit and the secondary objective is to balance venue utilization. The effectiveness of the proposed model is demonstrated by utilizing a dataset consisting of 24 different events with five audience types scheduled over a 20-day season. Based on our findings, we also highlight potential directions for future research.

Keywords: scheduling, lexicographic, mixed-integer, optimization, resource-constrained, event

1. INTRODUCTION

Recently, as part of the Vision 2030 initiative, Saudi Arabia has demonstrated a significant increase in the number and scale of events, conferences, and international exhibitions, particularly within the entertainment sector. This rapid growth has increased the need for efficient event scheduling, especially for large-scale seasonal events such as the Riyadh Season, EXPO 2030, and the 2034 FIFA World Cup. Coordinating multiple events while considering the availability of limited resources, diverse audience types, and daily timelines requires careful planning and scheduling to ensure that all objectives are achieved.

Managing event schedules at such a large scale introduces several challenges for decision makers. Accordingly, event organizers must allocate multiple constrained resources, including venues that vary in capacity, suitability, and the types of events they can support. Furthermore, they must account for different types of audiences to ensure that individuals who wish to attend several related events are not forced to choose between overlapping events, which could result in significant revenue losses. Such conflicts become even more critical during long event seasons, in which the overarching goal is to maintain continuity and avoid idle days, thereby ensuring consistent audience engagement and maximizing overall seasonal performance.

To address these problems, we developed a multi-objective mixed-integer programming model for resource-constrained event scheduling over a multi-day (i.e., seasonal) planning horizon. The proposed model considers several key features, including the incorporation of a class-based event structure in which events are categorized according to their importance to a certain type of audience, thereby preventing events within the same

class from occurring on the same day. Additionally, venue feasibility is captured by the requirement that the expected attendance of an event should reach at least 75% of a venue's capacity for that venue to become a potential candidate location for hosting that specific event. Furthermore, the proposed model enforces seasonal continuity to eliminate undesirable gaps (i.e., time between events) during the season. A lexicographic multi-objective method is adopted, where the primary objective is to maximize overall profit and the secondary objective is to balance the utilization of venues to avoid excessive reliance on a few locations while others remain underused.

The remainder of this paper is organized as follows. Section 2 reviews relevant scheduling literature and studies related to the proposed model. Section 3 details the proposed optimization model. Section 4 presents numerical results and discusses key insights. Finally, concluding remarks are presented in Section 5, along with promising directions for future research.

2. LITERATURE REVIEW

Recent research on resource-constrained scheduling has highlighted the challenges of allocating activities when time, eligibility, and resource constraints must be satisfied simultaneously. Naderi et al. (2022) extended the classical resource-constrained project scheduling problem (RCPSP) by incorporating a reliability function for renewable resources into a mixed nonlinear model that was linearized and solved using both exact and genetic metaheuristic approaches. Their work addressed multi-objective tradeoffs; however, it focused on project activities, rather than

large-scale seasonal event planning. Other studies have examined dual-resource environments. Kimura et al. (2024) proposed a decoding-based solution for the dual-resource-constrained scheduling problem, demonstrating how operator ability and eligibility influence feasible allocations. Although their work is relevant to conflict structures, they did not address multi-day event occupancy or audience-based conflicts.

The multi-objective aspects of scheduling have also received considerable attention. Su et al. (2025) developed a reactive, multi-objective, multi-skilled project-scheduling model that can respond to resource disruptions and dynamic updates. Their model demonstrates how skill and resource variability affect the scheduling process but does not consider event duration, venue eligibility, or lexicographic priority rules. At the methodological level, lexicographic optimization has been examined in several domains. Lai et al. (2020) introduced a mixed Pareto–lexicographic framework based on priority chains, revealing how strict hierarchical priorities can be embedded within multi-objective optimization while preserving Pareto efficiency across other criteria. Sun et al. (2024) applied lexicographic ordering to an event-triggered model predictive control framework for ship path following, where primary objectives were optimized first, and then secondary objectives were addressed within the restricted feasible region. In the context of production planning, Demir (2023) developed an efficient lexicographic solution method for a multiport fabric dyeing scheduling problem, offering exact and heuristic strategies for handling hierarchical objectives. Although these studies demonstrate the application of lexicographic structures, they are not applicable to resource-constrained event scheduling.

The studies mentioned above cover reliability-based RCPSP, dual-resource conflicts, multi-skilled scheduling, and lexicographic decision structures; however, they do not combine audience type conflicts, class-based event hierarchies, all-day seasonal coverage, location capacity eligibility, and a lexicographic multi-objective approach in a single framework. In this study, motivated by these gaps, focused on the development of an optimization model for resource-constrained event scheduling that incorporates real-world conflicts and supports multiple objectives for large-scale event planning.

3. OPTIMIZATION MODEL

In this section, we detail the proposed optimization model for the resource-constrained event-scheduling problem. Our model is formulated as a mixed-integer program that aims to determine the optimal location assignment for all events and the starting day for each event within a season. We adopt the following assumptions: (i) all events belong to one main season or a large-scale event, (ii) each location can hold only one event per day, (iii) visitors attend an event from its starting day until the end of the event duration, and (iv) all events follow the same ticket pricing for the same event days (i.e., the price per day of an event is the same throughout the duration of the event). Without loss of generality, ticket prices can be treated as adjustable input parameters that vary across events or days to reflect demand fluctuations driven by factors such as peak periods (e.g., weekends and holidays) or high-demand conditions (e.g., trendy events or the presence of

Table 1. Sets and parameters in the proposed optimization model

E	Set of events
L	Set of available locations
D	Set of days, planning horizon, $D = \{1, \dots, \delta\}$
p_e	Ticket price per person for event $e \in E$
a_e	Audience type for event $e \in E$
b_e	Class of event $e \in E$
α_e	Expected number of attendances for event $e \in E$
q_e	Duration of event $e \in E$
c_l	Capacity of location $l \in L$
t_l	Type of location $l \in L$
β_{el}	Binary indicator equal to one if event $e \in E$ can be held at location $l \in L$ and equal to zero otherwise

famous performers), and the model can be adjusted accordingly. However, in this study, a uniform pricing structure was adopted to focus on core scheduling decisions.

3.1. Notation. The sets and parameters considered in the proposed optimization model are listed in Table 1, while decision variables are listed in Table 2. The assignment and timing of events across the planning horizon are represented by a set of mixed-integer decision variables. The starting day decision for each event $e \in E$ is captured by the binary variable x_{el}^d , which is equal to one if event e begins on day $d \in D$ at location $l \in L$, and equal to zero otherwise. Because an event occupies a location for the entire duration of its schedule, the daily usage of each location is represented by the binary variable y_{el}^d , which indicates whether event e is taking place at location l on day d . To ensure consistency between these two decision variables, the integer variable z_e determines the starting day of event e , thereby guaranteeing that event occupancy aligns with the assigned duration. Consequently, these variables enable the model to accurately reflect how events utilize locations over time, capture real-world conflicts among events specifically aimed at similar classes or audience types, and enforce the daily operational structure of the season.

Table 2. Decision variables in the proposed optimization model

x_{el}^d	Binary variable equal to one if event $e \in E$ starts on day $d \in D$ at location $l \in L$ and equal to zero otherwise
y_{el}^d	Binary variable that is equal to one if event $e \in E$ is occupying location $l \in L$ on day $d \in D$ and equal to zero otherwise
z_e	Integer variable indicating the starting day of event $e \in E$
v_l^+	Continuous non-negative variable representing the positive deviation of the usage of location $l \in L$ above the average usage
v_l^-	Continuous non-negative variable representing the negative deviation of the usage of location $l \in L$ below the average usage

3.2. Mathematical Model. The mathematical model described below provides a structured representation of the proposed scheduling model by defining how events interact with locations and planning horizons. This formulation represents the logical relationships connecting the starting day of an event with its occupancy over time, thereby ensuring consistency throughout the season. The model also includes structural constraints that reflect operational rules such as location eligibility requirements and conflict relationships between events of different classes or audience types. By combining these elements, the proposed model captures the complete behavior of the targeted scheduling system. The detailed objective functions and constraints controlling these relationships are presented in the following subsections.

3.2.1. Objective functions. The proposed model has multiple objective functions: (i) maximizing the expected revenue from all events, and (ii) minimizing the total imbalance in location usage. Therefore, we adopt a lexicographic multi-objective optimization approach to prioritize scheduling objectives in a strictly hierarchical manner. In a lexicographic structure, the primary objective is optimized independently first, and the secondary objective is considered strictly after obtaining the optimal value of the primary objective, which is then considered as a constraint (Fiaschi and Cococcioni 2022, Demir 2023 and Sun et al. 2024).

The first objective of the proposed model is to maximize the expected revenue generated from all scheduled events. Revenue is obtained when event $e \in E$ is assigned to location $l \in L$ on day $d \in D$, and is calculated by multiplying the ticket price per person by the expected number of attendees for that event, as expressed in Eq. (1).

$$\text{Max } Z^* = \sum_{e \in E} \sum_{l \in L} \sum_{d \in D} \alpha_e p_e y_{el}^d \quad (1)$$

After determining the optimal revenue value Z^* , this value is fixed as an additional constraint to ensure revenue optimality. The model is then reoptimized to address the secondary objective, which minimizes the total imbalance in location usage. Imbalance is defined as the sum of the positive deviations v_l^+ and negative deviations v_l^- from the average usage of each location $l \in L$, as shown in Eq. (2).

$$\text{Min } \sum_{l \in L} (v_l^+ + v_l^-) \quad (2)$$

This lexicographic structure guarantees that maximizing revenue remains the top priority, while the secondary objective enhances the fairness of location utilization.

3.2.2. Constraints. The constraints in the proposed model ensure that events are scheduled feasibly and consistently across the planning horizon. Each constraint reflects a key requirement, as outlined below.

Constraint (3) indicates that each event must be scheduled no more than once, implying that an event must have at most one starting day at one location. This constraint prevents the assignment of a single event to multiple locations or starting days.

$$\sum_{l \in L} \sum_{d \in D} x_{el}^d \leq 1, \quad \forall e \in E \quad (3)$$

Constraint (4) ensures that each location hosts a maximum of one event on a given day. Because a venue cannot physically accommodate multiple events simultaneously, this condition prevents the overlapping usage of the same location.

$$\sum_{l \in L} y_{el}^d \leq 1, \quad \forall e \in E, d \in D \quad (4)$$

Constraint (5) ensures that each day in the planning horizon hosts at least one event across all available locations. This constraint prevents idle days during the season, leading to the continuous engagement of visitors and enhancing the overall attractiveness of the event schedule.

$$\sum_{e \in E} \sum_{l \in L} y_{el}^d \geq 1, \quad \forall d \in D \quad (5)$$

Constraint (6) ensures that the starting day recorded for each event accurately reflects the day on which the event is scheduled to begin at its assigned location.

$$z_e = \sum_{l \in L} \sum_{d \in D} d x_{el}^d, \quad \forall e \in E \quad (6)$$

Constraint (7) ensures that no more than one event targeting the same audience type is scheduled on the same day, thereby preventing overlapping events targeting the same attendees and ensuring that the individuals interested in each event category have the opportunity to attend all relevant events without facing conflicts.

$$\sum_{d \in D} y_{el}^d \leq 1, \quad \forall a_e, e \in E, l \in L \quad (7)$$

Constraint (8) ensures that at most one event belonging to a given event class is scheduled on any day, preventing the splitting of the intended audience for any event class and increasing the revenue stream.

$$\sum_{d \in D} dy_{el}^d \leq 1, \quad \forall a_e, e \in E, l \in L \quad (8)$$

Constraint (9) states that an event cannot appear in the schedule on a certain day unless it starts on that day, ensuring that each event may continue running throughout its assigned duration.

$$y_{el}^d \leq \sum_{n=d}^{d+q_e-1} z_{el}^n, \quad \forall e \in E, l \in L, d \in D \quad (9)$$

Constraint (10) ensures that if event e begins at location l on day d ($x_{el}^d = 1$), then the location must remain occupied by the same event for every day n of its duration.

$$y_{el}^n \geq x_{el}^d, \quad \forall e \in E, l \in L, n \in (d, d + q_e), n \leq \delta \quad (10)$$

Constraint (11) is added after solving the first optimization stage to obtain the maximum possible revenue Z^* . This constraint fixes the revenue at the optimal value, ensuring that no

solution found in the second stage will reduce the revenue.

$$\sum_{e \in E} \sum_{l \in L} \sum_{d \in D} x_{el}^d p_e a_e \geq Z^* \quad (11)$$

Constraints (12) and (13) define the usage structure required for the secondary objective, which aims to balance the utilization of locations across seasons. Constraint (12) computes the total number of occupied days for each location $l \in L$ by summing all days on which each location is used by any event. Constraint (13) determines the average usage across all locations by dividing the total number of event days by the number of available locations, thereby providing a benchmark against which deviation can be measured.

$$U_l = \sum_{e \in E} \sum_{d \in D} y_{el}^d, \forall l \in L \quad (12)$$

$$\bar{U} = \frac{\sum_{e \in E} q_e}{|L|} \quad (13)$$

Constraint (14) introduces the deviation variables v_l^+ and v_l^- , linking them to the difference between the actual usage of each location and the corresponding average usage. This relationship ensures that positive and negative deviations are correctly captured and used to optimize the secondary objective.

$$U_l - \bar{U} = v_l^+ - v_l^-, \quad \forall l \in L \quad (14)$$

Constraints (15) to (17) define the data domains of the decision variables in the proposed model.

$$x_{el}^d, y_{el}^d \in \{0, 1\}, \quad \forall e \in E, l \in L, d \in D \quad (15)$$

$$z_e \in \mathcal{Z}, \quad \forall e \in E \quad (16)$$

$$v_l^+, v_l^- \geq 0 \quad \forall l \in L, \quad (17)$$

3.2.3. Parallel events of the same class. In many large-scale seasonal festivals, including Formula-E fan zones, entertainment seasons, or Expo-style events, it is often desirable to allow parallel events within the same class if they target different audience types. In such scenarios, the strict formulation of Constraint (8), which prohibits more than one event within the same class from occurring on the same day, may be overly restrictive.

To address this limitation, we introduce Constraint (18), which is a relaxed version of Constraint (8), in which the non-parallelism rule for classes is applied within class-audience combinations, rather than across the entire class. To formalize this relaxation, we define the set $E_{b,a} = \{e \in H | b_e = b, a_e = a\}$, which represents events belonging to class b and targeting audience a . Under this structure, the model continues to prevent two events belonging to the same class and targeting the same audience group from occurring simultaneously, while allowing parallel events within the same class if they target different audiences. This approach enables richer daily schedules and improves operational flexibility.

$$\sum_{l \in L} \sum_{e \in E_{b,a}} y_{el}^d \leq 1, \quad \forall b, a, d \in D \quad (18)$$

4. RESULTS AND DISCUSSION

This section presents the experimental results of the proposed scheduling model using the constructed event dataset and defined constraints.

4.1. Data. To demonstrate the effectiveness of the proposed optimization model, the model was tested on a dataset consisting of 24 events with five different target audiences, scheduled across a 20-day season. Each event has a defined duration of two days, an audience group, a class category (A, B, or C), and an expected attendance level. Four locations were considered in our experiment, and they differed in both capacity and type, which introduced eligibility constraints regarding where each event could be assigned. The locations belonged to three main types, namely stadium, hall, and park, with capacities of 100, 80, and 90, respectively. Each event could only be scheduled at locations matching its allowed type.

For consistency, all ticket prices and expected attendance values were assumed to be constant when solving the proposed model. Therefore, our focus remains on understanding scheduling behavior driven by conflicts, feasibility rules, and the lexicographic objective structure. Furthermore, this dataset enabled us to evaluate how the proposed model handles capacity limits, class and audience conflicts, location eligibility, and multi-objective tradeoffs over the planning horizon (i.e., a 20-day season).

4.2. Experiment. The proposed resource-constrained scheduling optimization model was solved using the PuLP optimization library in Python. All computational experiments were conducted using a laptop with an 11th-Gen Intel® Core™ i5-1135G7 CPU running at 2.40 GHz, 16 GB of RAM, under the Windows 11 (64-bit) OS. The optimal solution to the proposed model was found with a CPU time of 4.38 s. The process of deriving the global optimal solution using the parameters listed in Table 3 began by generating multiple optimal schedules. First, the model considers the highest-priority objective function to maximize profit without considering balancing location utilization across seasons. The resulting schedules are presented in Figure 1, highlighting solution feasibility and objective-driven profit behavior.

After identifying all optimal event schedules satisfying the model constraints and maximizing the primary objective of profit, we introduce balancing constraints to ensure a fair distribution of events across locations throughout the season. Accordingly, the second-stage optimization searches for the most balanced assignment, subject only to additional balancing constraints (11–14), within the subset of solutions optimized with respect to profitability. This structure represents the direct application of lexicographic multi-objective optimization.

The secondary objective is to minimize the imbalance in location utilization throughout the season, ensuring that each location hosts approximately the same number of events. This objective promotes a more even distribution of operational load and improves the overall structure and practicality of the seasonal schedule. The final schedule, satisfying both objectives and all model constraints, is illustrated in Figure 2, which clearly shows the allocation of events across days and locations.

To further evaluate the effect of the proposed recovery strategy, the optimization model was re-solved using the relaxed

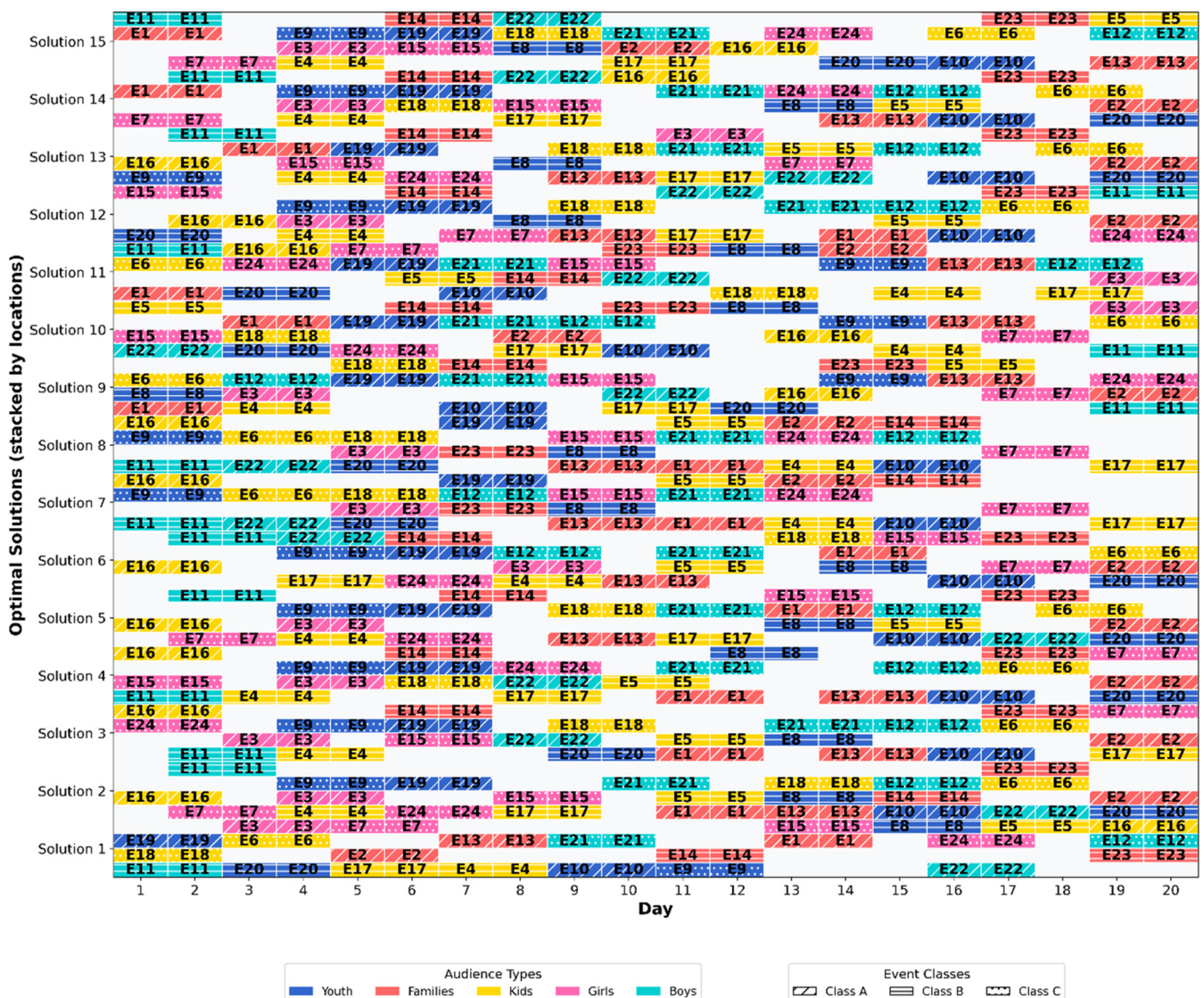


Figure 1. All optimal event schedules under the profit-driven objective

formulation in Constraint (18), which replaces the original class-level non-parallelism rule in Constraint (8). As observed in Figure 3, the relaxed constraint allows several parallel activities that were not feasible in the baseline schedule while still satisfying all resource, location, and audience constraints.

The comparison between the baseline and recovery schedules highlights the benefits of relaxing overly restrictive scheduling rules. By allowing events of the same class to occur in parallel when targeting different audiences, the recovery formulation increases scheduling flexibility and produces a richer and more diverse event program while maintaining all feasibility conditions. This demonstrates how controlled constraint relaxation can improve the practical applicability of optimization-based event scheduling models for large-scale seasonal events.

4.3. Discussion. The computational results of our model demonstrate that it can build a feasible event schedule that (i) assigns events on all days of the season, (ii) respects both audience and class conflicts, (iii) assigns events to the proper location

type and utilizes at least 75% of the location capacity, and (iv) keeps location utilization relatively balanced while maximizing the expected ticket revenue. This structure is consistent with event-based formulations that have been successful in resource-constrained project scheduling; however, our application focuses on seasonal event planning with different venue types and audience groups instead of traditional project activities. For example, Koné et al. (2011) developed event-based mixed-integer linear programming models for the RCPSP and demonstrated that event-based time representations can provide strong formulations with good computational performance. Our model follows the same philosophy by defining binary variables that capture when an event occupies a location on a specific day. However, we extend this concept to handle multi-day events, audience conflicts, and the requirement for full seasonal coverage.

Compared with the classical RCPSP variants summarized by Hartmann and Briskorn (2010), and broader surveys on project scheduling and multimode extensions by Weglarz et al. (2011), our study has a different focus. Most RCPSP studies optimize the

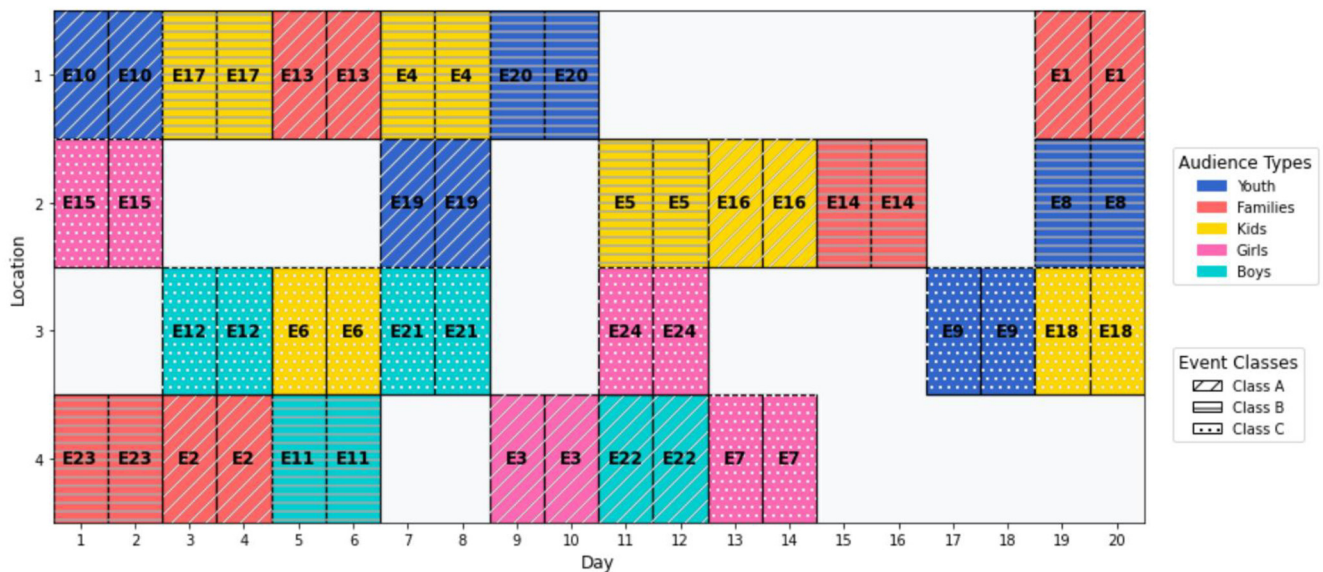


Figure 2. Optimal event schedule obtained from the baseline model

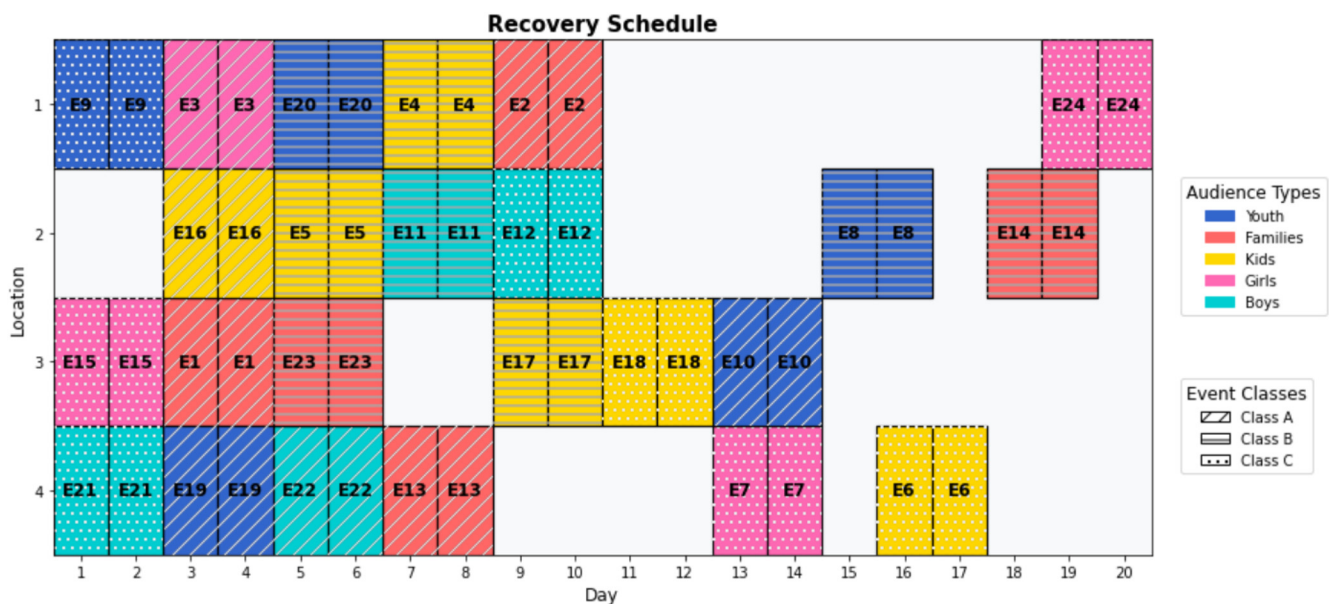


Figure 3. Optimal event schedule obtained from the recovery model

project makespan or cost, considering activities that must respect precedence relations and renewable resource limits. In our context, events do not follow a precedence network but are linked through audience type conflicts, class-based importance levels, capacity eligibility rules, and the requirement that every day in the season must host at least one event. The constraints that prevent parallel events of the same class or those targeting the same audience on the same day also have no direct counterpart in the standard RCPSP variants surveyed in the literature.

At the modeling level, our work is related to extensions from the dynamic scheduling perspective discussed by Vanhoucke (2012), where more complex constraints and project control aspects are added to the basic RCPSP. However, previous studies focused on projects, rather than public events with heterogeneous audiences and venues. Our class-based constraints can be considered as a practical version of priority rules between subsets

of activities; however, they are directly motivated by the need to protect high-importance events from internal competition within the same class. Recent studies on multimode and stochastic RCPSP highlight the importance of uncertainty and complex resource structures. Gao et al. (2024) proposed a metaheuristic-oriented iterative greedy algorithm for a multimode RCPSP under uncertainty and demonstrated that hybrid exact heuristic approaches can handle large instances with uncertain durations and resource parameters.

Beyond classical RCPSP formulations, many allocation and scheduling problems are based on direct revenue considerations, implicitly assuming that prioritizing the most profitable option at each decision point leads to near-optimal outcomes. Ferguson et al. (2011) studied disposition decisions for product returns and demonstrated that many firms rely on margin-based rules that always prioritize the highest-profit option. Their analysis

Table 3. Sets and parameters used in the experiment

E	$\{1, 2, 3, \dots, 24\}$
L	$\{1, 2, 3, 4\}$
D	$\{1, 2, 3, \dots, 20\}$
p_e	100 SAR
a_e	Families: $\{1, 2, 13, 14, 23\}$, Kids: $\{4, 5, 6, 16, 17, 18\}$, Girls: $\{3, 7, 15, 24\}$, Boys: $\{11, 12, 21, 22\}$
b_e	Class A: $\{1, 2, 3, 10, 12, 14, 16, 18, 20, 22\}$, Class B: $\{4, 5, 8, 11, 13, 17, 19, 23, 24\}$, Class C: $\{6, 7, 9, 15, 21\}$
α_e	80 attendees
q_e	Two days
t_l	Location 1: Stadium, Locations 2 & 4: Hall, Location 3: Park
c_l	Location 1: 100, Locations 2 & 4: 80, Location 3: 90
β_{el}	1 for the following events that can be held in location type: Stadium: $\{1, 4, 7, 9, 10, 11, 13, 17, 18, 20, 22, 24\}$ Hall: $\{2, 3, 6, 8, 11, 14, 15, 16, 18, 19, 20, 22, 23\}$ Park: $\{1, 3, 5, 6, 7, 9, 11, 12, 13, 18, 19, 20, 21, 23, 24\}$

revealed that such profit-only objectives can lead to inefficiencies, with average profit losses of approximately 7% compared with more structured optimization approaches and even higher losses over longer planning horizons. Although their application differs from seasonal event scheduling, the underlying decision logic is comparable to that of scheduling events with the goal of maximizing ticket revenue. Although such profit-focused approaches are attractive owing to their simplicity and ease of implementation, they typically overlook the importance of operational interactions, which can limit their practical effectiveness in large-scale and complex scheduling environments.

According to the value-focused thinking paradigm, decision makers organize objectives hierarchically by distinguishing between (i) fundamental objectives, which capture the core values of the decision problem, and (ii) objectives that describe how these values can be achieved (Keeney, 1994). In large-scale event-planning contexts, economic objectives such as profit maximization are commonly identified as fundamental objectives, whereas operational performance measures serve as means objectives. This theoretical foundation supports the adoption of a hierarchical objective structure in which profit is prioritized over secondary operational considerations. For large-scale problems, alternative approaches such as metaheuristic or hybrid solution frameworks can be employed to enhance the quality of derived schedules.

5. CONCLUSIONS

This paper presented an event-based optimization model for resource-constrained seasonal scheduling that integrates class-based conflict rules, audience compatibility conditions, and a

lexicographic structure to balance multiple objectives. The proposed mathematical formulation leverages mixed-integer programming to schedule events by maximizing total seasonal profit while reducing the imbalance in location utilization. The results demonstrated that including structured conflict rules, particularly within each class, improves the practicality of scheduling by avoiding audience overlap and maintaining event diversity across days.

Beyond numerical improvements, the proposed model offers a flexible foundation for planners to explore different scheduling priorities, including fairness, utilization balance, and revenue maximization, which are considered to be a valuable feature for large-scale seasons, where operational constraints, stakeholder profit expectations, and audience segmentation must be balanced effectively.

Future studies could extend this work by incorporating the uncertainty of event attendance to examine how demand-related factors such as marketing campaigns or weather conditions might affect scheduling decisions within the same modeling framework. Furthermore, heuristic- or metaheuristic-based solutions could be developed to handle large-scale event portfolios in a timely manner. The proposed model could also be applied to real-world events such as the Riyadh Season or FIFA World Cup datasets to validate its performance further and highlight its potential practical impact.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Shimaa Alwadai – Department of Industrial & Systems Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0009-0005-7422-9611
Email: S202181790@kfupm.edu.sa

Corresponding Author

Yasser Almoghathawi – Research Mentor, Associate Professor, Department of Industrial & Systems Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0000-0003-0527-4820
Email: moghathawi@kfupm.edu.sa

ACKNOWLEDGEMENTS

The authors acknowledge the support provided by the Industrial and Systems Engineering Department at King Fahd University of Petroleum and Minerals (KFUPM). This research was funded by the Undergraduate Research Department at KFUPM. The authors also thank the JURI editorial team for their assistance in reviewing and improving this manuscript.

REFERENCES

- (1) Sun, Y., Cao, Y., Shan, Q. & Zhang, H. Ship path following control using event-triggered lexicographic ordering multi-objective model predictive control. Proc. Int. Conf. Intell. Syst. Technol. (ICIST), 268–274 (2024).

- (2) Demir, Y. An efficient lexicographic approach to solve multi-objective multi-port fabric dyeing machine planning problem. *Appl. Soft Comput.* 144, 110541 (2023).
- (3) Naderi, S., Vaez-Ghasemi, M. & Movahedi Sobhani, F. Optimizing resource-constrained project scheduling problem considering the reliability function. *Discrete Dyn. Nat. Soc.* 2022, 7711383 (2022).
- (4) Kimura, M., Ishigaki, A. & Tanabe, N. Decoding scheme considering differences in operator ability for dual resource-constrained scheduling problem. *Proc. Int. Conf. Intell. Inform. Anal.*, 549–555 (2024).
- (5) Su, Y., Xu, Z. & Liu, D. A model and algorithm for reactive multi-objective multi-skilled project scheduling under resource disruptions. *Comput. Ind. Eng.* 203, 111043 (2025).
- (6) Lai, L., Fiaschi, L. & Cococcioni, M. Solving mixed Pareto-lexicographic multi-objective optimization problems: The case of priority chains. *Swarm Evol. Comput.* 55, 100687 (2020).
- (7) Fiaschi, L. & Cococcioni, M. A non-Archimedean interior point method and its application to the lexicographic multi-objective quadratic programming. *Mathematics* 10, 4536 (2022).
- (8) Koné, O., Artigues, C., Lopez, P. & Mongeau, M. Event-based MILP models for resource-constrained project scheduling problems. *Comput. Oper. Res.* 38, 41–50 (2011).
- (9) Hartmann, S. & Briskorn, D. A survey of variants and extensions of the resource-constrained project scheduling problem. *Eur. J. Oper. Res.* 207, 1–14 (2010).
- (10) Weglarz, J., Józefowska, J., Mika, M. & Waligóra, G. Project scheduling with finite or infinite number of activity processing modes – A survey. *Eur. J. Oper. Res.* 208, 177–205 (2011).
- (11) Vanhoucke, M. Project management with dynamic scheduling: Baseline scheduling, risk analysis and project control. Springer (2012).
- (12) Gao, Z., Zhang, L., Yu, P., Zhang, Z. & Li, Z. A matheuristic-oriented iterated greedy algorithm for multi-mode resource-constrained project scheduling problem under uncertainty. *Eur. J. Oper. Res.* (2024).
- (13) Keeney, R. L. Creativity in decision making with value-focused thinking. *Manage. Sci.* 40, 1637–1654 (1994).
- (14) Ferguson, M., Guide Jr., V. D. R., Koca, E. & Souza, G. C. A profit-maximizing approach to disposition decisions for product returns. *Decis. Sci.* 42, 773–798 (2011).

Blocking Humidity Ingression in FAPbI₃ Thin Film for Perovskite Solar Cell Application

Ghadah Khalid Mobaireek, Muhammed P.U. Haris, and Amir Al-Ahmed*

Cite <https://doi.org/10.64589/juri/218959>

Submitted: November 27, 2025 Revised: January 13, 2026 Accepted: January 28, 2026

ABSTRACT

Owing to the increased global need for renewable energy, cost effect solar energy generation has become key to energy sustainable. Among the new thin film solar cells perovskite solar cells are considered as the suitable alternatives to the market giant c-Si based technology. Among the different perovskites FAPbI₃ is the most popular one for their better efficiency, cost-effectiveness, thermal stability, direct bandgap, and improved light absorption compared to other perovskites. However, the commercialization of FAPbI₃ is hindered by its phase shift from the photoactive cubic to the photo-inactive hexagonal phase as well as stability issues in humid conditions. This study aims to overcome the limitations of humidity effect on this perovskite layer using a multifunctional additive engineering strategy. 2-Aminoethylthiol was added to the FAPbI₃ precursor solution and active layer was spin coated on the pre-cleaned FTO and FTO/SnO₂ glass substrates. Through various structural, opto-electronic, and morphological analyses, we found that the additive-modified films showed increased light absorption, photoluminescence emission, and crystallinity by 20%, 400%, and >300%, respectively, promising a higher solar cell current density. The percentage improvements were measured by comparing the peak intensities of the data obtained from the modified film with those obtained from the control film. When these films aged for 200 h at 85% relative humidity and 85 °C temperature, the target film retained photophysical integrity >600% better than the control film, proving their robustness in harsh environments.

Keywords: FAPbI₃ perovskite, thermal stability, humidity ingression, perovskite solar cell, photophysical properties

1. INTRODUCTION

Photovoltaics research has undergone a substantial transformation since the discovery of halide perovskite materials with suitable optoelectronic properties such as low exciton binding energy, direct bandgap, defect tolerance, and high absorption coefficient¹⁻⁴. Perovskite solar cells (PSCs) have recently emerged as a prospective third-generation alternative owing to their rapid advancement and comparable efficiencies to conventional silicon. Since their first trial in 2009, their power conversion efficiency has increased from 3% to over 27% within <15 years⁵⁻⁷.

Currently, FAPbI₃ perovskite demonstrates substantial potential owing to its favorable optoelectronic properties and relatively high efficiency. A significant challenge is posed by its phase instability, particularly its propensity to transition from the favorable cubic (3C) phase to the photo-inactive hexagonal (2H) phase, and subsequently degrade to final PbI₂. This process is further accelerated under humid conditions^{8,9}. In addition to destabilizing the perovskite structure, water vapor adsorption induces morphological changes, such as a weakening of the grains and an increase in defect density which further impedes the charge-transport pathways of solar cells, potentially leading to a decline in their performance^{10,11}.

Therefore, understanding and mitigating these effects is critical for the development of perovskite solar technology. As in Table 1, various strategies including compositional engineering, additive assisted modifications, process engineering, and passivation strategies have been utilized in the literature to prevent the unfavorable phase transition^{6,7,9}. Incorporating long-chain organic cations or Lewis bases into FAPbI₃ precursor solutions has emerged as an effective approach to enhancing its stability. For example, long-chain cations such as phenylethyl ammonium (PEA⁺) enhance particle size and morphology, thereby potentially improving electronic properties. The hydrophobic properties of these additives establish a moisture-resistant barrier, effectively minimizing the likelihood of phase transitions. In addition to regulating the crystallization process, these compounds are crucial for mitigating the effects of temperature and humidity fluctuations¹²⁻¹⁶. Notwithstanding these endeavors, the humidity resistance and photovoltaic potential of perovskites have not been adequately investigated.

In this study, we utilized 2-Aminoethylthiol (AET) as a novel additive in a perovskite precursor solution and investigated its effect on the structural, optoelectronic, morphological, and stability properties of the perovskite layer.

Table 1. Comparison of FAPbI₃ film properties and stability benchmarks in this study and recent high-impact literature

Performance metric	Xiong et al. (2025) ⁷	Zhou et al. (2024) ⁶	Haris et al. (2023) ⁹	This study
Material target	Cl-doped FAPbI ₃	FAPbI ₃ (Buried interface)	Pure FAPbI ₃	FAPbI ₃ thin films
Primary strategy	Alkali metal oxalates	Benzothiadiazole derivatives	Sulfur heteroatoms (Pb-S)	Multifunctional additives
Stability focus	Continuous MPPT / thermal	Ambient storage stability	Water immersion stability	Accelerated damp heat (85/85)
Structural effect	Defect passivation	Energy level regulation	2H to 3C phase shift	>300% Crystallinity increase
Optoelectronic gain	Certified 27.2% PCE	Non-radiative suppression	Improved PL lifetime	+400% PL; +20% Absorption
Humidity resistance	Not the primary metric	1000h (Ambient)	Water-stable	200h at 85% RH / 85 °C

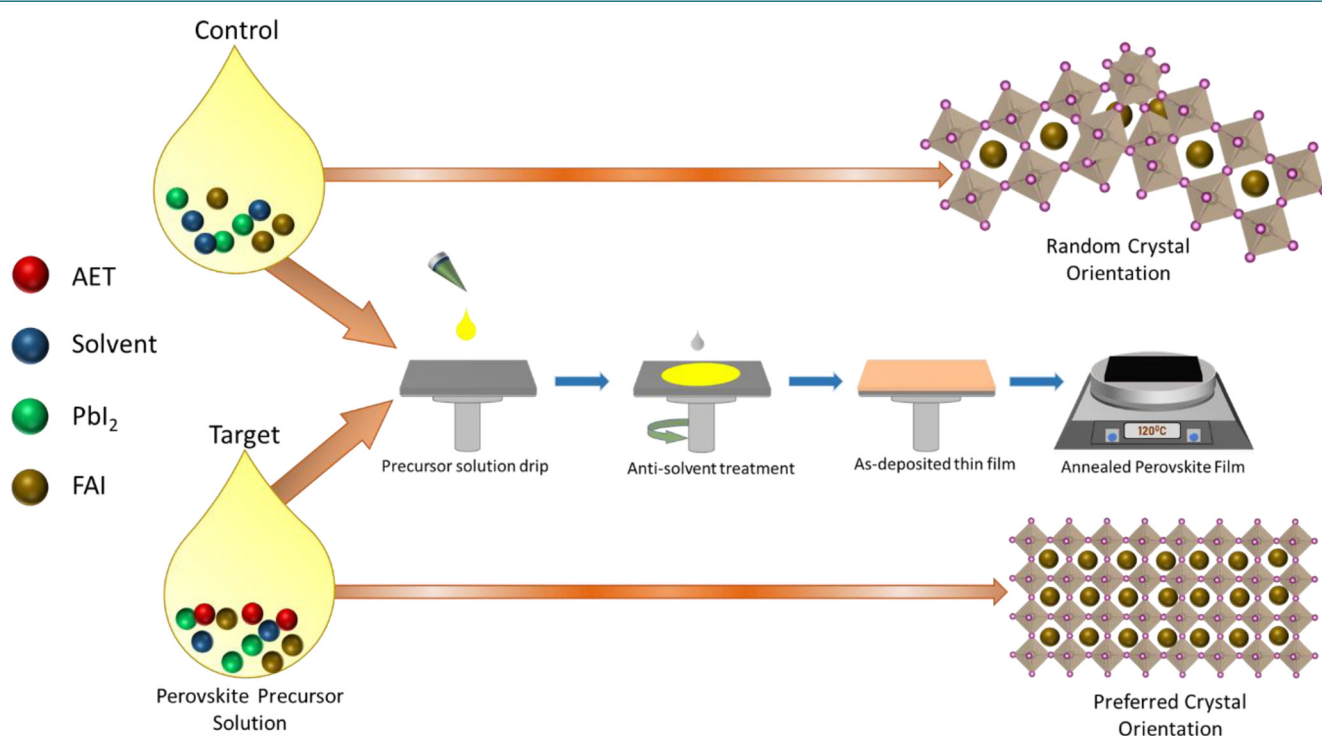
2. METHODOLOGY

The preliminary stage of this investigation entailed meticulous and systematic cleaning of glass substrates. This involved sonication in a soap solution, followed by rinsing with deionized water, acetone, and isopropyl alcohol. The cleaned substrates were treated with ultraviolet (UV)-ozone plasma to eliminate residual organic impurities and enhance their wettability, after which the perovskite precursor solution for FAPbI₃ was prepared by dissolving equimolar quantities of suitable precursor components in a N,N-Dimethylformamide-Dimethyl sulfoxide (DMF-DMSO) solvent blend. The solution was agitated at 60 °C for 2 h to produce a transparent precursor solution. Then, it was placed onto the cleaned glass substrates via spin-coating, after which the spin-coated thin films were annealed at 150 °C for 30 min to achieve full crystallization (Figure-1). All processes were conducted inside a glove box. The perovskite films were then characterized using UV-visible (UV-Vis) and photoluminescence (PL) spectroscopy to evaluate their optical characteristics. Furthermore, structural analysis was conducted by performing powder X-ray diffraction (PXRD) and scanning electron microscopy (SEM) to assess the morphology and grain size; the

crystallinity and phase purity were also analyzed by PXRD. The acquired data were investigated using Origin software to better understand the film quality. The optimized control and modified perovskite thin films were tested for their humidity resilience and thermal stability by applying 85% RH and 85 °C stresses for 200 h.

3. RESULTS AND DISCUSSION

We chose AET as the additive molecule to explore the interaction between the thiol group and the perovskite structure. We prepared the perovskite precursor solution by dissolving synthesized FAPbI₃ perovskite powder in a mixed solvent comprising DMF and DMSO in a 4:1 volume ratio. The precursor solution was subsequently spin-coated onto the substrates and annealed at 120 °C for 30 min to finalize the deposition. The additive molecules were immediately integrated into the precursor solution before deposition for additivization. All preparatory procedures were performed in a controlled glove box to preserve material integrity. The deposited perovskite films were later utilized for further characterization.

**Figure 1.** Schematic representation of thin film deposition procedure and the impact of the additive on crystal orientation

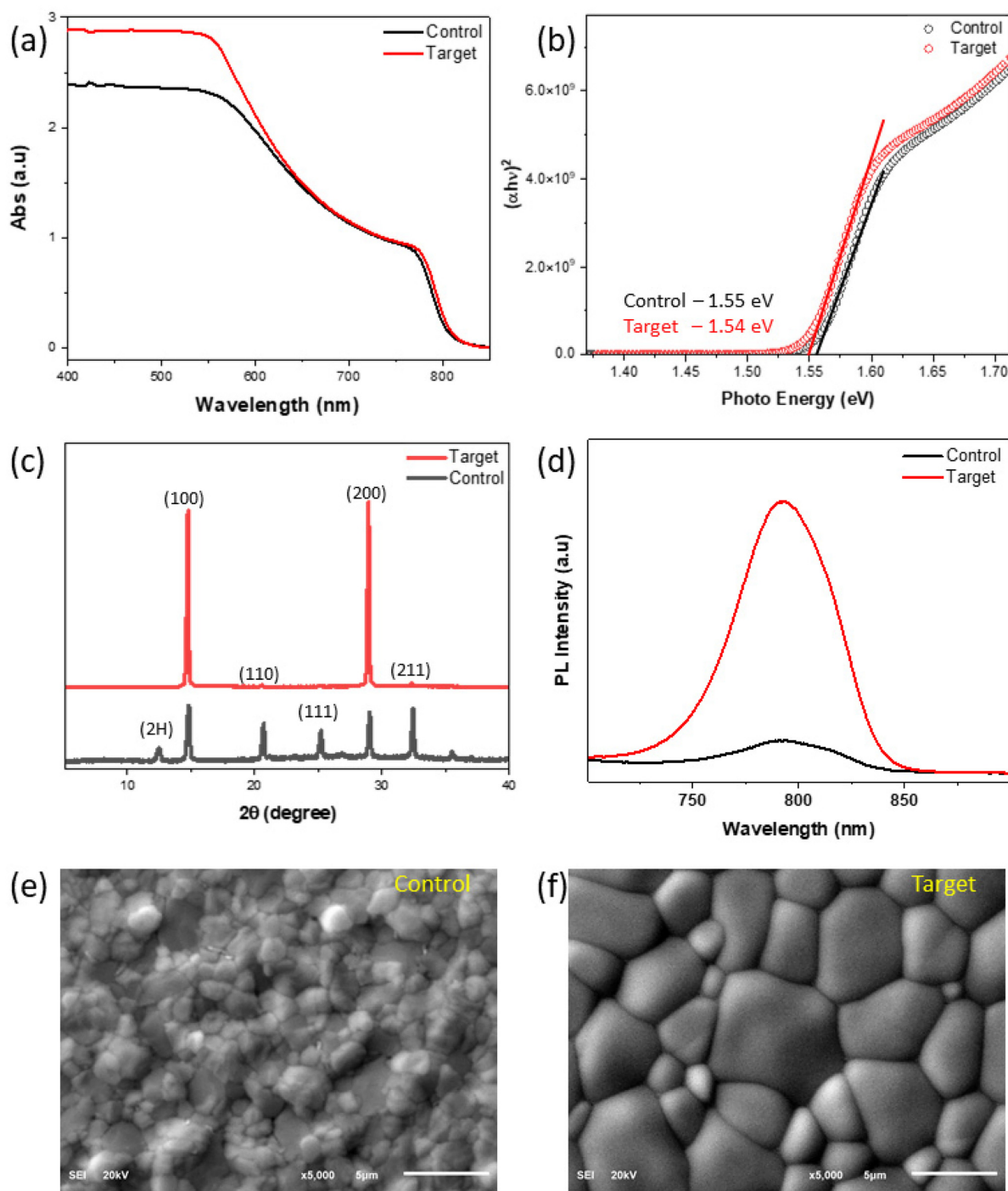


Figure 2. (a) UV-Vis absorption spectra of perovskite thin films; (b) tauc plot calculated from UV-Vis transmission spectra for control and target perovskite thin films; (c) PL spectra for control and target perovskite thin films; (d) PXRD reflections for control and target perovskite films; (e) SEM images for control perovskite thin film; (f) SEM images for target perovskite thin films

3.1. Opto-electronic Analysis. Improved photophysical characteristics are essential for the development of prospective photovoltaic materials. UV-Vis absorption spectra are essential for clarifying the light-absorption characteristics of a material. We performed measurements on both the control FAPbI₃ perovskite thin film (hereinafter referred to as the control) and the additive-modified target perovskite film (hence referred to as the target). Figure 2A shows that the target film had significantly enhanced light absorption in both the visible and UV region compared to

the control film. The increase in light absorption is expected to correlate with improved efficiency of solar cells because higher energy conversion is inherently linked to greater light absorption levels⁹. The modifications observed in the UV absorption spectra indicate that the additive modified the structural characteristics of the perovskite, leading to enhanced light absorption efficiency. This modification is a feasible approach to improving the effectiveness of devices utilizing these modified perovskite sheets, which may lead to more efficient energy conversion processes.

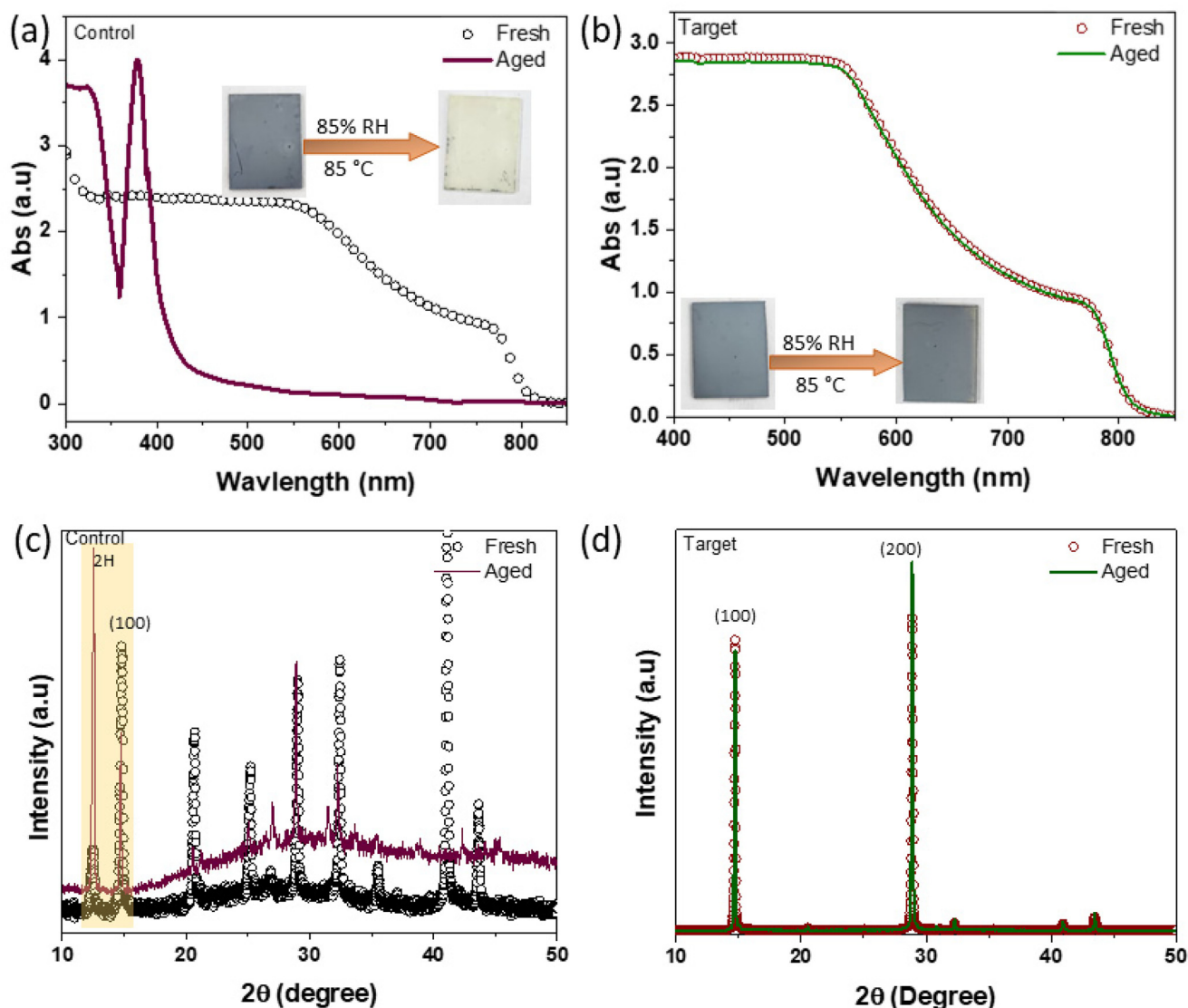


Figure 3. (a) UV-Vis spectra recorded before and after stability test for control perovskite film; (b) UV-Vis spectra recorded before and after stability test for target perovskite film; (c) PXRD diffractogram recorded before and after stability test for control perovskite film; (d) PXRD diffractogram recorded before and after stability test for target perovskite thin films

Furthermore, we calculated the optical bandgap of the perovskite layer from τ_{auc} plot. As shown in Figure 2B, the bandgap was calculated as 1.55 eV and 1.54 eV for the control and target films, respectively, confirming the absorption of more light by the target film. The literature indicates that the peak intensity of the PL emission in perovskite thin films is significantly influenced by the numerous defects present within the perovskite lattice^{6,8,9}. Defects such as vacancies, interstitials, or impurities can create trap states inside the lattice, resulting in nonradiative recombination, which adversely affects the performance of perovskite solar cells. A reduced defect concentration results in less non-radiative recombination and enhanced PL emission. The PL emission spectra obtained from both the control and target perovskite thin films are shown in Figure 2C. The PL emission peak of the control film was observed at approximately 795 nm, whereas the target film displayed an emission peak at approximately 796 nm. Notably, the target film exhibited a substantial

increase in PL emission intensity, indicating a large decrease in lattice defects. The improved light absorption, reduced bandgap, and increased PL emission intensity validated the efficacy of our additive engineering approach for use in perovskite solar cells.

3.2. Structural Evaluation. Structural studies are key to develop perovskite thin films as they reveal their phase purity, crystal orientation, defect density, and degradation potential. This study employed PXRD to evaluate the structural properties of the perovskite thin films, as shown in Figure 2D. The control perovskite film exhibited most of the perovskite (3C) peaks, indexable to $2\theta \sim 14^\circ(100)$, $20^\circ(110)$, $24^\circ(111)$, $28^\circ(200)$, and $35^\circ(211)$. The lower intensities of the (110) and (211) peaks in the target film were due to the AET additive facilitating the transition from random crystal growth to a preferred (h00) orientation. In addition to the perovskite peaks, the control film exhibited another additional peak around $2\theta = 12.8^\circ$, and it is identified for

the presence of a degraded phase of perovskite⁹. This 2H phase is regarded as the non-perovskite phase of FAPbI₃, and it has a non-photovoltaic nature; this phase transition is detrimental to the performance of perovskite solar cells. By contrast, the target thin film exhibited highly oriented PXRD reflections with high intensity, indicating excellent crystallinity compared with the control film. Further, it exhibited high-intensity peaks at the (100) and (200) planes, indicating that the modification resulted in highly oriented crystallinity. Moreover, the degraded 2H phase was not visible in the target film. The increased peak intensity and favored (h00) orientation at the macroscopic level indicate an enhancement in crystallinity, potentially resulting from the reduced defect density due to the AET additive.

3.3. Morphological Analysis. Furthermore, we analyzed the surface uniformity of the perovskite films using SEM; the recorded surface images are shown in Figure 2 E and F. The control film exhibited a rough surface with smaller grains, and is substantiated by the lower absorption and crystallinity characteristics observed above^{4,17}. On the other hand, the target film exhibited a significant enhancement in the grain size and smoothness of the surface. This is in agreement with the enhanced light absorption and crystallinity observed in the UV-Vis and PXRD data explained above.

3.4. Damp Stability. In addition, we investigated the effect of our additivation technique on the humidity stability of perovskite layers. To assess this, we obtained digital images, UV-Vis absorption spectra, and PXRD diffractograms before and after the exposure (aging) of both the control and target films to elevated humidity conditions (85% relative humidity) for 200 h, as illustrated in Figure 3 (A to D). The digital image shows that the control film underwent a phase shift to the photoinactive 2H phase (yellow) within several minutes, whereas our target film maintained its photoactive Figure 3C (black) structure even after 200 h of aging at 85% relative humidity. The UV-Vis absorption spectra also substantiate this observation. XRD patterns were acquired for both the fresh and aged materials, enabling a direct comparison of their stability under humid conditions. The results demonstrated no significant formation of the 2H phase characterized by the peak at $2\theta = 11.8^\circ$. In addition, the peak intensities corresponding to the perovskite phase remain stable in the aged films, as shown in Figure 3D. The results highlight the remarkable humidity stability conferred by the additive modification, demonstrating that the structural integrity and performance characteristics of the altered perovskite films were preserved even under adverse environmental conditions^{6,9,12}. This tolerance to humidity stress is crucial for practical applications in solar cell technology, as environmental variables may significantly affect device lifetime and performance.

The improved light absorption, increased crystallinity, and remarkable humidity stability of the additive-modified FAPbI₃ perovskite films indicate their potential for superior performance in solar applications.

4. CONCLUSIONS

In summary, we investigated the impact of the aminoethylthiol molecule as an additive for FAPbI₃ perovskite thin films.

By applying our additive engineering approach, we were able to enhance the light absorption properties, crystallinity, and morphology of the perovskite coatings, while also reducing intrinsic defects that could lead to non-radiative recombination. After aging for 200 h, the target perovskite film retained its stability under high humidity (85% RH) and temperature (85 °C). The remarkable stability and robust optoelectronic properties of this perovskite layer render it a promising candidate for application in PSCs. The fabrication and evaluation of PSCs has been not been performed in this study. Future research could focus on investigating the long-term photovoltaic performance and efficiency of these additive-modified FAPbI₃ solar cells in fully assembled devices under operational conditions.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Ghadah Khalid Mobaireek – Department of Materials Science and Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0009-0001-5529-3045
Email: s202248500@kfupm.edu.sa

Author

Muhammed P.U. Haris – Research Mentor, Postdoctoral Fellow Interdisciplinary Research Center for Sustainable Energy Systems, King Fahd University of Petroleum & Minerals, Dhahran, 31261, Saudi Arabia;  0000-0003-3696-8622
Email: m.haris@kfupm.edu.sa

Corresponding Author

Amir Al-Ahmed – Research Mentor, Research Scientist-I, Interdisciplinary Research Center for sustainable Energy Systems, King Fahd University of Petroleum & Minerals, Dhahran, 31261, Saudi Arabia; Chemistry Department, King Fahd University of Petroleum & Minerals, Dhahran, 31261, Saudi Arabia;  0000-0001-9686-2374
Email: aalahmed@kfupm.edu.sa

ACKNOWLEDGEMENTS

The authors acknowledge the support provided by the Department of Materials Science and Engineering and the Interdisciplinary Research Center for Sustainable Energy Systems at King Fahd University of Petroleum & Minerals. This research was funded by Interdisciplinary Research Center for Sustainable Energy Systems, KFUPM (INSE-2501).

REFERENCES

- (1) Green, M. A., Ho-Baillie, A., & Snaith, H. J. The emergence of perovskite solar cells. *Nat. photonics*, **8**(7), 506–514 (2014).
- (2) Su, P., et al. Pb-based perovskite solar cells and the underlying pollution behind clean energy: dynamic leaching of toxic substances from discarded perovskite solar cells. *J. Phys. Chem. Lett.*, **11**(8), 2812–2817 (2020).
- (3) Mitzi, D. B. Introduction: perovskites. *Chem. Rev.* **119**(5), 3033–3035 (2019).

- (4) Liu, S., et al. A review on additives for halide perovskite solar cells. *Adv. Energy Mater.* **10**(13), 1902492 (2020).
- (5) Kojima, A., Teshima, K., Shirai, Y., & Miyasaka, T. Organometal halide perovskites as visible-light sensitizers for photovoltaic cells. *J. Am. Chem. Soc.*, **131**(17), 6050–6051 (2009).
- (6) Zhou, H., et al. Buried interface management for FAPbI₃-based perovskite solar cells via multifunctional benzothiadiazole derivatives. *Chem. Eng. J.* **156614** (2024).
- (7) Xiong Z., et al. Homogenized chlorine distribution for >27% power conversion efficiency in perovskite solar cells, *Science*, **390**, 6773, 638–642 (2025).
- (8) Yang, G., Zhang, H., Li, G., & Fang, G. Stabilizer-assisted growth of formamidinium-based perovskites for highly efficient and stable planar solar cells with over 22% efficiency. *Nano Energy*, **63**, 103835 (2019).
- (9) Haris, M. P., Ruiz, E., Kazim, S., & Ahmad, S. Lead-sulfur interaction induced damp and water stability in pure formamidinium lead triiodide. *Cell Rep. Phys. Sci.* **4**(8), 101516 (2023).
- (10) Chen, T., et al. Entropy-driven structural transition and kinetic trapping in formamidinium lead iodide perovskite. *Sci. Adv.*, **2**(10), e1601650 (2016).
- (11) Lee, et al. Rethinking the A cation in halide perovskites. *Science*, **375**(6583), eabj1186 (2022).
- (12) Xu, C., et al. Manipulation of perovskite crystallization kinetics via Lewis base additives. *Adv. Funct. Mater.* **31**(13), 2009425 (2021).
- (13) Sutanto, A. A., et al. In situ analysis reveals the role of 2D perovskite in preventing thermal-induced degradation in 2D/3D perovskite interfaces. *Nano let.* **20**(5), 3992–3998 (2020).
- (14) Wharf, I., Gramstad, T., Makhija, R., & Onyszchuk, M. Synthesis and vibrational spectra of some lead (II) halide adducts with O-, S-, and N-donor atom ligands. *Can. J. Chem.* **54**(21), 3430–3438 (1976).
- (15) Fei, C. et al. Highly efficient and stable perovskite solar cells based on monolithically grained CH₃NH₃PbI₃ film. *Adv. Energy Mater.* **7**(9), 1602017 (2017).
- (16) Varma P. C. R. Low-Dimensional Perovskites Perovskite Photovoltaics: Basic to Advanced Concepts and Implementation, 197–229 (2018).
- (17) Noman, M., Khan, Z., & Jan, S. T. A comprehensive review on the advancements and challenges in perovskite solar cell technology. *RSC Adv.*, **14**(8), 5085–5131 (2024).

Low-Temperature Electrowinning of Iron from Hematite Suspension in Sodium Hydroxide Solution: Toward Sustainable Ironmaking

Mohammed A. Bahamdan and Nestor K. Ankah*

Cite <https://doi.org/10.64589/juri/218961>

Submitted: November 29, 2025 Revised: January 07, 2026 Accepted: January 28, 2026

ABSTRACT

Electrochemical ironmaking has re-emerged as a promising low-carbon alternative to conventional carbothermic processes, yet most prior studies have focused on high-charge or optimized systems aimed at maximizing iron yield. As a result, the early-stage reduction behavior of iron oxides under low-temperature, charge-limited conditions remains insufficiently understood. This study investigates the low-temperature electrochemical reduction of iron from a hematite suspension in concentrated sodium hydroxide as a mechanistic step toward low-carbon ironmaking. Experiments were conducted at 95–110 °C using a three-electrode configuration with a graphite working electrode, a platinum counter electrode, and an Ag/AgCl reference. Cyclic voltammetry (CV) indicated the onset of iron oxide reduction near -1.10 V vs. Ag/AgCl, and the voltammetric features were consistent with limited electrochemical reversibility under the tested conditions. Potentiostatic electrolysis performed at -1.33 V produced a porous iron-rich deposit containing residual oxide. X-ray diffraction combined with Rietveld refinement identified a two-phase product comprising approximately 63 wt.% Fe_3O_4 and 37 wt.% α -Fe, with crystallite sizes in the range of 40–45 nm. SEM and EDS analyses revealed a nanostructured, porous morphology dominated by metallic iron with localized oxygen-rich regions associated with retained magnetite. Charge analysis showed that the supplied 47.4 C corresponded to only a small fraction of the theoretical charge required for the complete reduction of the 5 g hematite feed, which explains the observed partial deoxygenation. The results demonstrate the initiation and governing mechanisms of alkaline electrowinning at reduced temperatures, while highlighting the constraints imposed by the limited charge input, hydrogen evolution, and mass transport under static slurry conditions. By clarifying reduction pathways, phase coexistence, and morphological evolution under incomplete conversion, this study establishes a mechanistic baseline that is critical for bridging fundamental electrochemistry and scalable low-temperature alkaline iron electrowinning.

Keywords: sustainable, electrowinning, hematite, rietveld analysis, cyclic voltammetry, iron oxide reduction

1. INTRODUCTION

Ironmaking is a basic industrial process that supports modern infrastructure, manufacturing, and energy systems, yet it remains among the most energy- and carbon-intensive metallurgical operations. Conventional iron production relies predominantly on carbothermic reduction in blast furnaces, which requires temperatures exceeding 1500 °C and consumes large quantities of fossil carbon. As a result, iron and steelmaking account for approximately 7–9% of global anthropogenic CO_2 emissions^{1,2}. This environmental burden has intensified the search for alternative ironmaking methods that decouple iron production from carbon-intensive thermal processes. Electrochemical reduction of iron oxides has re-emerged as a promising low-carbon alternative, as it replaces carbon with electricity as the primary reducing agent. In particular, alkaline electrochemical routes enable iron oxide reduction at temperatures below 120 °C, several orders of magnitude lower than those required for conventional smelting. In strongly alkaline environments, iron oxides can undergo reduction through soluble ferrite and hydroxo-complex intermediates,

while oxygen evolution occurs at the anode, producing no direct carbon emissions^{3–6}. Sodium hydroxide (NaOH) is especially attractive as an electrolyte owing to its high ionic conductivity, ability to stabilize $\text{Fe}^{3+}/\text{Fe}^{2+}$ species, and compatibility with graphite and metallic electrodes under alkaline conditions. Extensive prior work by Allanore et al.^{5,6} and Lopes et al.^{7,8} established the fundamental mechanisms governing alkaline electroreduction of hematite and magnetite, including the sequential $\text{Fe}_2\text{O}_3 \rightarrow \text{Fe}_3\text{O}_4 \rightarrow \text{Fe}$ pathway, the critical role of electrolyte concentration, and the competition between iron deposition and hydrogen evolution. Subsequent studies have demonstrated that high current efficiencies and near-complete reduction are achievable through optimized cell design, controlled slurry transport, and large charge inputs. However, these studies primarily focused on process optimization and yield enhancement, often under tightly controlled hydrodynamic or engineered electrode conditions^{9–12}. By contrast, comparatively little attention has been paid to the early-stage reduction behavior of hematite suspensions under charge-limited conditions at low temperatures (≤ 110 °C). This regime is characterized by

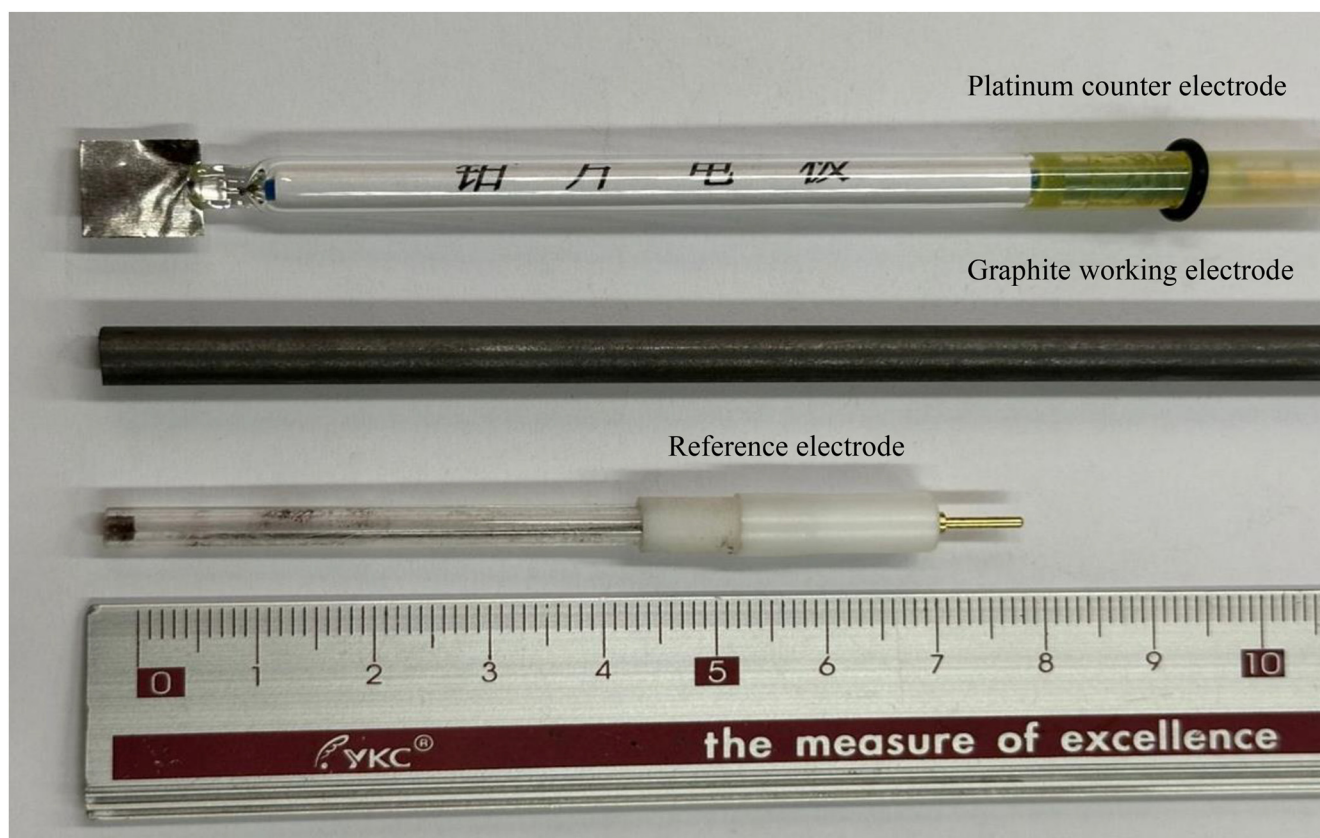


Figure 1. Image of the working, counter, and reference electrodes utilized in this work

partial deoxygenation, phase coexistence, and porous iron formation, all of which strongly influence the downstream processing, sintering behavior, and mechanical integrity of the reduced product^{13–15}. Understanding iron nucleation, phase evolution, and morphology development under such conditions is essential for bridging the gap between fundamental electrochemical behavior and scalable low-carbon ironmaking technologies^{16,17}.

Accordingly, the present study investigates the low-temperature electrochemical reduction of hematite suspended in concentrated NaOH at 95–110 °C using a three-electrode cell with a graphite cathode, a platinum counter electrode, and an Ag/AgCl reference with explicit emphasis on reduction pathways, phase composition, and microstructural evolution under limited charge input. The recovered solids were examined using scanning electron microscopy (SEM), energy-dispersive X-ray spectroscopy (EDS), and X-ray diffraction (XRD) to evaluate their morphology, composition, and phase structure. Unlike prior studies aimed at maximizing the iron yield or faradaic efficiency, this work was designed as a mechanistic investigation, examining how incomplete charge transfer influences the formation of mixed $\text{Fe}_3\text{O}_4/\alpha\text{-Fe}$ products and porous iron morphologies. By clarifying the limitations and governing processes in this low-temperature low-conversion regime, the study provides insights relevant to the development of future alkaline electrowinning strategies for sustainable iron production.

The remainder of this paper is arranged as follows: Section 2 introduces the materials and methods used in the study; Section 3 presents the obtained results and discusses their significance; and conclusions are drawn in Section 4.

2. METHODOLOGY

2.1. Materials. All materials were used as received without further purification. Sodium hydroxide pellets (NaOH, purity $\geq 97.0\%$, Fisher Chemical™) and iron (III) oxide powder (Hematite Fe_2O_3 , purity: $\geq 99.995\%$ trace metals basis, Sigma-Aldrich) were employed as the electrolyte and iron source, respectively. Deionized water with a resistivity of $18.2 \text{ M}\Omega\cdot\text{cm}$ was used for all solution preparations. In the electrochemical cell, a graphite rod (CS915, 6 mm in diameter) was used as the working electrode (WE) for iron deposition; a platinum plate electrode (CS911, 10 mm $\times$ 10 mm $\times$ 0.1 mm) was used as the counter electrode (CE); and a standard reference electrode (RE), silver/silver chloride, filled with a 4 M KCl solution (Ag/AgCl 4 M), was used in the electrochemical cell measurements to provide a stable and constant potential. The graphite electrode was selected owing to its chemical stability in concentrated alkaline media and its widespread use in alkaline iron electrowinning studies. Figure 1 illustrates the working, counter, and reference electrodes utilized in this work.

2.2. Electrolyte and Slurry Preparation. The alkaline electrolyte was prepared by dissolving 36 g of NaOH pellets in 50 mL of deionized water, corresponding to an approximate NaOH concentration of 18 M. The solution was magnetically stirred for 15 min to ensure complete dissolution and homogeneity, after which it was allowed to equilibrate thermally. Subsequently, 5 g of hematite powder was added to the NaOH solution, yielding an approximately 10 wt.% Fe_2O_3 suspension relative to the solution mass. The suspension was magnetically stirred for an

additional 15 min to promote uniform particle dispersion prior to electrochemical testing. No active stirring was applied during CV and potentiostatic electrolysis. The experiments were conducted under static slurry conditions to examine the reduction behavior under limited mass transport. Because of the short experimental duration, no pronounced macroscopic sedimentation was observed; however, the influence of particle settling and diffusion limitations is acknowledged as a constraint of the present study.

2.3. Electrochemical Test. All electrochemical measurements and electrowinning experiments were conducted using a CorrTest instrument potentiostat (CS300M Potentiostat/Galvanostat, by Wuhan Corrtest Instruments Corp., Ltd., Wuhan, China). The experiments were performed in a custom three-electrode electrochemical cell with a total electrolyte volume of 100 mL. A schematic of the electrochemical cell is presented in Figure 2. The electrolyte temperature was controlled using a Stuart US 152D digital hot plate and maintained at 110 °C; the reported ± 5 °C reflects an instrumental limitation rather than intentional variation. The working, counter, and reference electrodes were immersed in the electrolyte before the measurements, with the reference electrode positioned close to the working electrode to minimize the ohmic drop. A magnetic stirring bar was placed at the bottom of the cell but remained stationary during electrochemical measurements. CV was conducted to probe the redox behavior of the iron species in the alkaline hematite suspension. The potential was scanned from 0 to -1.33 V vs. Ag/AgCl in the cathodic direction and subsequently reversed to 0 V in the anodic direction at a scan rate of 50 mV s^{-1} for three consecutive cycles. Only a single scan rate was employed in this study; therefore, conclusions regarding irreversibility are limited to the observed behavior under this specific condition, and the need for multi-scan-rate analysis is acknowledged. The lower potential limit was selected based on prior studies showing that this range is sufficient to drive Fe_3O_4 and Fe formation in concentrated alkaline media while avoiding

excessive electrolyte decomposition. Potentiostatic iron deposition was performed by applying a constant potential of -1.33 V vs. Ag/AgCl for 30 min, and during this process the current was recorded continuously. After electrolysis, the working electrode was removed, and the deposited particles were collected from the graphite surface using a magnetic stirring bar. The particles were stored in a container filled with ethanol to minimize oxidation. Post-experiment inspection and EDS analysis showed no evidence of graphite degradation or carbon contamination in the deposited material.

2.4. Characterization. The morphology and composition of the reduced particles were examined using SEM coupled with EDS on a JEOL JSM-6610LV microscope equipped with an Oxford Instruments X-Max detector. XRD analysis was performed using a Rigaku Mini Flex X-ray diffractometer operating at 30 kV and 30 mA using $\text{Cu K}\alpha$ radiation ($\lambda = 0.1541 \text{ nm}$). The XRD patterns were collected over a 2θ range of 10° – 90° with a step size of 0.01° and 1 s per step. Rietveld refinement was applied to quantify phase fractions and estimate crystallite sizes.

3. RESULTS AND DISCUSSION

3.1. Electrochemical (CV/Chronoamperometry) Analysis. Figure 3(a) shows the cyclic voltammogram recorded at 110 °C in concentrated NaOH using an Ag/AgCl reference electrode. The current remains small until the potential approaches approximately -1.10 V, which marks the onset of the electrochemical reduction of the suspended Fe_2O_3 particles. As the potential becomes more cathodic, the current rises sharply and reaches a pronounced minimum between -1.25 and -1.30 V. This feature corresponds to overlapping reduction processes in which Fe_2O_3 converts to Fe_3O_4 , and Fe_3O_4 further reduces toward metallic Fe¹⁷. Hydrogen evolution also becomes significant in this range, contributing to the increase in cathodic current. The steep decrease and shallow minimum reflect the rapid nucleation of metallic Fe on partially reduced oxide grains. The return sweep shows no distinct oxidation peak apart from a weak

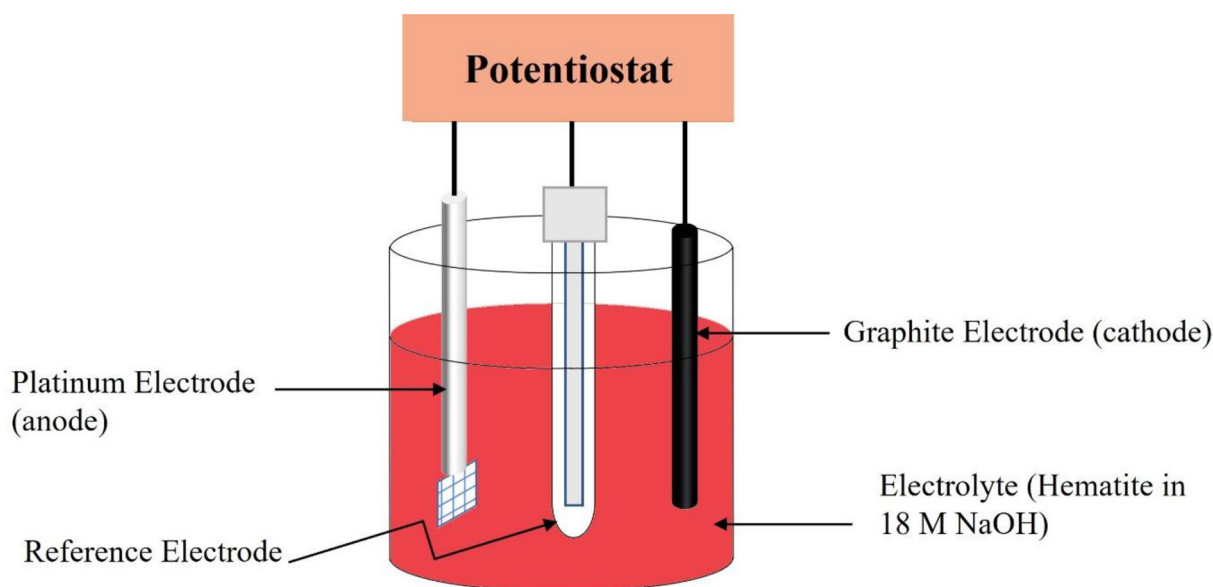


Figure 2. Schematic of the electrochemical cell for CV experiments

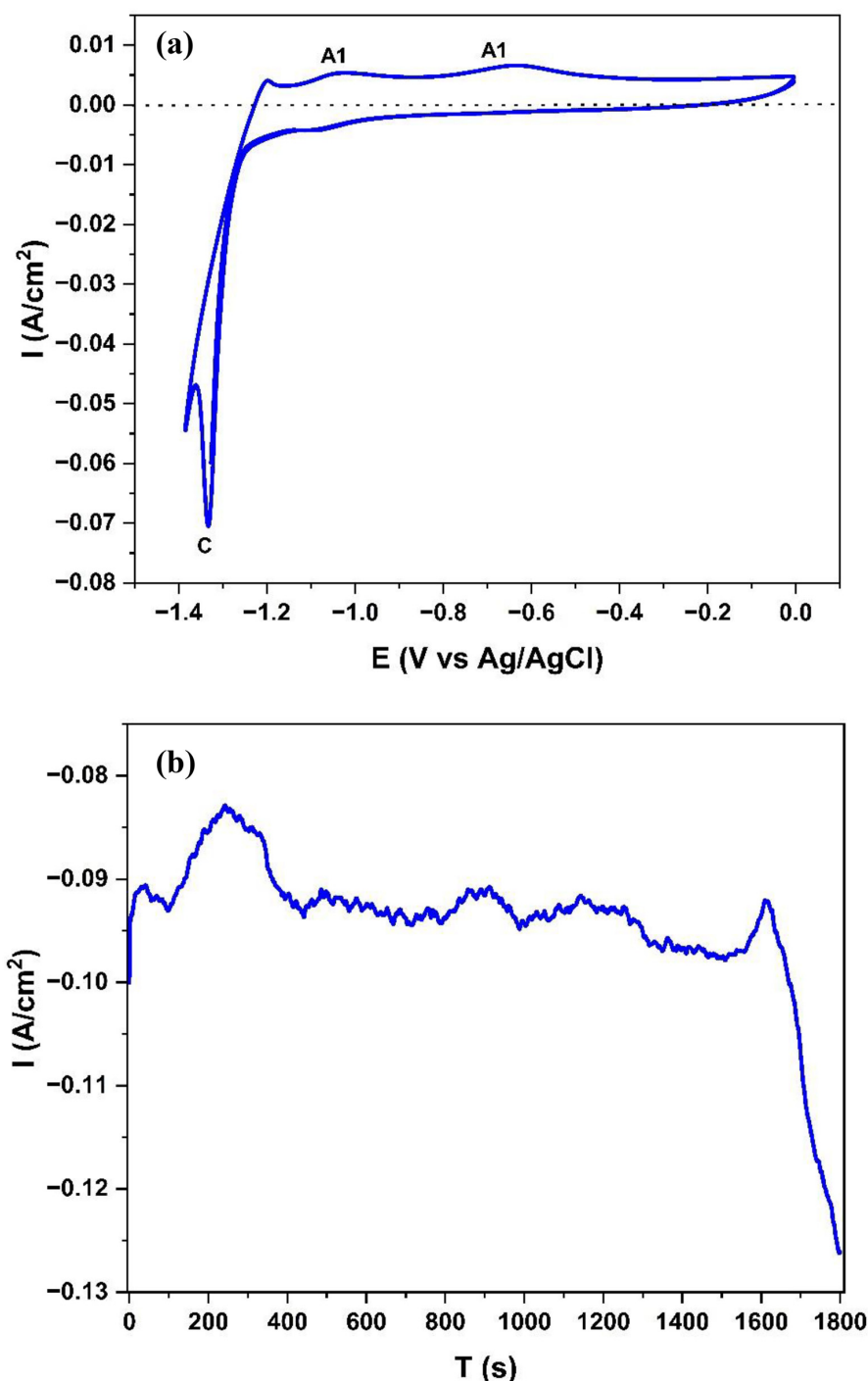


Figure 3. Electrochemical measurements carried out at 110 °C: (a) CV performed at -1.38 V and scan rate of 50 mVs^{-1} after 30 min of Fe deposition from an NaOH suspension of Fe_2O_3 , and (b) chronoamperometry curve recorded at -1.33 V

anodic feature. This strong hysteresis is consistent with an electrochemically irreversible process in which the newly formed Fe and Fe_3O_4 are not reoxidized within the scanned potential window¹⁸. The CV therefore captures the known $\text{Fe}_2\text{O}_3 \rightarrow \text{Fe}_3\text{O}_4 \rightarrow \text{Fe}$ pathway in an alkaline solution, with hydrogen evolution acting as a competing reaction.

The potentiostatic current–time profile in Figure 3(b) provides a complementary insight into the kinetics of the reduction process. Following the potential step to -1.33 V, the current exhibits an initial transient and stabilizes at approximately -0.09

to -0.10 A cm^{-2} for most of the 30-min electrolysis run. This behavior indicates a regime in which reduction is limited by the supply of oxide particles to the electrode surface and by simultaneous hydrogen evolution^{10,12}. Toward the end of electrolysis, the current again increases in magnitude, reaching -0.12 to -0.13 A cm^{-2} , likely due to intensified hydrogen evolution after local oxide depletion. Integration of the current–time curve yields a total cathodic charge of 47.4 C. By contrast, full reduction of the original 5 g Fe_2O_3 feed would require approximately 1.81×10^4 C. Thus, only approximately 0.26% of the theoretical

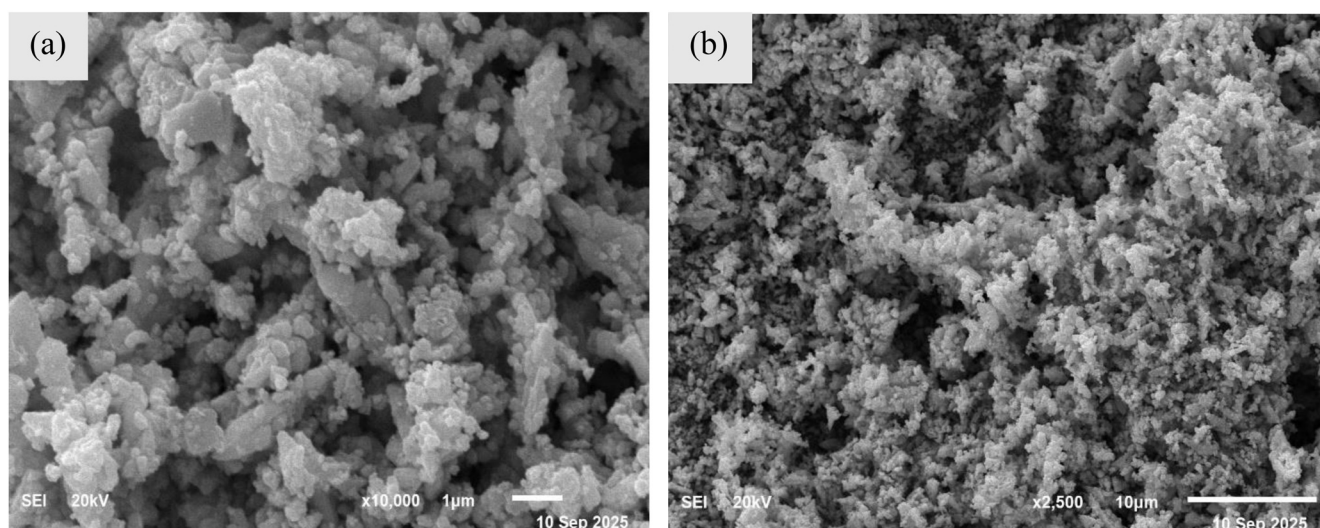


Figure 4. Microstructural analysis of the iron deposits, showing SEM images acquired from the hematite suspension at two magnifications: (a) 10,000 $\times$ with a 1 μ m scale, and (b) 2,500 $\times$ with a 10 μ m scale

charge was supplied. This limited charge explains why only partial deoxygenation occurred and why the final product contained a mixture of Fe_3O_4 and metallic Fe. The combined CV and chronoamperometric results show that the electrode operated in a mixed regime of solid-state oxide reduction and hydrogen evolution, with incomplete conversion strongly constrained by the low charge input¹⁹.

3.2. SEM & EDS Analysis. SEM images of the reduced product are shown in Figure 4. Observation at higher magnification (Figure 4a) shows that the deposit consists of irregular, porous agglomerates composed of nanoscale crystallites. This morphology is characteristic of electrochemical iron formation under conditions where nucleation occurs rapidly on partially reduced oxide particles and growth is disrupted by concurrent hydrogen evolution. The absence of well-defined faceted grains

suggests that solid-state reduction and localized nucleation dominate over sustained metallic growth²⁰. At lower magnification (Figure 4b), the reduced material forms a highly porous, interconnected network extending over tens of micrometers²¹. Such porosity is commonly reported in alkaline electrowinning systems and arises from a combination of incomplete deoxygenation, gas evolution, and fragmentation of the parent oxide particles. The observed morphology is consistent with the charge-limited nature of the experiment and aligns with the mixed-phase composition identified by XRD.

EDS analysis (Figure 5) indicates that the deposit is iron-rich, with approximately 93.0 wt.% Fe and 7.0 wt.% O. The residual oxygen is attributed to the retained Fe_3O_4 and the partially reduced oxide domains rather than to surface oxidation alone. No extraneous elements were detected, confirming that the electrolyte and electrode materials did not introduce measurable contamination. Importantly, no carbon enrichment was

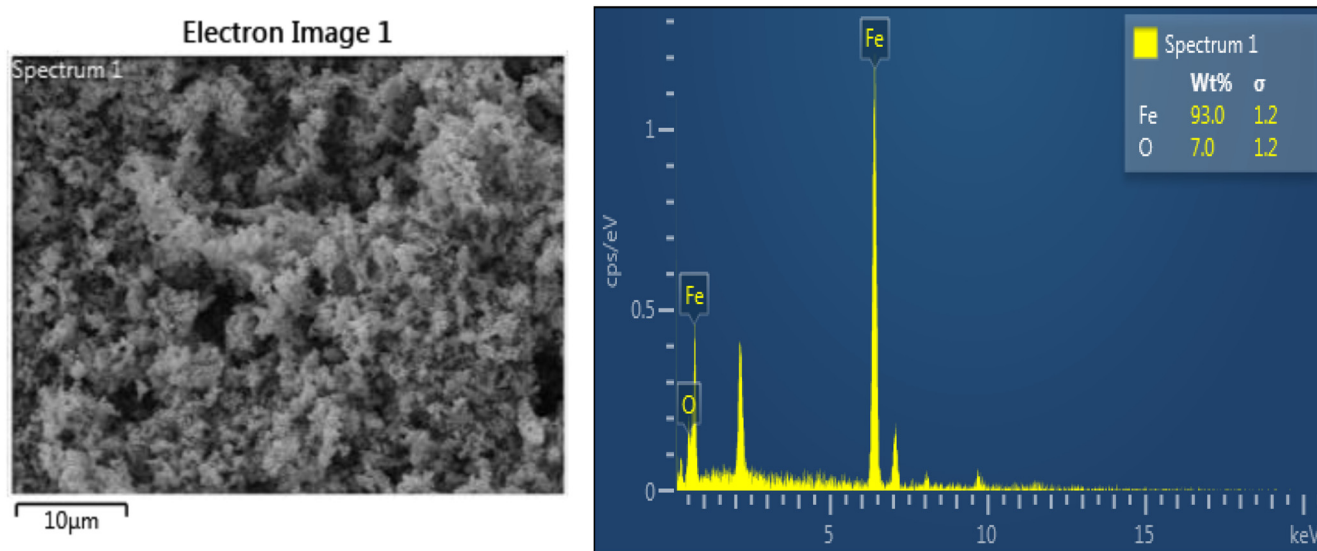


Figure 5. Energy dispersive X-ray spectroscopy (EDS) spectrum of iron deposits showing the quantitative elemental composition by weight (Fe 93.0 wt.% and O 7.0 wt.%)

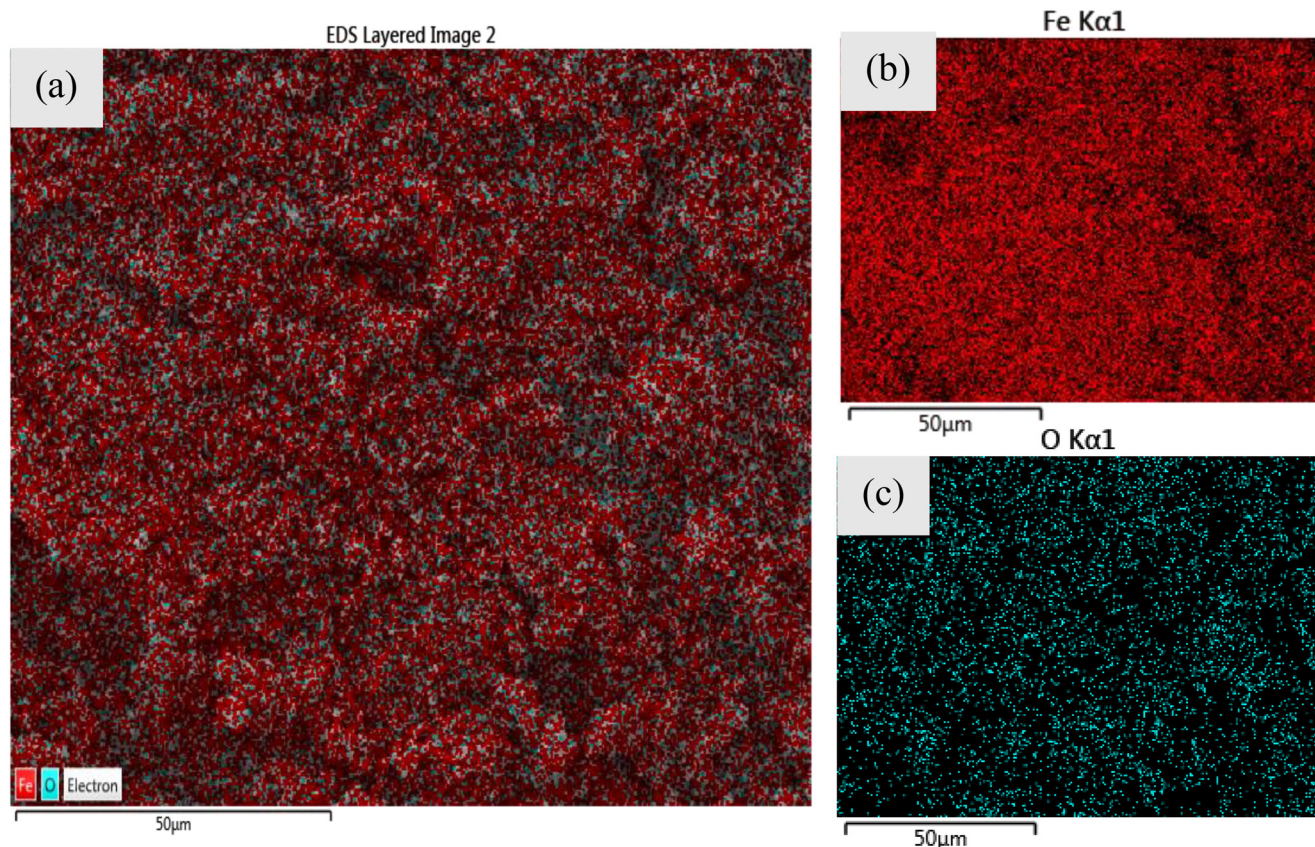


Figure 6. EDS elemental maps of the iron deposits: (a) Combined Fe–O map, (b) Fe, and (c) O distributions

observed beyond the graphite substrate background, indicating negligible cathode degradation during electrolysis. Elemental mapping (Figure 6) further supports these observations. The Fe map shows a relatively uniform distribution across the analyzed region, whereas the O map reveals localized, discrete oxygen-rich areas. These features correspond to magnetite-rich regions embedded within a metallic iron matrix, reflecting the spatially heterogeneous reduction under a limited charge input.

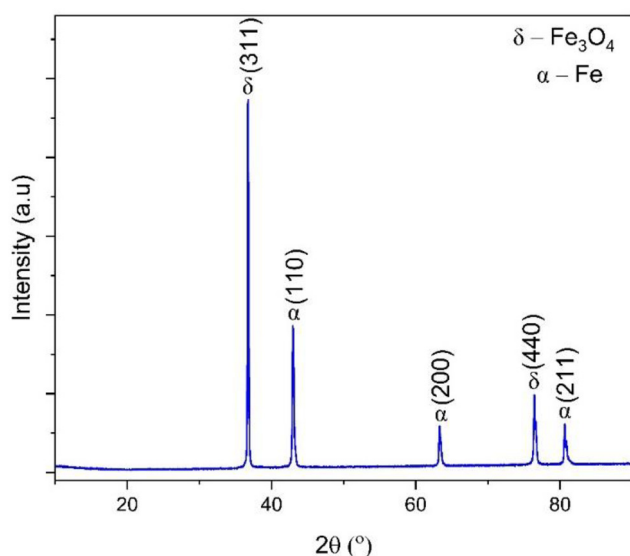


Figure 7. XRD pattern showing the presence of Fe_3O_4 and $\alpha\text{-Fe}$ after electrowinning

3.3. XRD. The XRD pattern of the reduced product is shown in Figure 7. The diffraction peaks at approximately 36.75° and 76.45° (2θ) correspond to the characteristic reflections of Fe_3O_4 , whereas the peaks at 42.97° , 63.31° , and 80.65° are indexed to the (110), (200), and (211) planes of body-centered cubic $\alpha\text{-Fe}$. No residual Fe_2O_3 peaks were detected, indicating that hematite was at least partially converted during electrolysis. Rietveld refinement indicates that the product consists of approximately 63 wt.% Fe_3O_4 and 37 wt.% $\alpha\text{-Fe}$, confirming partial reduction consistent with the limited charge supplied. The refined lattice parameters are close to the reference values, indicating well-defined crystallinity of both phases. The crystallite sizes estimated from peak broadening fall in the range of 40–45 nm, whereas the microstrain values on the order of 10^{-4} suggest that lattice strain contributes minimally to peak broadening^{1,6,8,21}.

The coexistence of magnetite and metallic iron reflects the incomplete reduction pathway expected under static slurry conditions and limited charge transfer. Rather than representing inefficiency, this mixed-phase product provides insights into the intermediate reduction states that are often bypassed in high-charge or highly optimized systems.

Unlike molten oxide electrolysis or hydrogen-based electroreduction routes, which are conducted at elevated temperatures and aim for complete conversion, the present alkaline electrowinning approach enables iron formation at temperatures below 110°C . While this significantly reduces the thermal energy requirements, the results demonstrate that charge transfer, mass transport, and hydrogen evolution impose strong constraints on

conversion under static conditions²⁵. The present study does not attempt to optimize the faradaic efficiency or scalability. Instead, it highlights key limitations inherent to low-temperature, charge-limited operation, including incomplete reduction, porous morphology, and phase heterogeneity. Addressing these challenges will require improved temperature control, enhanced slurry transport, suppression of hydrogen evolution, and systematic efficiency evaluation.

4. CONCLUSIONS

This study investigated the electrochemical reduction of a hematite suspension in concentrated sodium hydroxide at temperatures between 95 and 110 °C using a graphite cathode, a platinum counter electrode, and an Ag/AgCl reference electrode. The work was conducted under charge-limited conditions to examine the early-stage reduction behavior rather than to optimize the iron yield or efficiency. The main conclusions can be summarized as follows:

1. CV revealed the onset of iron oxide reduction at approximately -1.10 V vs. Ag/AgCl, followed by increased cathodic current at more negative potentials. Under the single scan rate employed, the voltammetric response exhibited limited electrochemical reversibility within the investigated potential window, consistent with overlapping iron oxide reduction and hydrogen evolution processes.
2. Potentiostatic electrolysis at -1.33 V vs. Ag/AgCl supplied a total charge of 47.4 C over 30 min, corresponding to only 0.26% of the theoretical charge required for the complete reduction of the 5 g Fe_2O_3 feed. This deliberately limited charge input resulted in partial deoxygenation rather than full conversion to metallic iron.
3. SEM analysis showed that the reduced product was formed as highly porous agglomerates composed of nanoscale crystallites, reflecting nucleation-dominated iron formation under conditions where hydrogen evolution and incomplete reduction disrupted sustained metallic growth. Such morphologies are characteristic of early-stage alkaline electrowinning and are strongly influenced by mass transport and charge limitations.
4. EDS confirmed an iron-rich composition with localized oxygen associated with residual magnetite domains. No evidence of graphite degradation or carbon contamination was found, indicating that the cathode remained chemically stable under the applied conditions.
5. XRD combined with Rietveld refinement identified a two-phase mixture consisting of approximately 63 wt.% Fe_3O_4 and 37 wt.% $\alpha\text{-Fe}$, with crystallite sizes in the range of 40–45 nm. The absence of residual hematite confirms progression beyond the initial reduction stage, while the coexistence of magnetite and metallic iron reflects incomplete reduction governed by limited charge transfer.

Overall, the results demonstrate that low-temperature alkaline electrowinning can initiate iron formation from hematite at temperatures well below those required for conventional ironmaking routes. However, the present findings also highlight critical limitations associated with static slurry conditions, hydrogen

evolution, and insufficient charge input. Addressing these constraints through improved temperature control, enhanced mass transport, suppression of competing reactions, and systematic efficiency analysis will be essential for advancing alkaline electrowinning toward scalable low-carbon ironmaking applications.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Mohammed A. Bahamdan – Department of Materials Science & Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0009-0005-2169-3031
Email: s202245900@kfupm.edu.sa

Corresponding Author

Nestor K. Ankah – Research Mentor, Postdoctoral Fellow, Interdisciplinary Research Center for Advanced Materials, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0000-0002-6059-5646
Email: nestor.ankah@kfupm.edu.sa

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support provided by the Department of Materials Science and Engineering, Interdisciplinary Research Center for Sustainable Energy Systems, and the Interdisciplinary Research Center for Advanced Materials at King Fahd University of Petroleum & Minerals. The authors also extend their appreciation to the technical staff and laboratory team for their assistance in specimen preparation and testing.

REFERENCES

- (1) Lopes, D. V., Quina, M. J., Frade, J. R., Kovalevsky, A. V & Wang, B. Prospects and challenges of the electrochemical reduction of iron oxides in alkaline media for steel production. 1–8 (2022). doi:10.3389/fmats.2022.1010156
- (2) Lopes, D. V., Kovalevsky, A. V & Quina, M. J. Electrochemical Deposition of Zero-Valent Iron from Alkaline Ceramic Suspensions of $\text{Fe}_2 - x\text{Al}_x\text{O}_3$ for Iron Valorisation Electrochemical Deposition of Zero-Valent Iron from Alkaline Ceramic Suspensions of $\text{Fe}_2 - x\text{Al}_x\text{O}_3$ for Iron Valorisation. 2–10 (2020). doi:10.1149/1945-7111/ab9a2b
- (3) Kim, H., Paramore, J., Allamore, A. & Sadoway, D. R. Electrolysis of Molten Iron Oxide with an Iridium Anode: The Role of Electrolyte Basicity. *J. Electrochem. Soc.* **158**, E101 (2011).
- (4) Allamore, A., Lavelaine, H., Valentin, G., Birat, J. P. & Lapicque, F. Electrodeposition of Metal Iron from Dissolved Species in Alkaline Media. *J. Electrochem. Soc.* **154**, E187 (2007).
- (5) Fan, Z. & Friedmann, S. J. Review Low-carbon production of iron and steel: Technology options, economic assessment, and policy. *Joule* **5**, 829–862 (2021).
- (6) Koutsoupa, S., Koutalidi, S., Balomenos, E. & Panias, D. Σ IDER-WIN — A New Route for Iron Production †. in *Materials Proceedings* **5**, 2–7 (2021).
- (7) Yuan, B. *et al.* Effects of Annealing and Oxidation on the Microstructure of Hot-Dipped Aluminum–Silicon Coating of 316L Stainless Steel. *J. Phase Equilibria Diffus.* **45**, 114–131 (2024).
- (8) Lopes, D. V *et al.* Electrochemical reduction of hematite-based ceramics in alkaline medium: Challenges in electrode design. *Electrochim. Acta* **327**, 135060 (2019).

- (9) Haarberg, G. M. & Yuan, B. Direct Electrochemical Reduction of Hematite Pellets in Alkaline Solutions. *ECS Trans.* **58**, 19–28 (2014).
- (10) Siebentritt, M., Volovitch, P., Ogle, K. & Lefèvre, G. Adsorption and electroreduction of hematite particles on steel in strong alkaline media. *Colloids Surfaces A Physicochem. Eng. Asp.* **440**, 197–201 (2014).
- (11) Ivanova, Y. A. *et al.* Designed porous microstructures for electrochemical reduction of bulk hematite ceramics. *Mater. & Des.* **122**, 307–314 (2017).
- (12) Lopes, D. V. *et al.* Electrochemical reduction of hematite-based ceramics in alkaline medium: Challenges in electrode design. *Electrochim. Acta* **327**, 135060 (2019).
- (13) Boehme, C., Strykevich, M., Constantinescu, G., Kovalevsky, A. V & Lopes, D. V. Low-Temperature Electrowinning of Iron from Mixed Hematite-Magnetite Alkaline Suspensions. *J. Electrochem. Soc.* **170**, 52502 (2023).
- (14) Monteiro, J. F., Ivanova, Y. A., Kovalevsky, A. V, Ivanou, D. K. & Frade, J. R. Reduction of magnetite to metallic iron in strong alkaline medium. *Electrochim. Acta* **193**, 284–292 (2016).
- (15) Majid, A. I., Finotello, G., van der Schaaf, J., Deen, N. G. & Tang, Y. On the formation of dendritic iron from alkaline electrochemical reduction of iron oxide prepared for metal fuel applications. *Chem. Eng. Sci.* **291**, 119931 (2024).
- (16) Lei, J. *et al.* Mechanism study on gas-based reduction swelling behavior of ultra-high grade pellets. *J. Mater. Res. Technol.* **26**, 823–836 (2023).
- (17) Meshram, A., Govro, J., OMalley, R. J., Sridhar, S. & Korobeinikov, Y. Modeling Isothermal Reduction of Iron Ore Pellet Using Finite Element Analysis Method: Experiments & Validation. *Metals (Basel)*. **12**, 2026 (2022).
- (18) LeDuc, J. A. M., Lofthfield, R. E. & Vaaler, L. E. Electrolytic Iron Powder from a Caustic Soda Solution. *J. Electrochem. Soc.* **106**, 659 (1959).
- (19) Cekerevac, M., Nikolic-Bujanovic, L. & Simicic, M. Investigation of electrochemical synthesis of ferrate, Part I: Electrochemical behavior of iron and its several alloys in concentrated alkaline solutions. *Hem. Ind.* **63**, 387–395 (2009).
- (20) Gorbunova, K. M. & Liamina, L. I. On the mechanism of iron reduction from alkaline solutions. *Electrochim. Acta* **11**, 457–467 (1966).
- (21) Monteiro, J. F., Ivanova, Y. A., Kovalevsky, A. V, Ivanou, D. K. & Frade, J. R. Reduction of magnetite to metallic iron in strong alkaline medium. *Electrochim. Acta* **193**, 284–292 (2016).

CO₂ Capture using Dual-Amine Functionalized MCM-41 Nanoparticles Synthesized via Silane Condensation

Maymounah Alshahrani and Ahmed Wasel A Alsmaeil*

Cite <https://doi.org/10.64589/juri/219228>

Submitted: February 12, 2025 Revised: February 17, 2026 Accepted: February 25, 2026

ABSTRACT

This paper reports the synthesis of dual-amine-functionalized MCM-41 nanoparticles via a silane condensation approach for enhanced CO₂ capture applications. Mesoporous MCM-41 was prepared using a surfactant-templated sol-gel method and subsequently functionalized with a 1:1 mixture of Hydrosil 2775 and Hydrosil 2999 to introduce cooperative amine functionalities onto the silica surface. Formation of the target structure and successful surface modification were confirmed by Fourier-transform infrared spectroscopy, while Brunauer-Emmett-Teller analysis revealed a decrease in surface area and pore volume after functionalization, consistent with amine grafting within the mesoporous framework. Thermogravimetric analysis performed under pure CO₂ at 30 °C demonstrated an adsorption capacity of 1.7 wt%, calculated from total mass uptake during CO₂ exposure. The observed performance reflects the balance between amine loading and pore accessibility, which is inherent to grafted systems and provides insight into the design of dual-amine-functionalized mesoporous silica materials for post-combustion carbon capture applications.

Keywords: MCM-41, dual-amine functionalization, mesoporous silica, CO₂ capture, surface modification

1. INTRODUCTION

As atmospheric carbon dioxide (CO₂) levels continue to rise, mitigating emissions from the industrial and energy sectors through carbon capture has become increasingly important¹. Although conventional liquid amine scrubbing processes are widely implemented for CO₂ capture, they suffer from serious drawbacks, including high regeneration energy requirements, solvent degradation, and equipment corrosion². These limitations have driven growing interest in solid sorbent technologies, which offer improved thermal stability and tunable surface chemistry for selective CO₂ adsorption. Porous solid materials can be engineered to overcome the many challenges faced by liquid solvents, rendering them promising candidates for more efficient post-combustion CO₂ capture systems. For instance, metal-organic frameworks (MOFs), activated carbon, and functionalized silica have been extensively studied owing to their high surface areas and favorable adsorption properties³. Mesoporous silica materials have attracted particular interest as CO₂ sorbents because of their ordered pore networks, narrow pore size distributions, and large surface areas, which facilitate effective gas diffusion and enable tailor-made surface functionalization⁴. MCM-41, a prototypical mesoporous silica with a hexagonally ordered pore structure and tunable pore size, represents an excellent high-surface-area scaffold for grafting functional groups. These characteristics render MCM-41 an adaptable platform for introducing active sites that enhance CO₂ affinity and uptake capacity⁵.

Amine functionalization is commonly employed to enhance the CO₂ capture performance of mesoporous silica through the introduction of reactive amino groups onto the pore surfaces. Amines bind chemically to CO₂ via carbamate or bicarbonate formation under dry and moist conditions, respectively⁶. This chemisorption mechanism enhances CO₂ uptake at the low partial pressures typical of flue gas and improves selectivity over purely physisorptive adsorbents⁷. Amines can be incorporated into silica via covalent grafting of amino silanes onto surface silanol (Si-OH) groups or impregnation of the pores with polyamine molecules⁸. While grafting minimizes amine loss during cyclical operation, impregnation introduces higher overall amine loadings at the cost of pore blocking or amine volatilization⁸. Accordingly, the choice of amine type, functionalization technique, and distribution within the pore network is critical for achieving optimal CO₂ adsorption capacity and sorbent stability.

Nevertheless, the performance of amine-functionalized MCM-41 remains highly sensitive to the nature and distribution of the grafted amine groups⁷. Although many studies have focused on polymeric amines impregnated into mesoporous supports and single small-molecule amino silanes grafted onto the surface, comparatively less attention has been given to tailored aminosilane coupling agents with diverse organic amine functionalities, limiting the rational design of improved amine-modified sorbents⁹⁻¹¹. It is therefore necessary to investigate the effects of different amine functional groups on CO₂ uptake in mesoporous silica materials⁸.

This study introduces a novel dual-amine functionalization approach for MCM-41, exploring, for the first time, the co-grafting of a 1:1 mixture of two different aminosilane agents onto the same silica framework. Specifically, MCM-41 is functionalized with an equimolar combination of Hydrosil 2775 (oligomeric alkylsilane bearing linear primary amine groups) and Hydrosil 2999 (oligomeric quaternary ammonium-functionalized silane) to integrate their complementary molecular characteristics. It is proposed that the linear amine groups will offer strongly chemisorptive sites for CO₂, while the quaternary ammonium groups will enhance amine loading. The dual-amine-functionalized MCM-41 is thoroughly characterized to confirm successful incorporation of both amine functionalities and to examine changes in its textural properties. Fourier transform infrared (FTIR) spectroscopy is employed to confirm successful modification of the silica framework, while N₂ adsorption-desorption isotherms provide insights into surface area and pore volume alterations following functionalization. Finally, CO₂ uptake tests are conducted using thermogravimetric analysis (TGA) to quantify the adsorption capacity of the sorbent and evaluate its CO₂ sorption behavior under simulated post-combustion conditions.

2. METHODOLOGY

2.1. Materials. MCM-41 was synthesized as described in Section 2.2. Hexadecyltrimethylammonium bromide (CTAB, 95%), tetraethyl orthosilicate (TEOS, 98%), NaOH (1 and 2 M), ethanol, deionized water (18.2 MΩ•cm at 25 °C), N₂ gas (99.99% purity), CO₂ gas (99.99% purity), Hydrosil 2775, and Hydrosil 2999 (purchased from Evonik, 0.200 kg net weight) were used as received without further purification.

2.2. Synthesis of MCM-41. MCM-41 was synthesized via a surfactant-templated sol-gel method under basic conditions. CTAB (1.0 g) was dissolved in deionized water (480 mL) at 70 °C, and NaOH (2 M, 3.5 mL) was added. TEOS (5 mL) was then added dropwise under continuous stirring. After stirring at this temperature for 2 h, the resulting white precipitate was collected by vacuum filtration without prior cooling, washed thoroughly with deionized water, and dried at 60 °C overnight. Surfactant removal was achieved by calcination at 550 °C for 2 h (heating rate = 10 °C min⁻¹) to obtain mesoporous MCM-41.

2.3. Dual-amine Functionalization. Surface functionalization was performed using a 1:1 mixture of Hydrosil 2775 and Hydrosil 2999. Calcined MCM-41 (75 mg) was suspended in deionized water (25 mL) and NaOH (1 M) was added to increase the solution pH to 10. Hydrosil 2775 (75 mg) and Hydrosil 2999 (75 mg) were then added to the solution and stirred continuously overnight at 70 °C. After this time, the resulting suspension was subjected to centrifugation (×2) and the precipitate was dried overnight at 50 °C to yield the functionalized MCM-41.

3. CHARACTERIZATION

3.1. Structural Characterization.

3.1.1. FTIR spectroscopy. FTIR spectroscopy was performed for the pure and functionalized MCM-41 samples using a Bruker Invenio spectrometer with a diamond ATR configuration (platinum ATR). Spectra were recorded over the range of 4000–400 cm⁻¹, with 2522 data points collected per scan. Each spectrum was obtained by averaging 64 scans.

3.1.2. Brunauer-Emmett-Teller (BET) analysis. The surface area and pore structure were analyzed using a Belsorp Max X-114 Brunauer-Emmett-Teller (BET) instrument with N₂ as the adsorbent at 77 K. Prior to analysis, the sample (~0.04326 g) was degassed at 120 °C for 6 h under vacuum. The manifold temperature during analysis was maintained at 50 °C. The BET surface area was determined using the linear fitting region selected by the instrument software, within a relative pressure range (P/P_0) of 0–0.35. The calculated BET C constant was 49.721. The pore size distribution was evaluated using the Barrett-Joyner-Halenda (BJH) method (adsorption branch) and the nonlocal density functional theory (NLDFT) model for cylindrical pores using N₂ at 77 K.

3.1.3. TGA for CO₂ uptake. CO₂ adsorption experiments were performed using a PerkinElmer Pyris TGA system and a platinum pan (~6 mg sample per measurement). Single runs were performed for both pure MCM-41 and the functionalized samples using identical thermal programs. Samples were heated at 30 °C under nitrogen purging (20 mL min⁻¹) for 1 min prior to increasing the temperature to 200 °C at a heating rate of 10 °C min⁻¹ (nitrogen flow, 20 mL min⁻¹). Subsequently, the temperature was held at 200 °C for 60 min (nitrogen flow, 10 mL min⁻¹) to remove physisorbed species. The sample was then cooled to 30 °C (cooling rate, 20 °C min⁻¹; nitrogen flow, 10 mL min⁻¹), followed by stabilization at 30 °C for 30 min (nitrogen flow, 10 mL min⁻¹). CO₂ adsorption was conducted at 30 °C for 30 min under a pure CO₂ flow (100 mL min⁻¹). The adsorption capacity was determined from the mass increase during 30 min of CO₂ exposure at 30 °C following stabilization under nitrogen. Desorption was performed by reheating the sample to 200 °C (heating rate, 10 °C min⁻¹; nitrogen flow, 10 mL min⁻¹), followed by holding at 200 °C for 30 min and final cooling to 30 °C (cooling rate, 10 °C min⁻¹).

4. RESULTS AND DISCUSSION

4.1. Structural Confirmation by FTIR Spectroscopy. FTIR analysis of both the pure and functionalized MCM-41 specimens was performed to demonstrate retention of the silica framework upon CTAB removal and to confirm successful amine functionalization. The FTIR spectrum of the calcined pure MCM-41 (Fig. 1) shows a flat baseline at high wavenumbers, indicating the absence of organic impurities. Intense bands appear in the fingerprint region below ~1300 cm⁻¹, with a prominent broad peak in the 1000–1200 cm⁻¹ range, characteristic of asymmetric Si–O–Si stretching vibrations, consistent with previous studies^{9–11}. The weaker band observed near ~800 cm⁻¹ was attributed to the symmetric stretching mode of Si–O–Si. Additionally, at lower wavenumbers (~450–500 cm⁻¹), peaks corresponding to O–Si–O bending vibrations

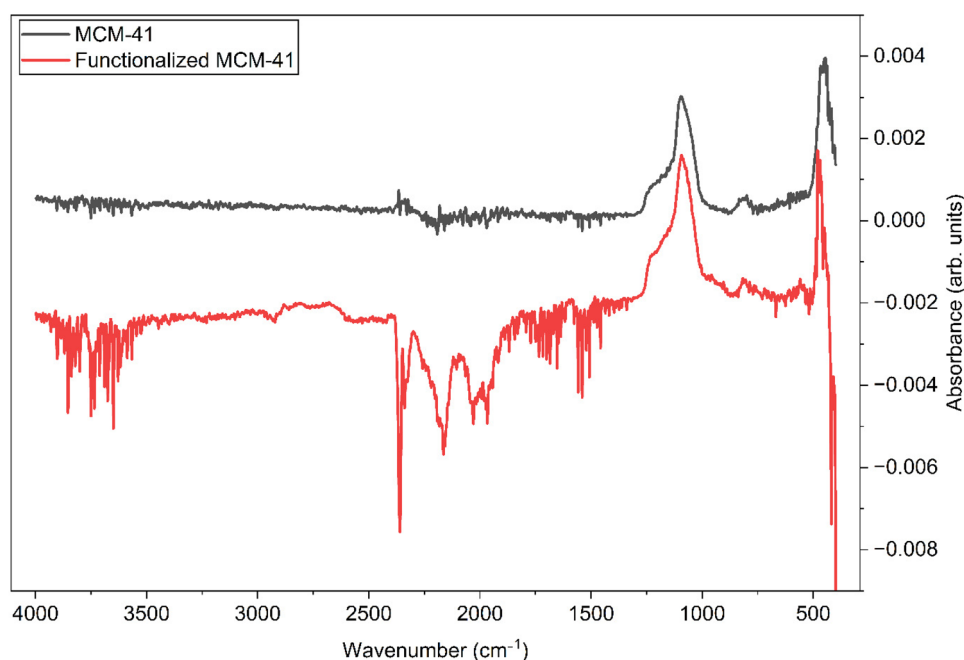


Figure 1. FTIR spectra of the pure MCM-41 and functionalized MCM-41 specimens

were observed, consistent with the expected lattice deformation modes in silica frameworks¹². Bands associated with surface hydroxyl (Si–OH) groups and adsorbed moisture were also present ($\sim 3200\text{--}3600\text{ cm}^{-1}$)¹³. The smaller band at $\sim 1640\text{ cm}^{-1}$ was attributed to the H–O–H bending vibrations of physisorbed water molecules. The presence of these bands confirms that MCM-41 retains its surface hydroxyl groups after calcination, which serve as anchoring sites for subsequent functionalization¹⁴. Notably, no significant C–H stretching peaks are present in the $2900\text{--}2850\text{ cm}^{-1}$ range, confirming effective CTAB removal during calcination¹⁵, consistent with previous reports¹⁴.

Upon comparison of the FTIR spectra recorded for the pure and functionalized MCM-41 samples, it is evident that the key fingerprint bands of the silica framework ($1200\text{--}800\text{ cm}^{-1}$) remain essentially unchanged after functionalization^{9,16}. This indicates that the hexagonal structure of mesoporous silica was preserved after functionalization, with no detectable framework degradation. However, in the functionalized sample, new weak features appeared in the C–H stretching region that were absent for the pure MCM-41. Specifically, bands at ~ 2920 and 2850 cm^{-1} correspond to the asymmetric and symmetric stretching vibrations of the $\text{--CH}_2\text{--}$ groups, respectively^{17,18}. These methylene stretches arise from the alkyl chains of the grafted organosilanes, confirming their introduction onto the silica surface. The low band intensities are consistent with the relatively modest organic loading. Since Hydrosil 2775 contains primary (and secondary) amine functionalities, the functionalized MCM-41 also showed evidence of N–H vibrations in the form of a broad, shallow feature in the $3300\text{--}3200\text{ cm}^{-1}$ region^{19,20}. These medium-intensity N–H signals overlap with the broad O–H band of the adsorbed water and silanol moieties, and their weakness reflects a low concentration of grafted amines. Likewise, the faint band at $\sim 1600\text{ cm}^{-1}$ was assigned to the N–H bending (scissoring) vibration of the primary amines. Although the N–H bending signal was difficult to distinguish due to overlap with the water

bending band ($\sim 1640\text{ cm}^{-1}$), its presence is consistent with the introduction of --NH_2 groups²⁰.

Functionalization with Hydrosil 2999 yielded an additional notable spectral feature. Specifically, a medium-intensity band appeared at $\sim 1470\text{ cm}^{-1}$, which was attributed to the combined $\text{N}^+\text{--CH}_2$ and C–N stretching/bending modes of the grafted quaternary ammonium group^{21,22}. Changes were also observed in the silanol region of the spectrum, wherein the shoulder at $\sim 960\text{--}970\text{ cm}^{-1}$ (assigned to the Si–OH stretching vibrations in pure MCM-41) was significantly diminished following amine functionalization. This observation is consistent with the reaction between silanols and organosilanes to form Si–O–Si–R linkages^{9,16}. This finding, together with the appearance of the new C–H and N–H bands, confirms successful amine functionalization and condensation of the organosilane molecules onto the MCM-41 surface while preserving the silica framework²².

4.2. Textural Evaluation of the Functionalized MCM-41. The textural properties of the functionalized MCM-41 were evaluated using N_2 adsorption–desorption isotherms at 77 K using a Belsorp Max X instrument. A sample mass of 0.04326 g was employed, as per the manufacturer’s instructions. Prior to analysis, the sample was degassed at 120°C for 6 h. The resulting adsorption–desorption isotherm (Fig. 2a) exhibits characteristics consistent with mesoporous materials (IUPAC classification), with a low-pressure region attributed to monolayer/multilayer adsorption, followed by pronounced uptake at intermediate relative pressures, consistent with capillary condensation within the mesopores.

BET analysis was performed on the adsorption branch over a relative pressure range (P/P_0) of 0.045–0.18, yielding a linear fit with a correlation coefficient of 0.999. From this fit, the monolayer capacity (V_m) was determined to be $87.2\text{ cm}^3(\text{STP})\text{ g}^{-1}$, the BET specific surface area (S_{BET}) was $379.54\text{ m}^2\text{ g}^{-1}$, and the BET constant (C) was 49.7. The total pore volume, measured at

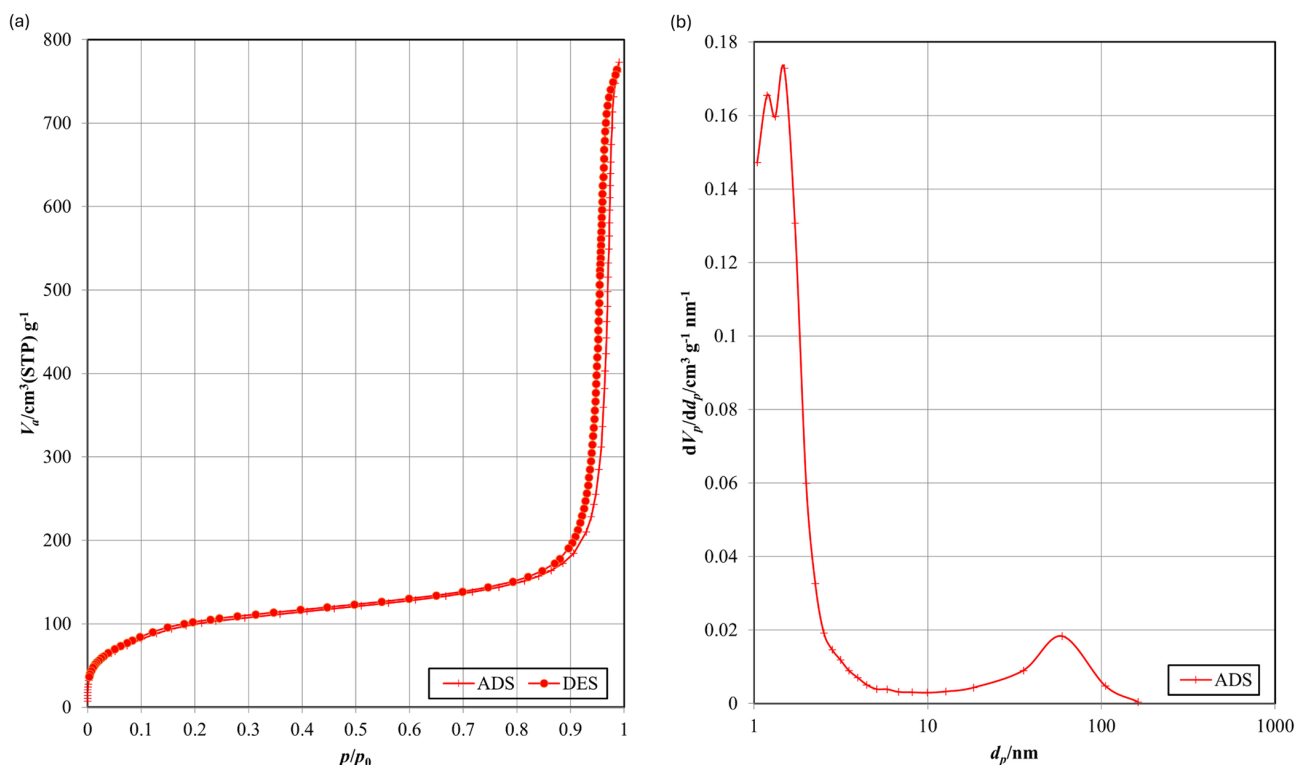


Figure 2. (a) N_2 adsorption/desorption isotherms, and (b) BJH pore size distribution (adsorption branch) for the functionalized MCM-41

$P/P_0 = 0.99$, was $1.1929\text{ cm}^3\text{g}^{-1}$, with an average pore diameter of 6.29 nm . These values indicate that the functionalized sample retains the substantial surface area and pore volume characteristic of the mesoporous silica; however, the surface area is lower than commonly reported ranges for unfunctionalized MCM-41 ($800\text{--}1500\text{ m}^2\text{g}^{-1}$ and $0.8\text{--}1.8\text{ cm}^3\text{g}^{-1}$), suggesting partial pore filling following amine condensation and grafting, consistent with prior reports^{23–25}. The pore size distribution on the adsorption branch was evaluated using the BJH method and NLDFT. BJH analysis (Fig. 2b) yielded a total mesopore volume of $1.2377\text{ cm}^3\text{g}^{-1}$, peak pore diameter of 8.57 nm , and BJH surface area of $577.7\text{ m}^2\text{g}^{-1}$. This pore diameter fell within the IUPAC-defined mesopore range ($2\text{--}50\text{ nm}$). This larger BJH-derived pore diameter relative to the BET-derived average diameter was consistent with the tendency of the BJH method to overestimate pore sizes in mesoporous silica^{26–28}.

NLDFT analysis (Fig. 3), performed using a cylindrical pore model for N_2 at 77 K , yielded a total pore volume of $1.2319\text{ cm}^3\text{g}^{-1}$, a main pore size of 2.34 nm , and an additional contribution in the macropore regime at $\sim 149.7\text{ nm}$. The dominant pore-size contribution in the $2\text{--}3\text{ nm}$ range was consistent with the mesoporous nature of MCM-41, whereas the macropore-scale contribution was attributed to interparticle voids arising from particle packing^{29,30}. The BJH and NLDFT pore volumes are in close agreement, confirming the overall mesopore nature, and revealing that the functionalized MCM-41 retains high pore volume and substantial surface area. This finding is relevant for accommodating grafted amine species and enabling CO_2 uptake within the porous structure^{31–33}.

4.3. CO_2 Uptake Analysis Using TGA. The CO_2 uptake behavior of the pure MCM-41 was evaluated for comparison with

the functionalized sample. Both samples were subjected to the thermal program described in Section 3.1.3. Considering only the CO_2 flow step, the corresponding results for the pure MCM-41 are shown in Fig. 4.

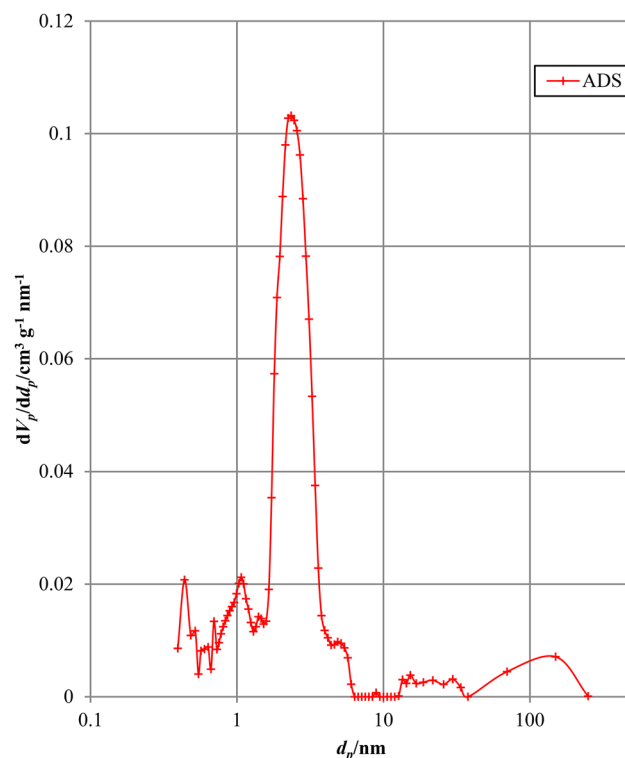


Figure 3. NLDFT/GCMC pore size distribution (adsorption branch)

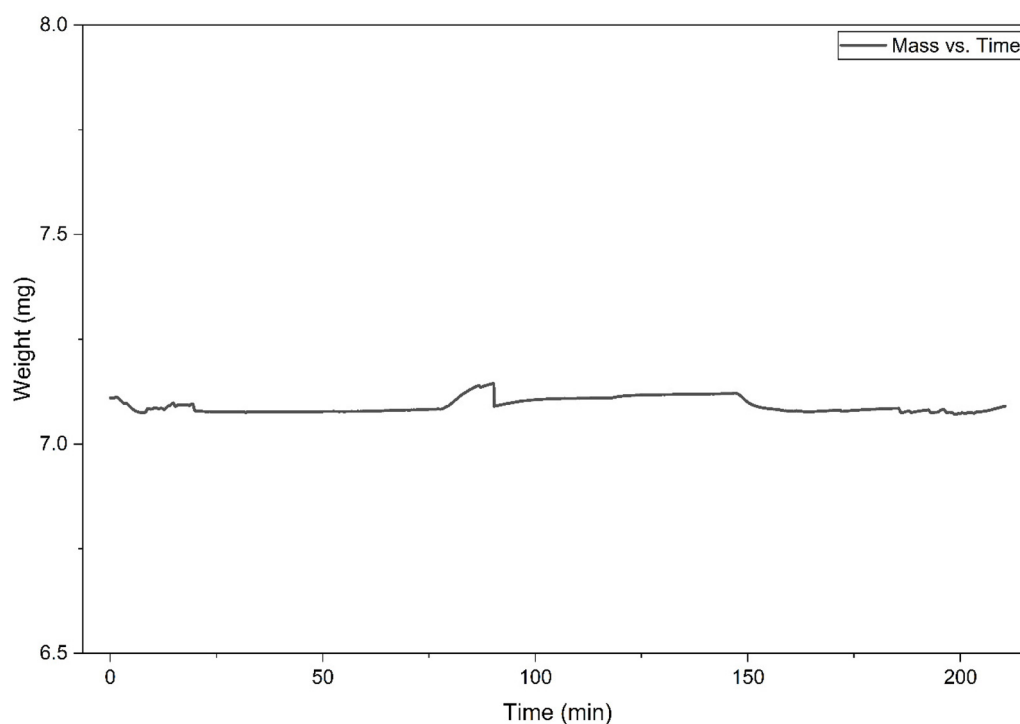


Figure 4. CO₂ adsorption by the pure MCM-41 during TGA

It can be observed that the weight of the pure MCM-41 increased only slightly (from 7.1056 to ~7.1204 mg), reflecting limited CO₂ adsorption, which is consistent with literature reports showing that MCM-41 exhibits weak CO₂ affinity at near-ambient temperatures. Considering that adsorption occurs mainly through weak physisorption interactions on the surface silanol groups, it is evident that MCM-41 serves primarily as a high-surface-area support material, which must be functionalized to achieve significant CO₂ capture capacity³⁴.

Figure 5 shows the CO₂ adsorption profile of the functionalized MCM-41 during the complete thermal program, including the degassing, stabilization, adsorption, and desorption steps. During CO₂ exposure (100 mL min⁻¹) at 30 °C, the mass of the functionalized sample increased progressively, reflecting CO₂ adsorption by the amine-modified surface.

From the CO₂ adsorption profile it is evident that the mass increased continuously throughout the 30 min CO₂ exposure period, without reaching a plateau. A net mass gain

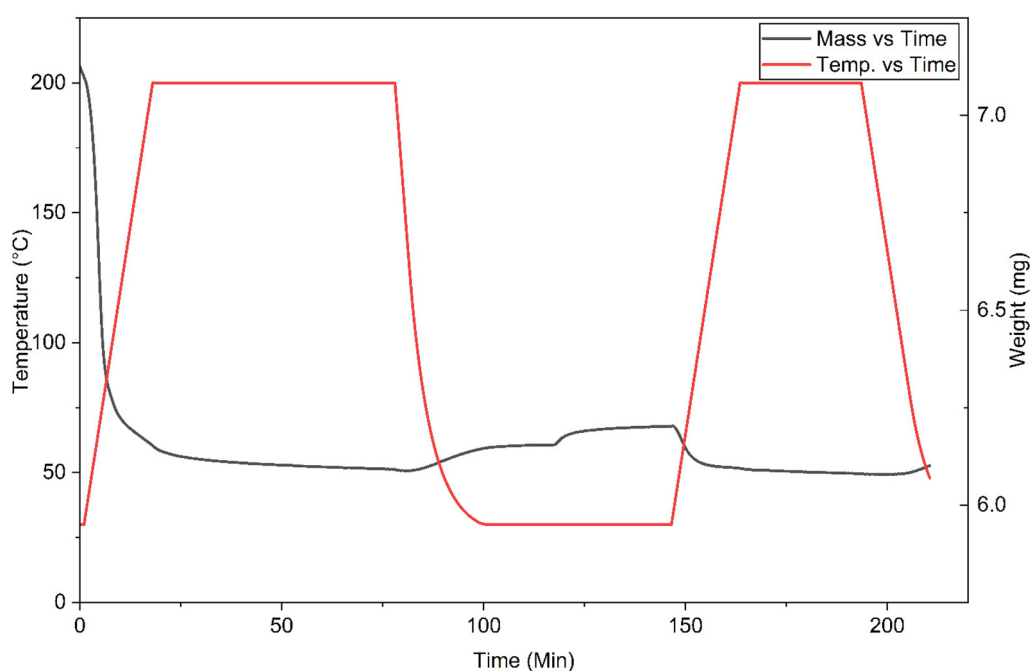


Figure 5. CO₂ adsorption of the functionalized MCM-41 during TGA

Table 1. Comparison of CO₂ capture performances of amine-functionalized sorbents

Support	Amine type and loading	CO ₂ uptake	Regeneration performance
MCM-41 ³⁵ (mesoporous silica)	APTS (30 wt% grafted) + TEPA (40 wt% impregnated)	3.50 mmol g ⁻¹ (ambient pressure, temperature not specified)	Negligible capacity loss over 10 adsorption–desorption cycles (high cyclic stability)
MCM-41 ³⁶	TEPA (30 wt%) + AMP (30 wt%) (co-impregnated blend)	3.01 mmol g ⁻¹ (at 70 °C, 15% CO ₂ feed)	~4% capacity drop after 15 cycles (3.01 → 2.88 mmol g ⁻¹), indicating good regenerability and thermal stability
SBA-15 (mesoporous silica) ³⁷	APTES (grafted) + TEPA (70 wt% impregnated)	5.69 mmol g ⁻¹ (at 75 °C, 20% CO ₂)	~6.1% capacity drop after 25 cycles (excellent stability under cycling)
MCM-41 ³⁷	EAE + TEPA (blended, total 60 wt% amine; optimal ratio: 50% EAE/10% TEPA)	4.25 mmol g ⁻¹ (at 25 °C, 15% CO ₂)	~9.4% capacity drop after 5 cycles (good reversibility; low heat of regeneration ~12 kJ mol ⁻¹ CO ₂)
KIT-6 (3D mesoporous silica) ³⁸	PEHA (pentaethylenehexamine, impregnated loading not specified)	~4.0–4.3 mmol g ⁻¹ (at 90–105 °C, 1 bar CO ₂)	<4% capacity loss over 10 cycles even after 6 months of use (very stable performance)
MCM-41 (current work)	Dual-amine functionalized Primary/secondary + tertiary amine (co-functionalized)	0.386 mmol g ⁻¹ (at 1.7 wt%, 30 °C, 1 bar CO ₂)	Fully regenerated by heating to ~150–200 °C

APTS = 3-aminopropyltrimethoxysilane (grafted surface amine); TEPA = tetraethylenepentamine; AMP = 2-amino-2-methyl-1-propanol (sterically hindered amine); APTES = 3-aminopropyltriethoxysilane; EAE = 2-(ethylamino)ethanol; PEHA = pentaethylenehexamine.

corresponding to 1.7 wt% was determined by calculating the difference between the stabilized baseline mass at 30 °C under N₂ and the mass following CO₂ exposure. Upon heating to 200 °C under N₂, the mass returned to its pre-adsorption value, demonstrating complete desorption of the adsorbed CO₂ and confirming the reversible nature of the adsorption process. The absence of a residual mass gain after desorption suggests that physisorption and weak chemisorption interactions dominated under the applied conditions.

To contextualize the measured CO₂ uptake, the performance of the functionalized MCM-41 was compared with reported values for amine-modified mesoporous silica materials (Table 1). Under dynamic TGA conditions (30 °C, flowing pure CO₂), the net mass gain of 1.7 wt% corresponds to a CO₂ uptake capacity of 0.386 mmol g⁻¹. As detailed in Table 1, reported CO₂ adsorption capacities for highly loaded impregnated systems typically range between 2 and 5 mmol g⁻¹ at 1 bar CO₂, depending on amine type, loading, and temperature. These results indicate that while the overall adsorption capacity remains modest relative to heavily impregnated systems, dual-amine grafting successfully introduces CO₂-adsorption functionality without compromising structural integrity.

5. CONCLUSIONS

In this study, the CO₂ adsorption behavior of MCM-41 functionalized with a dual-amine mixture of Hydrosil 2775 and Hydrosil 2999 (1:1 ratio) was investigated. Fourier-transform infrared spectroscopy analysis confirmed successful surface modification through the introduction of organic functionalities, while preserving the characteristic Si–O–Si framework of mesoporous silica. Nitrogen adsorption–desorption measurements

demonstrated that the functionalized material retained its mesoporosity with a high specific surface area, consistent with partial pore occupation by grafted organosilanes. Dynamic thermogravimetric analysis experiments performed under a flow of pure CO₂ at 30 °C revealed that the functionalized material achieved a net uptake of 1.7 wt% (0.386 mmol g⁻¹), with complete mass recovery upon thermal regeneration under N₂. Although the measured uptake is lower than values reported for highly loaded amine-impregnated MCM-41 systems, the current results confirm that dual-amine grafting successfully modifies surface chemistry and enables reversible CO₂ interaction without structural degradation. Overall, this study demonstrates the feasibility of Hydrosil-based dual functionalization of MCM-41 for CO₂ capture applications. Future studies should focus on optimizing amine loading, evaluating adsorption under equilibrium conditions, and further characterizing the adsorption mechanisms to enhance uptake performance.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Maymounah Alshahrani – Department of Materials Science and Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0009-0001-4369-4394
Email: s202251060@kfupm.edu.sa

Corresponding Author

Ahmed Wasel A Alsmaeil – Research Mentor, Professor, Department of Materials Science and Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0000-0001-7121-9876
Email: ahmed.smaeil@kfupm.edu.sa

ACKNOWLEDGEMENTS

The authors acknowledge support provided by the Department of Materials Science and Engineering and College of Petroleum Engineering and Geosciences Research Center at King Fahd University of Petroleum & Minerals, Saudi Arabia. The authors also extend their appreciation to the technical staff and laboratory team for their assistance with specimen preparation and testing.

REFERENCES

- (1) Gunawardene, O. H. P., Gunathilake, C. A., Vikrant, K. & Amaraweera, S. M. Carbon Dioxide Capture through Physical and Chemical Adsorption Using Porous Carbon Materials: A Review. *Atmosphere* vol. 13 Preprint at <https://doi.org/10.3390/atmos13030397> (2022).
- (2) Amaraweera, S. M. *et al.* Carbon Capture Using Porous Silica Materials. <https://doi.org/10.20944/PREPRINTS202306.0497.V1> (2023)
- (3) Bae, J. Y., Jang, S. G., Cho, J. & Kang, M. Amine-Functionalized Mesoporous Silica for Efficient CO₂ Capture: Stability, Performance, and Industrial Feasibility. *International Journal of Molecular Sciences* vol. 26 Preprint at <https://doi.org/10.3390/ijms26094313> (2025).
- (4) Kamarudin, K. S. N. & Alias, N. Adsorption performance of MCM-41 impregnated with amine for CO₂ removal. *Fuel Processing Technology* **106**, 332–337 (2013).
- (5) Ajumobi, O. *et al.* Design of Nanostraws in Amine-Functionalized MCM-41 for Improved Adsorption Capacity in Carbon Capture. *Energy and Fuels* **37**, 12079–12088 (2023).
- (6) Yildiz, M. G., Davran-Candan, T., Günay, M. E. & Yildirim, R. CO₂ capture over amine-functionalized MCM-41 and SBA-15: Exploratory analysis and decision tree classification of past data. *Journal of CO₂ Utilization* **31**, 27–42 (2019).
- (7) Bae, J. Y. *et al.* Amine-Functionalized Mesoporous Silica for Efficient CO₂ Capture: Stability, Performance, and Industrial Feasibility. *International Journal of Molecular Sciences* 2025, Vol. 26, **26**, (2025).
- (8) Rao, N. *et al.* CO₂ Adsorption by Amine-Functionalized MCM-41: A Comparison between Impregnation and Grafting Modification Methods. *Energy and Fuels* **32**, 670–677 (2018).
- (9) Wang, X., Guo, Q., Zhao, J. & Chen, L. Mixed amine-modified MCM-41 sorbents for CO₂ capture. *International Journal of Greenhouse Gas Control* **37**, 90–98 (2015).
- (10) Wang, X., Chen, L. & Guo, Q. Development of hybrid amine-functionalized MCM-41 sorbents for CO₂ capture. *Chemical Engineering Journal* **260**, 573–581 (2015).
- (11) Muchan, P., Saiwan, C. & Nithitanakul, M. Carbon dioxide adsorption/desorption performance of single- and blended-amines-impregnated MCM-41 mesoporous silica in post-combustion carbon capture. *Clean Energy* **6**, 424–437 (2022).
- (12) Zhai, Q.-Z. & Li, X.-D. Preparation and Characterization of (MCM-41)-Eu 2 O 3 Nanocomposite Materials. *Asian Journal of Chemistry* vol. 24 (2012).
- (13) Torabinejad, A., Nasirizadeh, N., Yazdanshenas, M. E. & Tayebi, H. A. Synthesis of conductive polymer-coated mesoporous MCM-41 for textile dye removal from aqueous media. *Journal of Nanostructure in Chemistry* 2017 7:3 7, 217–229 (2017).
- (14) Kang, Y. *et al.* 3D reconstruction and material analysis of Neolithic wattle-and-daub houses at Fenghuangzui site in China. *npj Heritage Science* 2026 14:1 **14**, 91- (2026).
- (15) Husna, N. *et al.* Magnetic Mesoporous Silica Composite for Enhanced Preconcentration of Selected Organophosphorus Pesticides in Fruits. *Indones. J. Chem* **2020**, 729–745.
- (16) Mersellem, K. *et al.* Synthesis and Characterization of Mesoporous Materials Functionalized with Phosphinic Acid Ligand and Their Capability to Remove Cd(II). *Molecules* **29**, 5199 (2024).
- (17) Hooshyar, M. R., Raygan, S. & Mehdiavaz aghdam, R. Investigating layer-by-layer chitosan-dextran sulfate-coated mesoporous silica as a pH-sensitive drug delivery system. *J. Mater. Sci. Mater. Med.* **35**, 29 (2024).
- (18) Pires, I. C. B. *et al.* Theranostics Using MCM-41-Based Mesoporous Silica Nanoparticles: Integrating Magnetic Resonance Imaging and Novel Chemotherapy for Breast Cancer Treatment. *International Journal of Molecular Sciences* 2024, Vol. 25, **25**, (2024).
- (19) Gomez, C. D., Rodriguez-Paez, J. E., Gomez, C. D. & Rodriguez-Paez, J. E. Photocatalytic Properties of Nb/MCM-41 Molecular Sieves: Effect of the Synthesis Conditions. *Coatings* 2015, Vol. 5, Pages 511–526 **5**, 511–526 (2015).
- (20) Ambrozewicz, D. *et al.* Fluoroalkylsilane versus Alkylsilane as Hydrophobic Agents for Silica and Silicates. *J. Nanomater.* **2013**, 631938 (2013).
- (21) Araújo, M. J. G. De *et al.* HDPE/Chitosan Blends Modified with Organobentonite Synthesized with Quaternary Ammonium Salt Impregnated Chitosan. *Materials* 2018, Vol. 11, **11**, (2018).
- (22) 24.10: Spectroscopy of Amines - Chemistry LibreTexts. https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Organic_Chemistry_%28Morsch_et_al.%29/24%3A_Amines_and_Heterocycles/24.10%3A_Spectroscopy_of_Amines?utm_source=chatgpt.com.
- (23) Infrared Spectroscopy-Analytical chemistry laboratory, ICT Prague.
- (24) Bajrami, D. *et al.* A scalable route to quaternary ammonium-functionalized AgCl colloidal antimicrobials inhibiting food pathogenic bacteria and biofilms. *Heliyon* **10**, e25260 (2024).
- (25) Liu, H., Guo, C., Cui, Y., Yin, J. & Li, S. Experimental and modeling investigation of organic modified montmorillonite with octyl quaternary ammonium salt. *Sci. Rep.* **12**, (2022).
- (26) Karczmarzka, A., Laskowska, W., Stróż, D. & Pawlik, K. Inside the Framework: Structural Exploration of Mesoporous Silicas MCM-41, SBA-15, and SBA-16. *Materials* **18**, 3597 (2025).
- (27) Serrano, R., Cuesta Zapata, P., Gonzo, E. & Parentis, M. AMINE-GRAFTED MESOPOROUS SILICA FOR CO₂ CAPTURE.
- (28) Barbosa, M. N. *et al.* Aminopropyltriethoxysilane functionalized MCM-41 and SBA-15 nanostructured materials for carbon dioxide adsorption. *Matéria (Rio de Janeiro)* **26**, e13085 (2021).
- (29) Thommes, M. *et al.* IUPAC Technical Report Physisorption of gases, with special reference to the evaluation of surface area and pore size distribution (IUPAC Technical Report). <https://doi.org/10.1515/pac-2014-1117> (2015).
- (30) Cychosz, K. A. & Thommes, M. Progress in the Physisorption Characterization of Nanoporous Gas Storage Materials. *Engineering* **4**, 559–566 (2018).
- (31) Sotomayor, F. J., Cychosz, K. A. & Thommes, M. Characterization of Micro/Mesoporous Materials by Physisorption: Concepts and Case Studies. *Acc. Mater. Surf. Res* **3**, 34–50 (2018).
- (32) Salmas, C. E. *et al.* An investigation of the physical structure of MCM-41 novel mesoporous materials using a corrugated pore structure model. *Appl. Catal. A Gen.* **216**, 23–39 (2001).
- (33) Carta, P., Cara, C., Cannas, C. & Scorciapino, M. A. Experiments-Guided Modeling of MCM-41: Impact of Pore Symmetry on Gas Adsorption. *Adv. Mater. Interfaces* **9**, 2201591 (2022).
- (34) Muchan, P., Saiwan, C. & Nithitanakul, M. Carbon dioxide adsorption/desorption performance of single- and blended-amines-impregnated MCM-41 mesoporous silica in post-combustion carbon capture. *Clean Energy* **6**, 424–437 (2022).

(35) Gholami, M., Talaie, M. R. & Aghamiri, S. F. CO₂ adsorption on amine functionalized MCM-41: Effect of bi-modal porous structure. *J. Taiwan Inst. Chem. Eng.* **59**, 205–209 (2016).

(36) Rao, N. *et al.* CO₂ Adsorption by Amine-Functionalized MCM-41: A Comparison between Impregnation and Grafting Modification Methods. *Energy and Fuels* **32**, 670–677 (2018).

(37) Karimi, M., Shirzad, M., Silva, J. A. C. & Rodrigues, A. E. Carbon dioxide separation and capture by adsorption: a review. *Environ. Chem. Lett.* **21**, 1 (2023).

(38) Amaraweera, S. M. *et al.* Carbon Capture Using Porous Silica Materials. *Nanomaterials* 2023, Vol. 13, Page 2050 **13**, 2050 (2023).

Platinum-loaded Molybdenum Phosphide Dispersed on Carbon Nanotubes for Improved Hydrogen Evolution Performance

Abdullah Alismail and Mohammad Qamar*

Cite <https://doi.org/10.64589/juri/219226>

Submitted: November 27, 2025 Revised: February 19, 2026 Accepted: February 25, 2026

ABSTRACT

The development of efficient and cost-effective electrocatalysts for the hydrogen evolution reaction (HER) is crucial for achieving sustainable hydrogen production. Molybdenum phosphide (MoP) represents a promising non-noble-metal catalyst whose HER overpotential is higher than that of Pt. In this study, Pt-loaded MoP dispersed on carbon nanotubes (Pt-MoP/CNTs) is fabricated via an oxalate-assisted molecular self-assembly route, followed by phosphorization and low-temperature Pt incorporation. The introduction of an ultralow Pt loading (2 wt%) significantly enhanced the HER performance in acidic media. At a current density of 20 mA cm^{-2} , the Pt-MoP/CNTs required an overpotential of 70 mV, which is markedly lower than that required by pristine MoP/CNTs (180 mV). The Pt-modified catalyst also exhibited a high double-layer capacitance of 532 s^{-1} at an overpotential of 28 mV, together with a large electrochemically active surface area, improved charge-transfer kinetics, and excellent stability over 48 h of continuous operation. The synergistic interaction between Pt and MoP effectively boosted catalytic activity while minimizing the use of noble metals. This work demonstrates an efficient strategy for the design of high-performance and economically viable HER electrocatalysts through trace-platinum modification of transition-metal phosphides.

Keywords: hydrogen production, electrocatalysts, energy conversion, transition metal phosphide, nanomaterials, water electrolysis

1. INTRODUCTION

Large-scale hydrogen production via electrochemical or photo-electrochemical methods has great potential owing to its low energy requirements, high hydrogen purity, and environmental friendliness¹. Although platinum is widely used as a catalyst in this process, the high cost of Pt and its instability in electrolyzer units renders its large-scale use challenging in hydrogen production². Consequently, extensive research efforts have focused on designing cathodes electrocatalysts composed of transition-metal to promote the hydrogen evolution reaction (HER). To date, several nonprecious metals, including metal carbides³, sulfides⁴, nitrides⁵, oxides⁶, borides⁷, and phosphides⁸, have been studied to reduce the HER overpotential⁹. Although significant progress has been made, considerable scope remains to further reduce it toward the level achieved by Pt. In recent years, transition metal phosphides have demonstrated high activity and stability in the HER, rendering them promising candidates for replacing noble-metal-based electrocatalysts⁸. In such systems, the overpotential can be reduced by carefully selecting a suitable electrocatalyst composition. Additionally, loading metal phosphides with small amounts of Pt could represent an effective strategy for reducing the overpotential and enhancing HER performance¹⁰. Owing to their distinctive d-band electronic structure, molybdenum-based materials have emerged as low-cost candidates for the HER. Notably, molybdenum phosphide (MoP), which is well known

for its catalytic activity in hydrosulfurization, has recently attracted attention as a promising HER electrocatalyst¹¹.

Specifically, Wang et al. were the first to demonstrate MoP as a possible HER electrocatalyst¹², revealing that MoP exhibits phase-dependent HER activity similar to that of Mo carbides. Specifically, following the synthesis of crystalline Mo_3P and MoP, they found that the metal-rich form (Mo_3P) required a significantly greater overpotential ($\sim 500 \text{ mV}$) than stoichiometric MoP (125 mV) to generate a current density of 10 mA cm^{-2} . They also reported that phosphorization alters the characteristics of the metal, with the degree of phosphorization affecting both its stability and activity. In another study, Zhao et al. reported that the addition of single Pt atoms to three-dimensional flower-like NiMoP increased the overpotential required to achieve a current density of 10 mA cm^{-2} from 200 to 35 mV¹³. While significant strides have been made in reducing overpotential, there is still substantial room to close the gap even further, aiming for performance that rivals pure platinum. Recent breakthroughs have highlighted transition metal phosphides (TMPs) integrated with platinum as a promising solution; these composites deliver high, stable Hydrogen Evolution Reaction (HER) activity with minimal overpotential, positioning them as a viable and more efficient alternative to traditional noble-metal catalysts.

This study therefore aims to assess the effectiveness of adding Pt to carbon nanotube (CNT)-supported MoP to promote the HER. Specifically, the molecular self-assembly of Mo—which

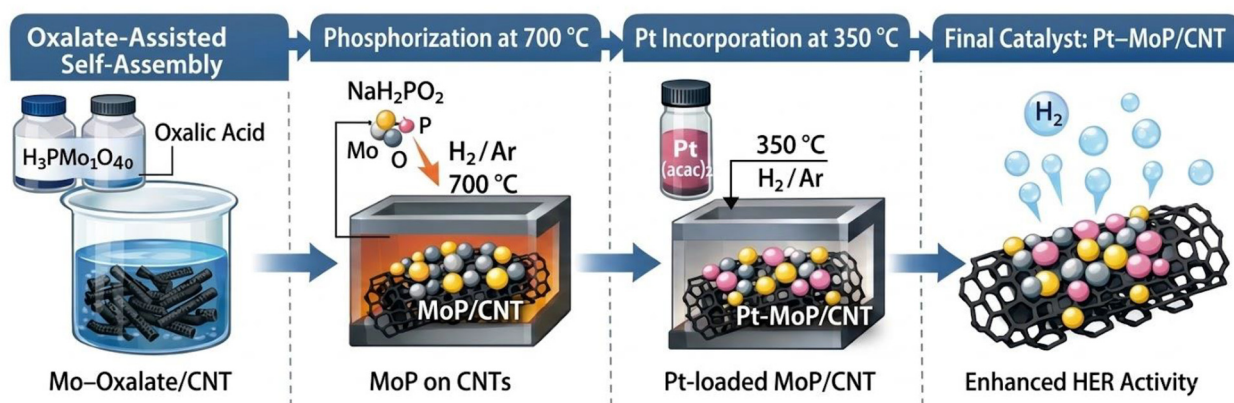


Figure 1. Schematic illustration showing synthesis of the Pt-MoP/CNT electrocatalyst

is coordinated with the CNT oxalate groups—is carefully controlled to enhance Mo immobilization and increase the Mo–Mo distance through steric hindrance. The electrocatalytic performance of the MoP/CNT system is examined in acidic media and compared with that of MoP/CNTs containing 2 wt% added Pt (Pt-MoP/CNTs). The roles of the highly specific and electrochemically active surface areas, high charge transfer kinetics, and low electrochemical impedance on defining the electrocatalytic performance are also evaluated².

2. MATERIALS AND METHODS

2.1. Synthesis of the MoP/CNT Electrocatalyst. To prepare the MoP/CNT electrocatalyst, a homogeneous suspension of CNTs in ethanol was prepared in a closed vial. Subsequently, a calculated amount of phosphomolybdic acid ($\text{H}_3[\text{PMo}_{12}\text{O}_{40}]$) was added to achieve a 1:1 ratio with the CNTs, and the mixture was stirred to promote complete dissolution of the precursor. Oxalic acid was then added and stirring was performed at 80 °C to evaporate ethanol and yield a black powder. This powder was dried overnight in an oven at 50 °C and subsequently subjected to phosphorization in a tubular furnace using a hypophosphite/Mo ratio of 20:1 under a H_2/Ar (10:90) gas flow at 700 °C (heating rate: 5 °C min^{-1}) for 2 h. During phosphorization, the gas flow was maintained at 100 mL min^{-1} using a mass flow controller (MFC, Alicat Scientific, USA).

After phosphorization, a calculated amount of Pt was used to achieve a loading of 2 wt% on the MoP surface. For this purpose, the Pt source and MoP powder were separately dispersed in ethanol and combined, and the mixture was stirred at 80 °C until the solvent completely evaporated, yielding a black powder. The resulting product was then dried in an oven at 50 °C overnight prior to treatment in a tubular furnace under a H_2/Ar (10:90) gas flow at 350 °C (heating rate: 5 °C min^{-1}) for 2 h to anchor Pt onto the MoP structure. The synthetic protocol is summarized in Figure 1. The black powder was collected then evaluated as a potential HER electrocatalyst.

3. CHARACTERIZATION

The microstructural and morphological attributes of the materials were determined using X-ray diffractometry (XRD, Rigaku

MiniFlex), scanning electron microscopy (SEM, TESCAN, Czech Republic), and electron diffraction spectroscopy (EDS, TESCAN, Czech Republic). Electrochemical measurements were conducted using an SP-300 potentiostat/galvanostat (Bio-Logic Science Instruments, France). Specifically, linear sweep voltammetry (LSV), electrochemical impedance spectroscopy (EIS), electrochemical stability tests, electrochemical surface area (ECSA) analysis, cyclic voltammetry (CV), and inductively coupled plasma optical emission spectroscopy (ICP-OES; Thermo Scientific, USA) were performed.

3.1. Evaluation of the Electrocatalytic Activity. A homogeneous ink solution was prepared by sonicating a suspension consisting of the electrocatalyst (7.5 mg), isopropanol (1 mL), and Nafion (16.5 μL) for 30 min. Subsequently, a fixed amount of the catalyst ink (12.5 μL) was drop-cast onto a cleaned glassy carbon (GC) disc electrode (0.07 cm^2 , Pine Instruments) and allowed to dry under ambient airflow. The deposition process was repeated to achieve the desired catalyst loading on the GC electrode. HER measurements were performed using a three-electrode setup connected to a potentiostat. Silver/silver chloride (Ag/AgCl) served as the reference electrode, and a coiled Pt wire was used as the counter electrode. A 0.5 M H_2SO_4 aqueous solution served as the working electrolyte. LSV was conducted at a scan rate of 5 mV s^{-1} . The Ag/AgCl reference electrode was converted against the reversible hydrogen electrode (RHE) scale through calibration. Unless otherwise noted, the current densities were calculated based on the geometric area of the GC electrode and reported after iR correction.

4. RESULTS AND DISCUSSION

A plausible mechanism for the molecular assembly, phosphorization, and subsequent Pt incorporation to yield the target Pt-MoP/CNT electrocatalyst was proposed. Initially, phosphomolybdic acid ($\text{H}_3[\text{PMo}_{12}\text{O}_{40}]$) dissolves in ethanol and rapidly interacts with the solvent, forming an intermediate species together with alcohol-derived byproducts. Upon the addition of oxalic acid, an ion-exchange process occurs, in which ethoxy groups are replaced by oxalate ligands, forming an Mo–oxalate complex. Progression of ligand exchange was confirmed visually from a change in solution color, indicating rearrangement of the Mo coordination environment¹¹.

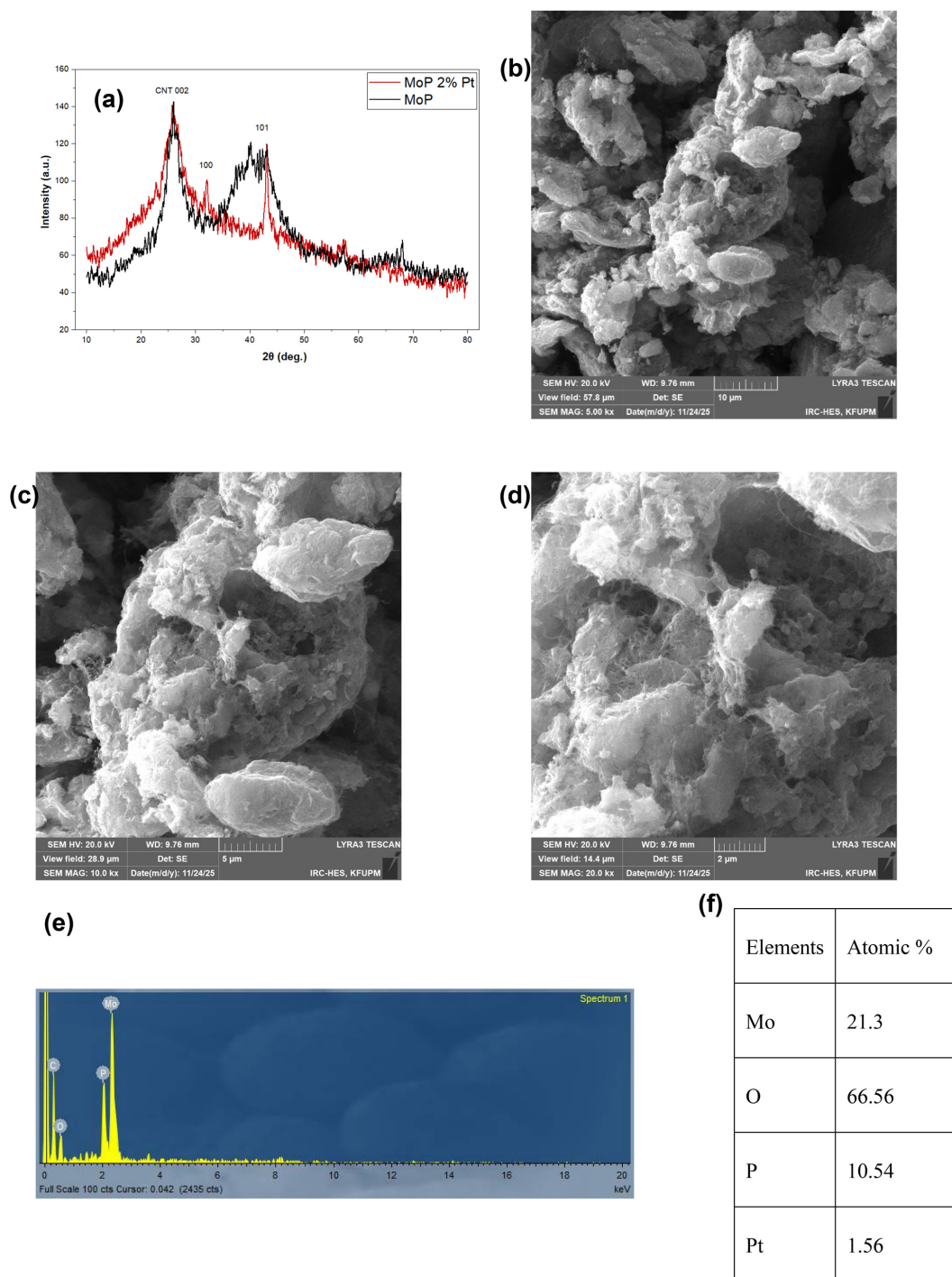


Figure 2. (a) XRD patterns of the MoP/CNT species with and without 2 wt% Pt loading. (b–d) SEM images of the Pt-MoP/CNT catalyst. (e) EDS spectrum showing the elemental composition of the Pt-MoP/CNT catalyst. (f) Elemental composition, as determined by EDS

The formation of this Mo–oxalate complex facilitated controlled molecular self-assembly. Specifically, steric bulk from the oxalate group increased the spacing between adjacent Mo centers, preventing early aggregation. Additionally, the oxygen-rich oxalate ligands interacted strongly with the surface functionalities of the CNTs, enhancing Mo attachment and ensuring uniform dispersion along the carbon framework. This organized molecular arrangement was expected to facilitate Mo clustering and guide more uniform nucleation and growth processes¹¹.

Upon heating, the Mo–oxalate complex underwent thermal decomposition. In the presence of sodium hypophosphite (NaH_2PO_2) as a phosphorus source, the Mo precursor reacted with $\text{PO}_2^-/\text{PO}_3^{3-}$ species to produce molybdenum phosphate intermediates. Under a reducing H_2/Ar atmosphere, these intermediates were gradually transformed into amorphous MoP. Upon increasing the temperature to 700 °C, the amorphous phase crystallized, yielding well-defined MoP nanoparticles distributed uniformly across the CNT surfaces¹¹.

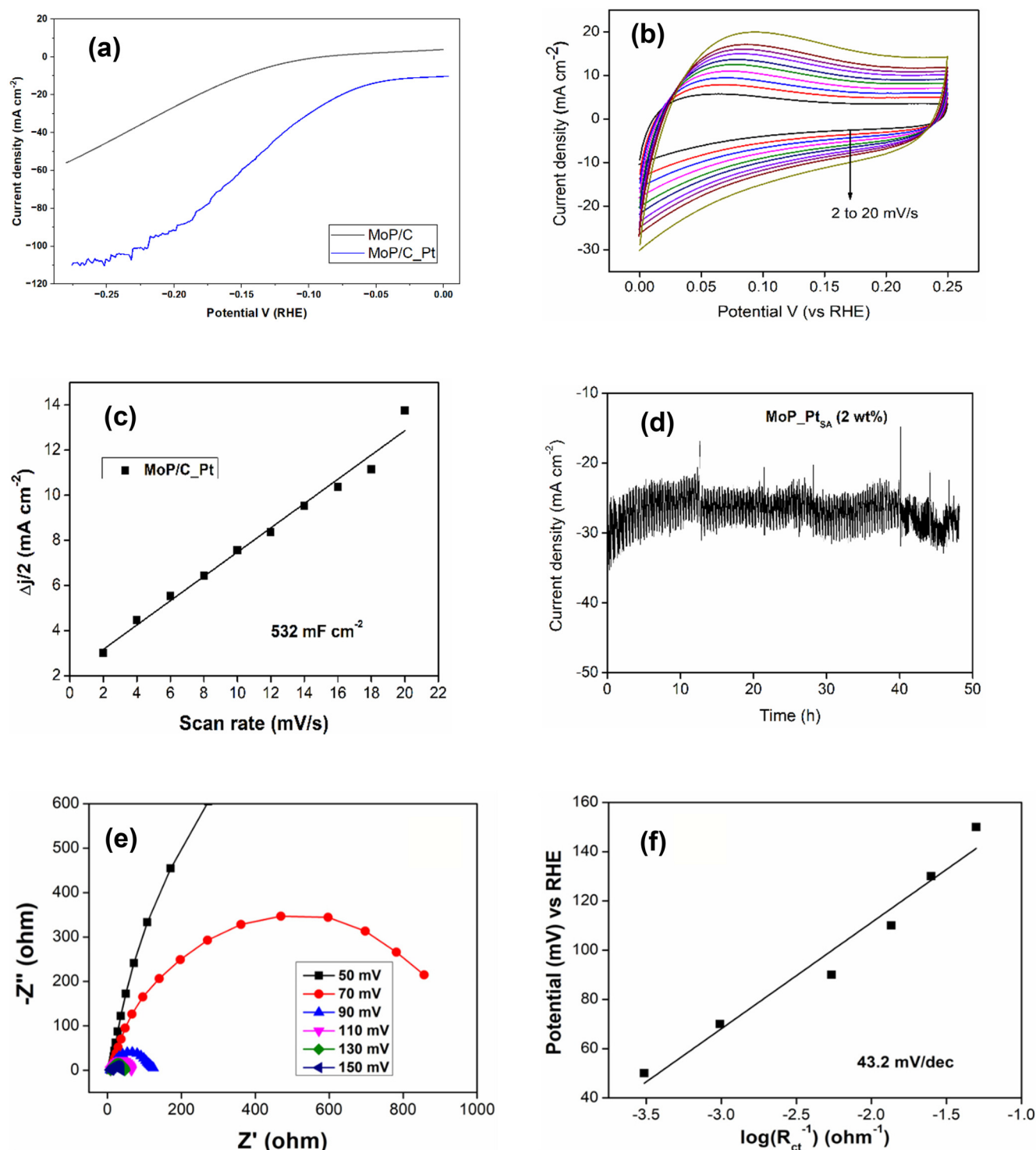


Figure 3. Electrocatalytic behavior of the electrocatalysts: (a) LSV results for the MoP/CNTs and Pt-MoP/CNTs, recorded at a scan rate of 2 mV s^{-1} ; (b) CV results for the Pt-MoP/CNT catalyst in H_2SO_4 ; (c) ECSA calculation; (d) evaluation of the chronoamperometric stability of the catalyst during a 48 h HER, showing sustained catalytic performance; (e) potential-dependent Nyquist plot recorded for Pt-MoP, and (f) the corresponding Tafel slope

Subsequently, Pt(II) acetylacetonate was mixed with the MoP/CNT precursor in ethanol. Strong interactions between the Pt species and the surfaces of the MoP/CNTs promoted the controlled adsorption of Pt complexes. Upon subsequent thermal treatment in a mildly reducing H_2/Ar environment at 350°C , the Pt precursor decomposed to generate metallic Pt nanoclusters and atomically dispersed Pt species on the MoP surface.

This lower treatment temperature ensured minimal sintering and preserved the atomic-level dispersion at 2 wt% Pt loading.

The incorporation of Pt modified the electronic structure of MoP by tuning the d-band center and creating electronic adjustments stemming from oxalate-assisted Mo dispersion, hypophosphite-driven phosphorization, and controlled Pt loading, collectively contributing to improved charge transfer

processes, enhanced electrochemically active surface area, and superior catalytic performance for the HER^{11,14}.

The powder XRD patterns of MoP prepared both with and without Pt loading are compared in Figure 2(a). For the sample prepared without Pt, the diffraction peak at 26.1° was assigned to the 002 plane of the CNTs, while the diffraction peaks centered at 27.9° (002), 31.2° (100), and 43.1° (101) were indexed to MoP with a hexagonal crystal structure¹⁴. A similar XRD pattern was observed for the specimen containing 2 wt% Pt, suggesting that the Pt loading is sufficiently low that it cannot be detected by XRD. Additionally, the increased temperature required for Pt incorporation appeared to have no influence on the XRD pattern, indicating that the electrocatalyst was stable at high temperatures. Notably, these XRD patterns are comparable to those reported by Adam et al.¹¹.

The SEM images shown in Figures 2(b)–2(d) show clusters of Pt-loaded MoP on the CNT support. The overall structure appeared rough and irregular, with the MoP particles appearing as bright and dense regions attached to or surrounding the nanotube network. The CNTs maintained their shape and acted as a porous scaffold for MoP particle anchoring. Some particles were embedded within the nanotube bundles, suggesting that growth occurs inside the tube structure. Overall, the SEM images confirmed that MoP was uniformly dispersed across the CNT surfaces to form a crystalline network. Notably, the Pt particles were not observable by SEM due to the low Pt loading.

Figures 2(e) and 2(f) show the EDS spectrum and quantitative elemental composition of the Pt-MoP/CNT electrocatalyst, respectively, confirming the presence of Mo (21.34%), P (10.54%), O (66.56%), and Pt (1.56%). The relatively high O content was attributed to surface oxidation caused by sample exposure to air. Although the Pt signal was not clearly distinguishable in the EDS spectrum owing to its extremely low concentration, quantitative analysis confirmed a Pt atomic percentage of 1.56%, consistent with the target ultralow Pt loading.

To accurately determine the Pt loading, ICP-OES was employed following complete acid-digestion of the catalyst. This yielded a measured Pt content of 11820.7 ppm, corresponding to 1.18 wt%, which is slightly lower than the nominal 2 wt% introduced during synthesis. This deviation from the theoretical value was attributed to Pt loss during the thermal treatment steps.

Subsequently, the HER performances of the prepared electrocatalysts were evaluated in acidic media, and the current density was calculated based on the geometric area of the GC electrode. Performance was compared using the same catalyst loading and conditions. It was revealed that Pt addition significantly enhanced the activity, with the Pt-MoP/CNT system requiring a lower potential to drive the HER, compared with the MoP/CNTs alone (Figures 4). This result confirms that the below-nominal Pt loading was sufficient to substantially improve catalytic activity, underscoring the efficient utilization of trace Pt in the designed catalyst system.

Figures 3 (a) shows the LSV results for the MoP/CNTs and Pt-MoP/CNTs measured in 0.5 M H_2SO_4 at a scan rate of 5 mV s^{-1} . The polarization curves were recorded as a function of the applied potential versus RHE, with the cathodic current density increasing at more negative potentials as a result

of the HER. Compared with the pristine MoP/CNT catalyst, the Pt-MoP/CNT catalyst exhibited significantly higher current densities across the entire potential range, indicating enhanced electrocatalytic activity. According to Figure 4(a), the overpotentials required to produce a current density of 20 mA cm^{-2} for the MoP/CNT and Pt-MoP/CNT catalysts were 180 and 70 mV, respectively. The Pt-modified catalyst achieved larger cathodic currents at lower overpotentials, reflecting improved reaction kinetics and reduced energy barriers for proton reduction. In contrast, the MoP/CNT catalyst required more negative potentials to reach comparable current densities, suggesting slower HER kinetics. This enhanced polarization behavior of the Pt-MoP/CNTs confirms that ultralow Pt incorporation effectively promotes charge-transfer and improves the overall HER performance.

Figures 3(b) presents a comparison of the cyclic CV curves recorded for the Pt-MoP/CNT electrode in acidic medium with a scan rate of 5 mV s^{-1} . Additionally, Figures 3(c) outlines determination of the ECSA based on the double-layer capacitance (C_{dl}), which was obtained from the linear relationship between $\Delta j/2$ and the scan rate. Specifically, upon increasing the scan rate from 2 to 20 mV s^{-1} , the $\Delta j/2$ value increased linearly, indicating that the electrode exhibits stable capacitive response. From the slope of the linear fit, the calculated C_{dl} was determined to be 532 s^{-1} , suggesting the presence of a high density of electrochemically accessible active sites in the Pt-MoP/CNT catalyst, thereby contributing to its strong electrochemical performance.

Chronoamperometric stability tests were subsequently performed over 48 h in acidic media at a scan rate of 5 mV s^{-1} . As shown in Figures 3(d), the current trend remained steady throughout the test, with no noticeable drop in current density. The overpotential required to produce the same current density remained essentially unchanged after 48 h, demonstrating excellent long-term electrode stability and endurance under acidic conditions.

Figures 3(e) illustrates the potential-dependent evolution of the Nyquist plots for Pt-MoP/CNT, which were constructed as a function of overpotential ($\eta = 50\text{--}150 \text{ mV vs RHE}$). The semicircular features in the impedance spectra reflect charge-transfer at the electrode–electrolyte interface. Smaller semicircular arcs correspond to faster interfacial charge transfer, whereas larger arcs indicate slower kinetics [1]. The obtained results therefore indicate accelerated electron-transfer kinetics with increasing overpotential. Moreover, the Tafel slopes derived from the EIS measurements are shown in Figures 3(f), wherein the semi-logarithmic plot of the inverse charge-transfer resistance ($1/R_{ct}$) versus overpotential shows a linear relationship with a defined slope. Consequently, the Tafel slope for Pt-MoP was determined to be 43.2 mV dec^{-1} , indicating that the HER proceeded predominantly via the previously reported Volmer–Heyrovsky mechanism on this electrocatalyst, consistent with our earlier findings^{15,16}.

5. CONCLUSIONS

In this study, a platinum-loaded molybdenum phosphide electrocatalyst supported on carbon nanotubes (Pt-MoP/CNTs) was synthesized via an oxalate-assisted molecular self-assembly route,

followed by controlled phosphorization and low-temperature Pt incorporation. The oxalate ligand played a key role in dispersing the Mo precursors and preventing aggregation, enabling uniform nucleation and growth of MoP across the CNT framework. Phosphorization using hypophosphite produced well-defined MoP nanoparticles, and subsequent treatment with a suitable Pt precursor under mild reducing conditions generated ultralow-load Pt species without altering the MoP crystal structure.

Comprehensive structural characterization confirmed the preservation of the MoP phase and its homogeneous distribution on the CNTs, while electrochemical measurements demonstrated the strong influence of Pt incorporation on catalytic behavior. Compared to the pristine MoP/CNTs, the Pt-modified catalyst exhibited significantly enhanced hydrogen evolution reaction (HER) activity in acidic media, including reduced overpotential, higher electrochemically active surface area, and notably lower charge-transfer resistance. The Pt-MoP/CNT catalyst also exhibited excellent operational durability, maintaining stable current density over extended testing periods.

Overall, these findings demonstrate that the incorporation of trace Pt substantially enhances the intrinsic activity of MoP through synergistic electronic interactions and the formation of highly active interfacial sites. This study therefore provides an effective, scalable, and economically attractive strategy for designing next-generation HER electrocatalysts that combine the robustness of transition-metal phosphides with trace noble-metal enhancement. Since the incorporation of platinum significantly enhances the electrochemical performance, this strategy is likely to open new avenues for the systematic investigation and optimization of platinum loading percentages to achieve superior synergistic effects for advanced energy conversion applications.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Abdullah Alismail – Department of Material Science and Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0009-0001-4055-7407
Email: s202176350@kfupm.edu.sa

Corresponding Author

Mohammad Qamar – Research Mentor, Associate Professor, Department of Material Science and Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0000-0002-5351-9872
Email: qamar@kfupm.edu.sa

ACKNOWLEDGEMENTS

The authors express their gratitude for the support received from the Interdisciplinary Research Center for Hydrogen Technologies and Carbon Management of King Fahd University of Petroleum & Minerals (KFUPM), Saudi Arabia. The authors also sincerely thank Dr. Umair Alam and Tsegaye Markos Damite

for their technical guidance, insightful discussions, and assistance with experimental work and data analysis throughout this research.

REFERENCES

- (1) Turner, J. A. Sustainable hydrogen production. *Science* (1979). **305**, 972–974 (2004).
- (2) Araújo, H. F., Gómez, J. A. & Santos, D. M. F. Proton-Exchange Membrane Electrolysis for Green Hydrogen Production: Fundamentals, Cost Breakdown, and Strategies to Minimize Platinum-Group Metal Content in Hydrogen Evolution Reaction Electrocatalysts. *Catalysts* **14**, (2024).
- (3) Joo, S. H. & Lee, J. S. Metal carbides as alternative electrocatalysts for energy conversion reactions. *J. Catal.* **404**, 911–924 (2021).
- (4) Hu, W. *et al.* The nature of active sites of molybdenum sulfide-based catalysts for hydrogen evolution reaction. *Coord. Chem. Rev.* **506**, 215715 (2024).
- (5) Park, S. H. *et al.* Highly active and stable nickel–molybdenum nitride (Ni₂Mo₃N) electrocatalyst for hydrogen evolution. *J. Mater. Chem. A Mater.* **9**, 4945–4951 (2021).
- (6) Zhang, M. *et al.* Electrocatalytic hydrogen evolution performance of RuO₂ nanorods grown on top of WO₃ nanotube arrays. *Ionics (Kiel)*. **31**, 7205–7216 (2025).
- (7) Zan, W. *et al.* Improved hydrogen evolution reaction performance of molybdenum boride via oxygen-functionalization: A first-principles study. *Int. J. Hydrogen Energy* **112**, 103–110 (2025).
- (8) Putri, L. K. *et al.* Engineering nickel phosphides for electrocatalytic hydrogen evolution: A doping perspective. *Chemical Engineering Journal* **461**, 141845 (2023).
- (9) Jiang, X. *et al.* Self-assembled nanoparticles with bilirubin/JPH203 alleviate imiquimod-induced psoriasis by reducing oxidative stress and suppressing Th17 expansion. *Chemical Engineering Journal* **431**, 133956 (2022).
- (10) Zhang, J. *et al.* Ultra-low Pt-loaded catalyst based on nickel mesh for boosting alkaline water electrolysis. *Appl. Catal. B* **325**, 122296 (2023).
- (11) Adam, A. *et al.* Rationally Dispersed Molybdenum Phosphide on Carbon Nanotubes for the Hydrogen Evolution Reaction. *ACS Sustain. Chem. Eng.* **6**, 11414–11423 (2018).
- (12) Xiao, P. *et al.* Molybdenum phosphide as an efficient electrocatalyst for the hydrogen evolution reaction. *Energy Environ. Sci.* **7**, 2624–2629 (2014).
- (13) Zhao, G. *et al.* Anchoring Pt single atoms on 3D flower-like NiMoP via PtNi bonding with enhanced hydrogen evolution reaction. *J. Colloid Interface Sci.* **688**, 233–240 (2025).
- (14) Deng, C. *et al.* Template-Preparation of Three-Dimensional Molybdenum Phosphide Sponge as High Performance Electrode for Hydrogen Evolution. *J. Journal of Materials Chemistry A*, 1–3 (2013).
- (15) Suliman, M. H., Adam, A., Siddiqui, M. N., Yamani, Z. H. & Qamar, M. The impact of microstructural features of carbon supports on the electrocatalytic hydrogen evolution reaction. *Catal. Sci. Technol.* **9**, 1497–1503 (2019).
- (16) Adam, A. *et al.* Controlled growth of small and uniformly dispersed Mo₂C on carbon nanotubes as high performance electrocatalyst for the hydrogen evolution reaction. *Int. J. Hydrogen Energy* **44**, 11797–11807 (2019).

Hybrid Statistical–Machine Learning Framework for Evaluating Drivers of Power Sector Decarbonization

Maryam Y. Alkhazraj and Mohamed Elsamadony*

Cite <https://doi.org/10.64589/juri/219332>

Submitted: November 26, 2025 Revised: February 22, 2026 Accepted: February 25, 2026

ABSTRACT

Decarbonizing electricity systems is essential for achieving global net-zero targets, yet the relative contributions of socioeconomic and environmental factors remain unclear. In this study, we developed a hybrid framework combining regression analysis and unsupervised principal component analysis (PCA) to assess the complex factors influencing power sector carbon intensity (g CO₂/kWh) across 170 countries. Fifteen indicators from social, environmental, and economic domains were selected and aligned based on their linear relationships with carbon intensity ($\beta \pm \text{SE}$). Globally, the share of renewable electricity ($\beta = -6.68 \pm 0.57$, $p < 0.001$) and low-carbon electricity ($\beta = -6.86 \pm 0.49$, $p < 0.001$) showed the strongest negative associations with carbon intensity. In contrast, primary energy per unit of GDP ($\beta = 87.5 \pm 23.5$, $p = 0.0003$) was linked to higher emissions intensity. PCA indicated that environmental indicators collectively showed the strongest negative correlation with carbon intensity ($r = -0.78$, $p < 0.001$), driven primarily by increases in low-carbon and renewables electricity (%) and decreases in CO₂ per energy unit (kg/kWh) and CO₂ per unit of GDP (kg/\$), underscoring the central role of clean-electricity deployment and CO₂ productivity in decarbonization. Group-level analysis revealed stratification, with high-income economies averaging lower emissions (384 g CO₂/kWh) than lower-middle-income countries (428 g CO₂/kWh). Clustering showed that Europe and North America were low-emission outliers, whereas Central and Southern Asia had high emissions. Network and chord analyses suggest that social progress and economic capacity aid decarbonization but are not direct drivers, underscoring the need to transform environmental systems.

Keywords: data-driven sustainability assessment, principal component analysis, energy system decarbonization, sustainability scoring framework

1. INTRODUCTION

The power sector is a crucial pillar of socioeconomic development and largest source of human-made greenhouse gas emissions, accounting for nearly 40% of global energy-related CO₂ emissions¹. This underscores the importance of meeting the Paris Agreement and net-zero goals by midcentury. Although renewable energy and technological progress have decreased local emissions, global decarbonization is progressing too slowly to achieve the Intergovernmental Panel on Climate Change (IPCC) 1.5 °C target². Continued fossil fuel use (mainly coal) in developing countries, along with increasing demand driven by industry, digitalization, and population growth, complicates efforts to decarbonize.

Understanding the factors affecting the carbon intensity of electricity generation (CI_e), defined as CO₂ emissions per kWh, is therefore essential. Research has often connected recent emissions drops in developed countries to fuel switching, such as transitioning from coal to natural gas, renewables, and nuclear power. For instance, Le Quéré et al.³ observed that renewables and low-carbon electricity sources accounted for nearly two-thirds of the CO₂ reductions in 18 advanced economies from 2010 to 2018, while Sevgi and Figen⁴ identified improvements in generation

efficiency as major contributors to declining CI_e in Organization for Economic Co-operation and Development (OECD) countries. However, most existing studies focus on specific regions or limited time periods, neglecting the broader systemic interactions among energy, economy, and society. Although high-income nations have reduced their carbon intensity through cleaner energy and policy changes, emerging economies still struggle with fossil fuel dependence, rising demand, and limited capacity. This highlights the need for comprehensive data-driven frameworks to evaluate the drivers of power sector carbon intensity across socioeconomic and environmental contexts.

An expanding body of literature links energy decarbonization to sustainability transitions, emphasizing that profound changes in energy, transportation, and industry require technological, institutional, and social transformations⁴. The Multi-Level Perspective interprets these as interactions among niche innovations, sociotechnical regimes, and landscape pressures, indicating that renewable energy develops through systemic interactions involving policies, markets, and social norms. However, applying these insights to quantitative cross-country evaluations remains challenging as most studies tend to focus on individual aspects or region-specific models, which limits their broader relevance. Conventional econometric and sustainability index

methods, such as the Energy Transition Index, Environmental Performance Index, and Energy Trilemma Index, offer benchmarks for national progress but rely on subjective weighting, limited coherence, or narrow coverage^{5,6}. In contrast, multi-criteria decision analysis (MCDA) frameworks (e.g., Analytic Hierarchy Process) improve transparency but depend on expert judgment and are sensitive to normalization⁷. Sustainability indicators are multicollinear and hierarchical, spanning interconnected areas in which linear models struggle to identify causal effects⁸.

Recent research has addressed these challenges by applying machine learning (ML) and dimensionality-reduction methods to sustainability assessments. For example, Chen et al.⁹ developed a principal component analysis (PCA) random forest model to evaluate the global progress toward the Sustainable Development Goals (SDGs). This approach combined unsupervised learning for indicator selection with imputation techniques to manage high-dimensional incomplete data. Similarly, Al Asbahi et al.⁶ demonstrated that PCA provides an objective variance-based weighting system that eliminates biases introduced by arbitrary aggregation in the Energy Trilemma Index. Neofytou et al.⁵ consolidated environmental, social, political, and technological factors into a single readiness index for sustainable energy transitions, demonstrating that the transition capacity of any country relies on complex interactions between institutional stability, economic resilience, and technological readiness. These studies agree that sustainable energy evaluations should integrate data-driven statistical techniques (e.g., regression and PCA) with system-level aggregation approaches (e.g., entropy and MCDA) to capture causal relationships and ensure accurate multidimensional coherence.

Additionally, variations in structural and developmental factors influence decarbonization trajectories. Aghababaei et al.¹⁰ analyzed 118 countries and observed that income level alone does not account for disparities in renewable energy adoption. Some lower-income nations achieved high renewable energy shares through targeted investments, whereas others still rely heavily on carbon-intensive methods, despite economic growth. Similarly, Tao et al.¹¹ and Wang & Jia¹² identified energy intensity, industrial clustering, and coal-based energy structures as factors that contribute to higher regional emissions through variations in labor, capital, and policy. These findings highlight the need for integrated frameworks that combine environmental, economic, and social indicators to understand the structural drivers of power sector emissions.

To address these challenges, this study developed a hybrid approach that combines statistical techniques with ML. It employs supervised regression and PCA to analyze indicators related to power-sector carbon intensity, grouping correlated indicators into social, environmental, and economic dimensions across 170 countries. This provides an empirical, transparent, and reproducible method for benchmarking decarbonization readiness, guiding sustainable energy policies, and supporting global sustainability efforts. Unlike conventional sustainability indices that construct composite scores by aggregating indicators using fixed-weighting frameworks, this approach introduces a data-driven framework that explicitly infers indicator directionality prior to dimensionality reduction using regression-based associations with CIE. This enhances interpretability by avoiding

arbitrary sign assignments and allows the resulting PCA-based pillars to be used as analytical tools for identifying structural decarbonization pathways, rather than as rank-oriented sustainability scores. The study also examines how the drivers of carbon intensity vary across World Bank income groups, capturing differences in economic development and decarbonization capacity.

2. METHODOLOGY

This study used a structured workflow to identify social, environmental, and economic factors influencing CIE across 170 countries (Figure 1). The framework combines data screening, normalization, regression alignment, and PCA to create three sustainability pillars. Table 1 lists the choice of each indicator based on its empirical relationship with the CIE.

2.1. Data Sources and Indicator Framework. Data at the national level for 170 countries were collected from 2022 to 2024 using *Our World in Data*. From this dataset, we selected 15 indicators across three sustainability dimensions.

- Social indicators (Clean cooking, Human Development Index [HDI], Access to electricity, Life expectancy) capture aspects of human development and inclusivity.
- Environmental indicators (PM2.5, air pollution, CO₂ emissions, low-carbon electricity, CO₂ per energy unit, renewable share of electricity [%], CO₂ embedded in trade, CO₂ per unit of GDP, and per capita GHG emissions) highlight electricity system characteristics, air quality, and carbon efficiency.
- Economic indicators (government expenditure, primary energy per capita, and primary energy per unit of GDP) represent the energy consumption, fiscal capacity, and economic efficiency, respectively.

These indicators were selected based on three criteria: accessibility, relevance to understanding sustainability transitions, and established associations with CIE in the literature (Table 1). All indicators are treated as cross-sectional observations drawn from the most recently available years within the period of 2022–2024. Accordingly, this analysis focused on cross-sectional associations rather than temporal dynamics. Potential indicators, such as electricity prices, carbon pricing instruments, and grid stability metrics, were not included because of limited global data availability, inconsistent cross-country comparability, or potential collinearity with the selected variables. This framework focuses on indicators with consistent global coverage and a clear interpretive relevance to CIE.

2.2. Data Processing and Directional Normalization.

To ensure comparability, each indicator “ X_j ” was then normalized to a 0–100 scale using min–max transformation, as shown in Eq. 1.

$$X_j^{(0-100)} = \frac{X_j - \min(X_j)}{\max(X_j) - \min(X_j)} \times 100 \quad (1)$$

This rescaling enabled the evaluation of indicators with different units, such as emissions, percentages, and economic intensities, using a standard metric⁷.

Table 1. Rationale for social, environmental, and economic indicators based on their directional relationships with the carbon intensity of electricity

Pillar	Indicator	Justification Based on Association With the Carbon Intensity of Electricity (CIE)
Economic	Primary energy per unit of GDP, kWh/\$	High energy consumption per unit of GDP indicates an inefficient economy reliant on fossil fuels. Countries with higher energy intensity burn more coal and oil, increasing CIE. Research shows energy intensity drives emissions in developing and fossil-fuel economies ²⁶ . Deichmann et al. ²⁷ found a negative correlation between GDP per capita and energy intensity, suggesting that higher GDP per capita generally correlates with better energy efficiency and less fossil fuel dependence.
	Government expenditure, % of GDP	Significant public investment in subsidies, grid improvements, and energy-efficiency initiatives underpins the shift to clean energy. Data indicate that policy-led adoption of renewable energy is crucial for decarbonization, mainly in high-income nations ²⁸ . Khan et al. ²⁰ found that higher public energy R&D budgets, higher incomes, and improved political frameworks accelerate clean energy transitions. Caglar et al. ²¹ demonstrated that national budgets dedicated to energy efficiency R&D promote policy diffusion and emission reductions, highlighting the importance of targeted public spending.
	Primary energy per capita, kWh/person	Higher energy use per person increases electricity demand, supplied mainly by fossil fuels, leading to higher CIE. Research indicates that fossil fuel consumption significantly raises CO ₂ emissions unless the grid is renewable. A 1% increase in per capita energy consumption results in approximately a 0.946% rise in CO ₂ emissions, reflecting many countries' reliance on fossil fuels to meet higher electricity demand ²⁹ . Chontanawat et al. ³⁰ confirmed a two-way causal relationship: greater fossil fuel use increases CO ₂ emissions unless electricity generation is renewable or low carbon.
Environmental	Low-carbon electricity, %	Electricity generated from solar, wind, hydro, or nuclear sources replaces coal and gas, effectively lowering emissions, especially in OECD countries. A 2024 study in Decarbonizing Energy found that a 1% increase in renewables results in a 0.87% decrease in fossil fuel generation, making this the most effective method for reducing emissions ³¹ . Rahman et al. ³² showed that deploying renewable energy significantly cuts CO ₂ emissions in countries heavily reliant on fossil fuels.
	Renewables share of electricity (%)	A recent multi-country analysis shows that increasing renewables strongly reduces fossil fuel dependence. Using data from 36 OECD countries (2000–2020), it finds that a 1% rise in renewables correlates with a 0.87% decrease in fossil fuel capacity, driving emission reductions in developed countries ³¹ .
	CO ₂ per unit of GDP, kg/\$	A strong long-term relationship exists between per capita CO ₂ emissions and the proportion of manufacturing value added to GDP. Higher CO ₂ emissions per unit of GDP are associated with large, energy-intensive manufacturing sectors that rely heavily on coal and other fossil fuels, thereby increasing carbon intensity and emissions. In addition, another study discovered a moderate-to-high positive correlation among CO ₂ emissions, manufacturing output, gross value added, and fossil fuel consumption, indicating that heavy industry and fossil fuels are major contributors to high emissions. Yang et al. ¹⁶ emphasized that manufacturing growth significantly increases CO ₂ emissions in developing, carbon-intensive economies, especially where electricity generation depends mainly on coal and oil.
	CO ₂ per energy unit, kg/kWh	Research shows that CO ₂ emissions per kilowatt-hour (kg CO ₂ /kWh) measure a country's energy carbon intensity. Coal systems emit around 800–1000 g CO ₂ /kWh due to high carbon content and low efficiency. Renewables and nuclear have much lower emissions, often below 50 g CO ₂ /kWh. This emissions intensity impacts the global carbon footprint of electricity, as confirmed by life-cycle assessments and international data ³³ .
	PM2.5 air pollution	PM2.5 pollution mainly results from burning coal, with research confirming coal-fired power plants as key sources. In China, coal makes up about 40–50% of PM2.5 in polluted areas, and global studies emphasize its central role in particulate emissions. The connection between PM2.5 and fossil-fuel electricity makes PM2.5 an indicator of coal power ¹⁷ .
	Per capita GHG emissions, tCO ₂ e/person	Countries with high per-capita GHG emissions often rely on fossil fuels for electricity, transportation, and industry, leading to higher CO ₂ emissions per kWh. Global studies show that fossil fuel consumption is correlated to both per capita emissions and the carbon intensity of electricity ^{18,19} .

Continued.

Table 1. *Continued*

Pillar	Indicator	Justification Based on Association With the Carbon Intensity of Electricity (CIE)
	CO ₂ embedded in trade	Many studies reveal that countries with cleaner electricity grids tend to export more embedded CO ₂ by shifting carbon-intensive manufacturing to nations with dirtier energy sources. Foong et al. ¹⁵ showed that import-dependent economies (e.g., Singapore, Hong Kong, the UAE, and Bahrain) have trade-adjusted emissions far higher than their production emissions, mainly due to energy-intensive imports from coal-reliant regions. Zhou et al. ¹³ observed that developed countries maintain low domestic carbon intensity while importing embodied carbon, effectively externalizing emissions.
	Annual CO ₂ emissions	High national CO ₂ emissions typically indicate a power sector reliant on fossil fuels, mainly coal and natural gas, which increases the CIE. This pattern is common globally among major emitters, where about 75% of power-related CO ₂ emissions come from fossil fuels. This link motivates efforts to shift to cleaner energy sources and reduce overall emissions ²⁵ .
Social	Life expectancy at birth, years	Higher life expectancy is associated with better air quality, stronger institutions, and more efficient energy regulation. Cutting PM _{2.5} by 10 µg/m ³ can extend lifespan by 0.35 years ³⁴ . Developed countries often use cleaner energy, lowering pollution and improving health. Global research confirms the link between cleaner energy, less pollution, and longer lives ³⁵ .
	Human Development Index, HDI	Countries with high HDI usually allocate more resources to renewable energy, enforce stricter pollution regulations, and are better prepared for decarbonization efforts. Studies show a positive connection between renewable energy use and HDI, indicating that cleaner energy helps human development. Econometric analyses reveal that these countries can separate electricity demand from its CIE, enabling economic growth while reducing emissions ³⁶ .
	Access to electricity, %	In low-income regions, expanding electricity access typically relies on fossil fuels, mainly coal, leading to a higher CIE. This "access-emissions" trade-off is common in emerging economies, where coal provides new power but also raises emissions. Studies highlight the importance of investing in renewable energy and grid improvements to overcome this trade-off and promote sustainable growth ²³ .
	Clean cooking, %	Clean cooking reduces health and environmental harm by replacing biomass or coal with cleaner fuels such as LPG or electricity. However, in areas powered by fossil fuels, higher demand can increase the CIE. Literature emphasizes this trade-off, noting that clean cooking must be combined with grid decarbonization to avoid increasing emissions. Studies indicate that better stoves and fuels reduce household emissions, but the climate advantages are most significant when paired with renewable energy sources ²⁴ .

2.3. Regression-Based Direction Inference and Correlation Analysis. To ensure empirical alignment between each indicator and the decarbonization target, ordinary least-squares regression was used to estimate the linear relationship between every indicator X_j and target Y (carbon intensity), as illustrated in Eq. 2.

$$Y = \beta_0 + \beta_1 X_j + \varepsilon \quad (2)$$

All indicators were normalized prior to regression to ensure comparability across scales. As regression was used solely to determine the direction of association rather than the effect magnitude, robustness was evaluated based on the consistency of the coefficient signs. The diagnostic checks did not alter the directional interpretation of the results. We report β_1 , its standard error (SE), p-value, sample size, and Pearson r . A negative coefficient ($\beta < 0$) indicates that higher indicator values are beneficial, associated with lower CIE. Conversely, a positive coefficient ($\beta > 0$) indicates an adverse association.

These supervised results (Table 1) provide an empirical orientation before unsupervised learning, avoiding arbitrary sign conventions and weights. Pairwise Pearson correlations among the indicators supported the network topology (Figure 3) and

income-group chord analysis (Figure 4). The univariate Ordinary Least Squares (OLS) regressions were not intended to estimate causal effects or relative magnitudes, but rather to examine the directional association between each indicator and carbon intensity of electricity generation. This step is used solely to determine whether indicators align positively or negatively with decarbonization outcomes prior to the PCA, thereby avoiding an arbitrary sign assignment. While regression coefficients may be sensitive to scale, variance, and omitted variables, these limitations are addressed by restricting the use of OLS to direction inference only and by subsequently relying on PCA to capture the multivariate structure and shared variance among correlated indicators. Alternative approaches, such as theoretical sign assignment or multivariate regression models, have been considered; however, theoretical assignment introduces subjective assumptions, and multivariate models are less suitable for highly collinear indicator sets. Some indicators showed insignificant univariate associations with carbon intensity. These indicators were retained because the objective of the framework is to capture broader social and economic contexts that interact with environmental factors in the multivariate PCA.

2.4. PCA and Composite Scoring. After directionally aligning and standardizing the data, PCA was used within each pillar to condense the multiple correlated indicators into a single underlying component. PCA is widely recommended for sustainability assessments because it identifies significant structures without subjective weighting⁶. The first principal component (PC₁) in each pillar explained 86.1% of the social variance, 35.0% of the environmental variance, and 54.0% of the economic variance, indicating strong internal coherence. Although the explained variance of Environmental PC₁ was lower than that of the Social and Economic pillars, it still captured the dominant shared structure among environmental indicators and provided a sufficiently representative composite dimension for cross-country comparison. Although several environmental indicators are conceptually related, they reflect the complementary dimensions of carbon performance, including both emission scale and structural efficiency, allowing the environmental pillar to capture broader system transformations. PC₁ was retained because it explained the largest share of the variance and captured the dominant pattern relevant to the analysis. Although PCA scores are used to summarize the dominant patterns within each pillar, they do not represent normative or exhaustive measures of sustainability performance. Instead, the PC₁ scores reflect the primary variance structure among the selected indicators and are interpreted as relative composite patterns rather than absolute sustainability rankings. This distinction highlights the analytical scope and limitations of the proposed PCA framework.

This study examined several interconnected environmental indicators because they reflect different facets of carbon pressure, including total emissions, emissions per capita, economic efficiency of carbon (CO₂ per unit of GDP), and carbon intensity of the energy system (CO₂ per energy unit). Rather than combining these indicators directly, PCA was used to address multicollinearity by transforming correlated variables into orthogonal components. Supplementary Table S1 shows that the first environmental principal component (PC₁) exhibits distributed loadings across indicators (e.g., +0.503 for low-carbon electricity, +0.391 for CO₂ per energy unit, -0.421 for CO₂ per unit of GDP, +0.223 for per-capita GHG emissions, and +0.181 for annual CO₂ emissions), indicating that no single metric dominates the composite representation. PC₁ accounted for 35.0% of environmental variance (Supplementary Table S2). Conversely, the subsequent components explained progressively smaller proportions, supporting the use of PC₁ as a parsimonious and robust summary of environmental system characteristics. The PC₁ scores were then rescaled to a 0–100 pillar score for comparability across countries. The composite pillar scores were then computed from principal component scores as mention in Eq. 3.

$$Score_1 = 100 \times \frac{PC1_i - \min(PC1)}{\max(PC1) - \min(PC1)} \quad (3)$$

The indicator loadings offered insights into the most significant factors within each domain. For instance, within the environmental pillar, low-carbon electricity (%) and renewable % electricity loaded strongly and positively, while carbon-efficiency indicators (CO₂ per unit of GDP and CO₂ per energy unit) exhibited contrasting loadings, reflecting the different dimensions of carbon productivity and energy-system structure^{3,13}.

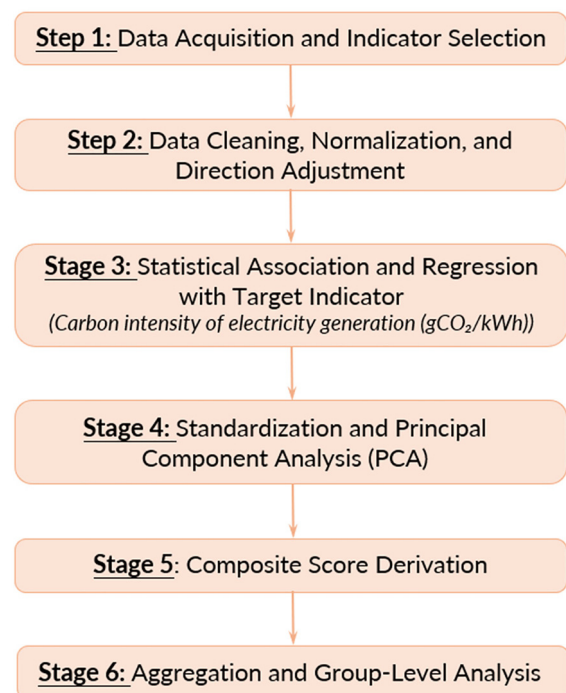


Figure 1. Workflow diagram illustrating the stages to derive indicators, perform principal component analysis, and compute scores, showing how metrics are aggregated into three sustainability pillars

Finally, country-level pillar scores were combined with World Bank income groups and UN-SDG regions (Figures 2, 5, and 6) to reveal global sustainability patterns, consistent with earlier comparative studies on transition readiness and structural constraints^{10,12}. World Bank income classifications were used as a comparative framework to examine structural differences in decarbonization patterns across countries at different stages of economic development.

As the analysis relies on cross-sectional observations from the most recently available year, the identified associations reflect current structural patterns rather than dynamic transition pathways over time.

3. RESULTS AND DISCUSSION

3.1. Global Carbon Intensity and Income Disparities.

The non-linear relationship between national income level and CI_e is clearly depicted in Figure 2. High-income countries (HICs) have a relatively low average intensity of 384.5 gCO₂/kWh, indicating mature energy systems that widely employ low-carbon and renewable technologies. Conversely, upper-middle-income countries (UMICs) exhibit the highest average intensity at 487.6 gCO₂/kWh and most considerable variation, with several nations surpassing 1,200 gCO₂/kWh. This heavy upper tail is typical of electricity systems still largely reliant on coal-fired generation and experiencing rapid growth in electricity demand, factors noted in many global assessments of emerging-economy energy transitions, where industrial expansion and coal dependence collectively increase the CI_e^{13,14}

Lower-middle-income countries (LMICs) have an average carbon intensity of 427.6 gCO₂/kWh, but this varies widely

due to diverse energy mixes, from hydro-rich systems to coal-dependent grids. Low-income countries (LICs) had an average of 376.9 gCO₂/kWh, similar to that of HICs. However, this similarity does not imply advanced decarbonization; many LICs have lower per capita electricity use, rely on hydropower and biomass, and lack large fossil fuel power plants. This pattern, in which low consumption masks structural carbon intensity, is typical of electrification and development studies^{12,15}. These patterns demonstrate that income alone does not determine carbon intensity. Instead, the CI_e is influenced by factors such as fuel mix, technology, efficiency, and demand growth. HICs have transitioned from coal with supportive policies, whereas middle-income countries (LMICs and UMICs) are experiencing increased demand and dependence on cheap coal, resulting in higher and more variable CI_e. The global differences in power sector emissions are primarily driven by fuel type, transition away from coal, adoption of renewables, and infrastructure development, rather than GDP alone^{3,10}.

3.2. Indicator Relationships with Carbon Intensity.

The regression and correlation analyses (Table 2), along with the inter-indicator network shown in Figures 3 and 4, demonstrate that environmental system indicators are the strongest correlates of global variation in CI_e. Conversely, economic and social indicators mainly serve as enabling conditions, rather than directly affecting emissions.

Among the environmental system indicators, the shares of low-carbon electricity and renewables share of electricity show the strongest significant negative correlations with CI_e ($\beta = -6.86 \pm 0.49$ and $\beta = -6.68 \pm 0.57$, $p < 0.001$), highlighting the importance of expanding clean electricity sources in decarbonizing power systems. Similar findings have been reported in China and other major emitting countries, which shows that replacing coal and gas with renewable, non-fossil electricity is the most effective mitigation strategy¹. They also align with studies in developed countries, where emission reductions are mainly driven by increased renewable energy adoption, structural changes in generation portfolios, and efficiency improvements³.

Environmental indicators that assess the carbon efficiency of energy and economic systems, such as CO₂ per energy unit (kg/kWh; $\beta = 1104.71 \pm 193.87$) and CO₂ per unit of GDP

(kg/\$; $\beta = 634.21 \pm 97.62$), show strong positive correlations with CI_e. These relationships are consistent with higher CI_e levels in countries heavily dependent on coal, with high industrial emissions, or with carbon-intensive supply chains. Previous studies have indicated that industrialization that relies on fossil fuels, especially in rapidly developing economies, significantly increases CO₂ emissions from electricity production^{12,16}. Similar patterns are expected in India, where the rising electricity demand and coal-based expansion may lead to substantial future increases in emissions¹⁴.

Considering the cross-sectional design, the regression and PCA results identified associations with CI_e rather than causal relationships. Causal inference requires longitudinal or instrumental variable methods, which are beyond the scope of this study. Network topology and chord diagrams were employed as complementary visualization tools to summarize the correlation structures among the indicators. The edges represent pairwise Pearson correlations exceeding a predefined threshold, with the edge thickness proportional to the correlation strength. Color scales were used to distinguish the correlation signs and magnitudes. These visualizations are not intended to replace correlation matrices but rather to provide an intuitive overview of clustering patterns and connectivity relationships that are less readily discernible in tabular form.

Note: Regression coefficients reported as 0.00 reflect rounding after standardization and do not indicate null associations. The correlation coefficients were independently reported.

The strong positive correlation between PM_{2.5} pollution ($\beta = 3.88 \pm 1.10$) and CI_e highlights its link with fossil fuels. Globally, coal burning is a key source of fine particulate pollution; in China, coal contributes approximately 40–50% of PM_{2.5}, in polluted regions¹⁷. As coal burning releases large amounts of both CO₂ and PM_{2.5}, these pollutants often increase together in areas with high-carbon electricity systems. Additionally, per capita greenhouse gas emissions are positively correlated with CI_e, indicating that countries that are still heavily reliant on fossil fuels for electricity, transportation, and industry tend to have higher emissions^{18,19}.

In contrast, government expenditure and HDI exhibit weak or insignificant direct effects on CI_e. This does not imply that social

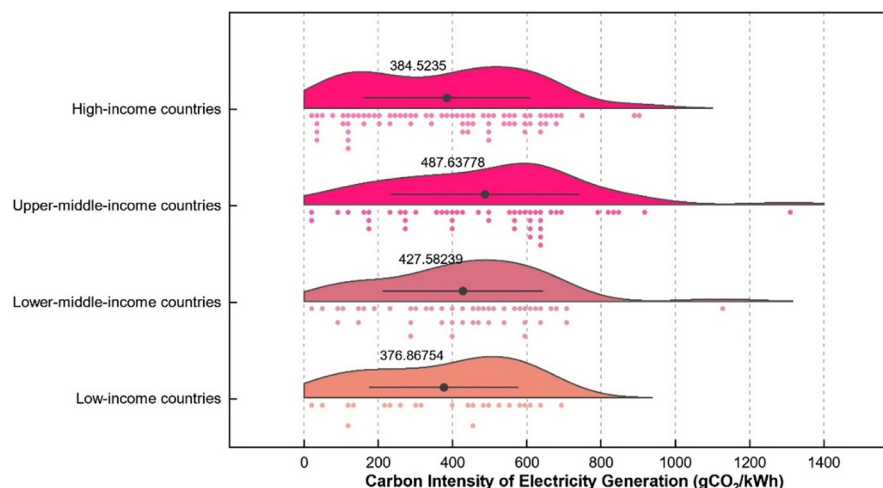


Figure 2. Global distribution of power-sector carbon intensity across World Bank income groups

Table 2. Global correlation matrix between CIE and composite indicators

Pillar	Indicator	β	SE	p_value	Pearson_r	Direction/sign
Economic	Primary energy per unit of GDP, kWh/\$	87.51	23.5185	0.0003	0.26	Adverse “-1” ($\beta > 0$)
	Government expenditure, % of GDP	-2.07	1.3201	0.1184	-0.11	Beneficial “1” ($\beta < 0$)
	Primary energy per capita, kWh/person	0.00	0.0005	0.2204	0.09	Adverse “-1” ($\beta > 0$)
Environmental	Low-carbon electricity, %	-6.86	0.4884	0.0000	-0.71	Beneficial “1” ($\beta < 0$)
	Renewables share of electricity %	-6.68	0.5688	0.0000	-0.65	Beneficial “1” ($\beta < 0$)
	CO ₂ per unit of GDP, kg/\$	634.21	97.6152	0.0000	0.42	Adverse “-1” ($\beta > 0$)
	CO ₂ per energy unit, kg/kWh	1104.71	193.8696	0.0000	0.38	Adverse “-1” ($\beta > 0$)
	PM2.5 air pollution	3.88	1.0968	0.0005	0.25	Adverse “-1” ($\beta > 0$)
	Per capita GHG emissions, t CO ₂ e/person	6.92	2.1010	0.0012	0.23	Adverse “-1” ($\beta > 0$)
	CO ₂ embedded in trade	0.00	0.0000	0.1408	-0.11	Beneficial “1” ($\beta < 0$)
	Annual CO ₂ emissions	0.00	0.0000	0.2583	0.08	Adverse “-1” ($\beta > 0$)
	Life expectancy at birth, years	-4.09	2.3565	0.0841	-0.12	Beneficial “1” ($\beta < 0$)
	Human Development Index, HDI	-109.11	111.4776	0.3289	-0.07	Beneficial “1” ($\beta < 0$)
Social	Access to electricity, %	0.68	0.7530	0.3665	0.07	Adverse “-1” ($\beta > 0$)
	Clean cooking, %	0.09	0.4957	0.8492	0.01	Adverse “-1” ($\beta > 0$)

and governance improvements are unimportant. Instead, it suggests that their impact is indirect, supporting the transition by enhancing institutional capacity, fostering innovation, and promoting policy adoption. Research on socio-technical transitions indicates that governance and social development shape a country's preparedness to adopt low-carbon technologies. However,

actual emission reductions occur only when these structural factors are combined with changes in the energy system^{4,5}. For instance, increasing public investment in energy R&D does not directly reduce emissions but creates an environment conducive to the adoption of energy efficiency measures and renewable technologies^{20,21}.

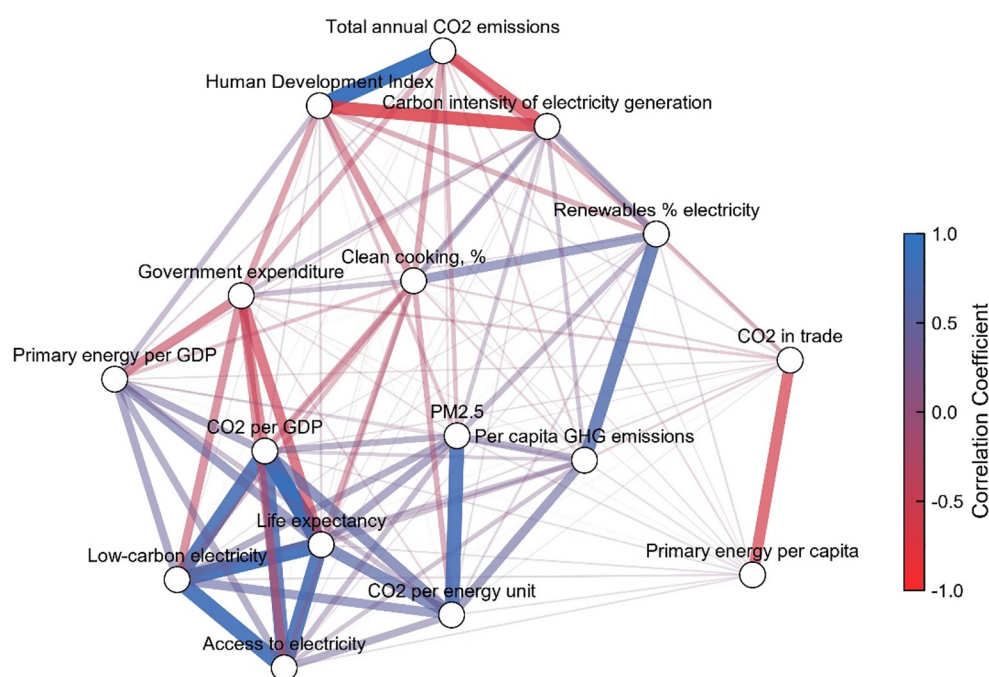


Figure 3. Network topology of inter-indicator relationships in the global electricity system, with blue and red edges representing positive and negative correlations, respectively

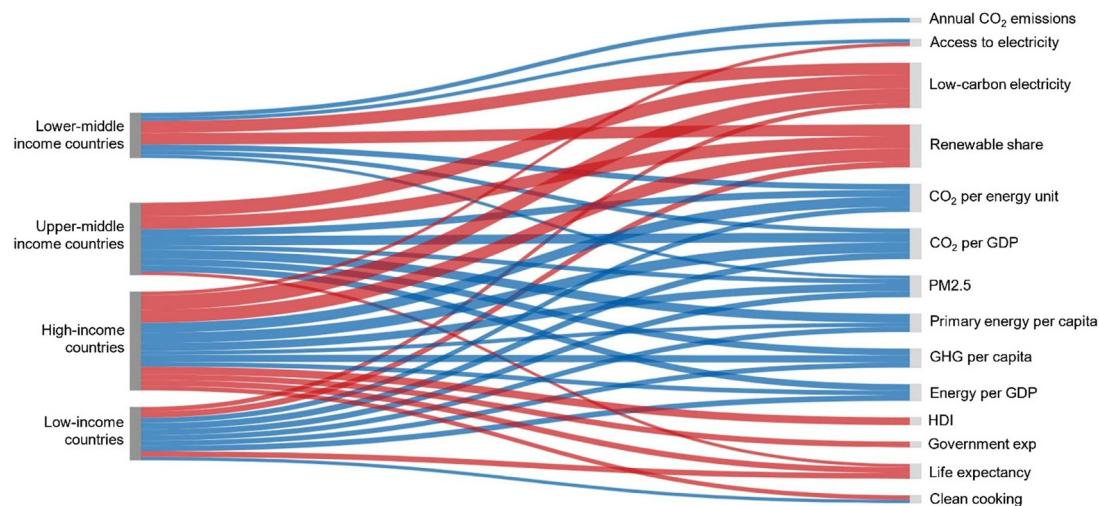


Figure 4. Cross-indicator correlation of carbon-intensity patterns by income category

From a policy perspective, decoupling electricity access from rising emissions in LICs requires avoiding a long-term reliance on coal during electrification. Expanding access through fossil fuel-based generation can meet short-term demand, but risks locking power systems into carbon-intensive pathways. Prioritizing renewable-based electrification, supported by appropriate policies and financing mechanisms, enables access for expansion while limiting increases in carbon intensity and maintaining alignment with long-term climate and sustainable development objectives²².

Similarly, life expectancy, access to electricity, and clean cooking show weak correlations with CI_e . Improvements in these social factors typically occur before decarbonization and are vital for welfare, but do not immediately reduce electricity sector emissions unless coupled with structural reforms in electricity generation. In LICs, expanding electricity access often relies initially on fossil-fuel-based power, leading to a temporary “access–emissions” trade-off until renewable capacity is established^{23,24}.

The network shown in Figure 3 reveals clear patterns: indicators for clean electricity and carbon efficiency cluster together and show strong positive connections with low-carbon electricity solutions. In contrast, reliance on fossil fuels, industrial emissions, PM2.5, and energy intensity form a separate cluster associated with higher carbon levels. This division illustrates two different socio-technical regimes, as described by the Multi-Level Perspective⁴: one characterized by mature, low-carbon electricity systems, and another shaped by coal dependence and inefficient energy use, which often become entrenched in high-carbon pathways.

Figure 4 illustrates how the indicator relationships vary by income group. HICs tend to cluster through low-carbon electricity pathways, reflecting earlier energy diversification and stronger policies. Middle-income countries are associated with carbon-intensive trajectories owing to rapid industrial growth and coal reliance^{11,12}. LICs, although seemingly less carbon-intensive, are characterized by low electrification and limited industries, suggesting suppressed emissions rather than active decarbonization. Our findings show that environmental indicators, such as clean electricity and carbon efficiency, are essential for decarbonizing

the power sector. Social and economic factors support the transition; however, they do not lower the CI_e .

3.3. Regional and Income-Level Sustainability Patterns.

The distribution of social, environmental, and economic PCA scores (Figure 5) shows clear differences in sustainability and progress toward decarbonization among income groups. HICs consistently scored the highest across all three pillars (87.4, 84.2, and 79.3, respectively), underscoring their continued strengths in technology, institutions, and resources. These patterns match earlier research, indicating that developed economies are more advanced in renewable energy use and energy efficiency, which have contributed to their historic reductions in CO₂ emissions³. In contrast, LICs tend to have lower scores on all pillars, averaging below 40, emphasizing ongoing inequalities in access to modern energy, human development, and economic capacity, which are key factors in transition readiness⁵. UMICs and LMICs occupy intermediate positions but show more variation, especially in their environmental (UMIC = 65.8; LMIC = 53.4) and economic (UMIC = 62.7; LMIC = 51.9) scores. This variation reflects the diversity of their energy paths: while some UMICs (e.g., Brazil and South Africa) are expanding renewables, others (e.g., China, India, and Indonesia) remain heavily reliant on coal. This transitional diversity is well-documented in studies showing that middle-income countries face the dual challenge of rapidly increasing electricity demand and fossil-fuel dependence^{12,14}. These structural pressures limit their ability to follow the emissions-reduction path observed in advanced economies, a challenge identified by van Soest et al.² as a significant barrier to achieving Paris-compatible net-zero trajectories.

The environmental pillar exhibited the clearest developmental gradient. HICs combine strong environmental PCA scores with some of the lowest observed CI_e values (~384 g CO₂/kWh). These results reflect mature decarbonization strategies, including the widespread use of hydro, wind, nuclear, and solar energy; strict air pollution controls; and effective coal phaseout policies^{1,2}. By contrast, LMICs and LICs exhibit significant environmental deficits and elevated carbon intensity, indicating systems

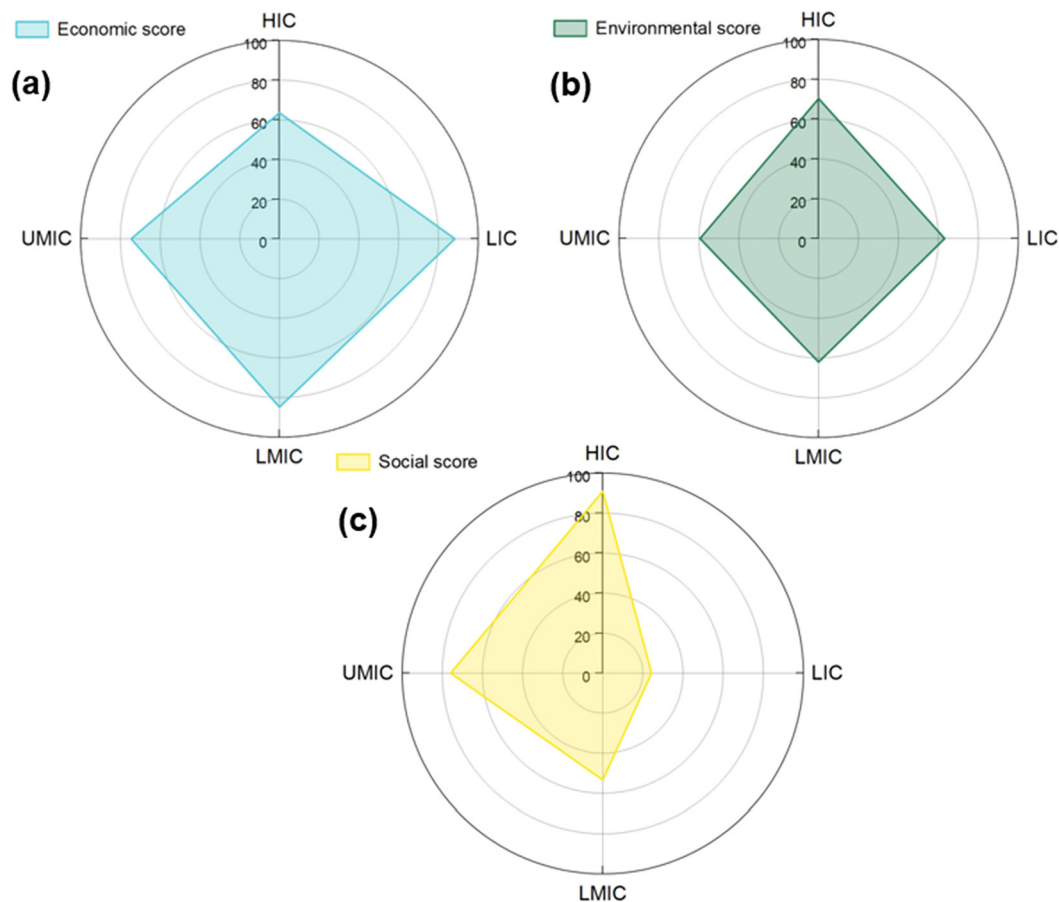


Figure 5. Comparative distribution of (a) economic, (b) environmental, and (c) social PCA scores classified based on income level (i.e., high-income (HIC); upper-middle-income (UMIC); lower-middle-income (LMIC); and low-income countries (LIC))

reliant on coal, diesel, or biomass for electricity with low efficiency. In the literature, this pattern is referred to as the “carbon development gap,” a period in which economic growth and industrialization increase energy demand faster than the adoption of clean technologies¹¹.

Figure 6 further illustrates these patterns using hierarchical regional clustering. Europe and North America form a distinct low-carbon electricity cluster marked by high Environmental PCA scores (>80) and a low CI_e (<400 g CO₂/kWh). These regions demonstrate the typical pathway outlined by Le Quéré et al.³, where emissions decline primarily through structural improvements in energy carbon intensity, fuel switching, and stronger policy measures. Australia and New Zealand cluster closely within this group, supported by a similar reliance on large-scale hydropower, wind, and gas, along with effective governance.

Intermediate clusters, including Oceania, Latin America, and the Caribbean, show strong renewable energy integration but have moderate social and economic scores. This pattern matches recent empirical analyses, indicating that regions rich in renewable energy can achieve significant emissions reductions even if their institutional capacity or economic diversification is limited¹⁰. Central and Southern Asia and Sub-Saharan Africa form the lowest-performing clusters, with environmental scores below 45 and CI_e values over 650 g CO₂/kWh. These regions face energy poverty, underinvestment in grid infrastructure, and

reliance on low-efficiency thermal generation, which hinders decarbonization^{13,25}.

These income and regional disparities show that the key to decarbonization is the transformation of environmental systems, and not just income levels. While social and economic development creates favorable conditions, such as better governance, innovation, and social readiness, reducing power sector emissions requires combining these with structural changes in the electricity mix, lowering the CO₂ intensity of energy sources, and ongoing policy efforts^{2,4}. This multilayered evidence shows that decarbonization is more than just an economic process; it is a sociotechnical transition driven by interactions among institutions, technology, investment, and infrastructure. Therefore, the global pattern of carbon intensity of electricity exposes both ongoing inequalities and differing capacities for environmental system change, posing the future challenge of aligning power systems with the pathways needed to meet the Paris Agreement’s goals.

The PCA-based scores reflect the current structural conditions captured by the 2022–2024 snapshot rather than the dynamic transition trajectories. Countries with rapidly improving indicators may, therefore, follow different decarbonization pathways over time than those implied by their present scores. These findings may also support policymakers in identifying priority areas for electricity sector decarbonization and tracking structural progress across income contexts.

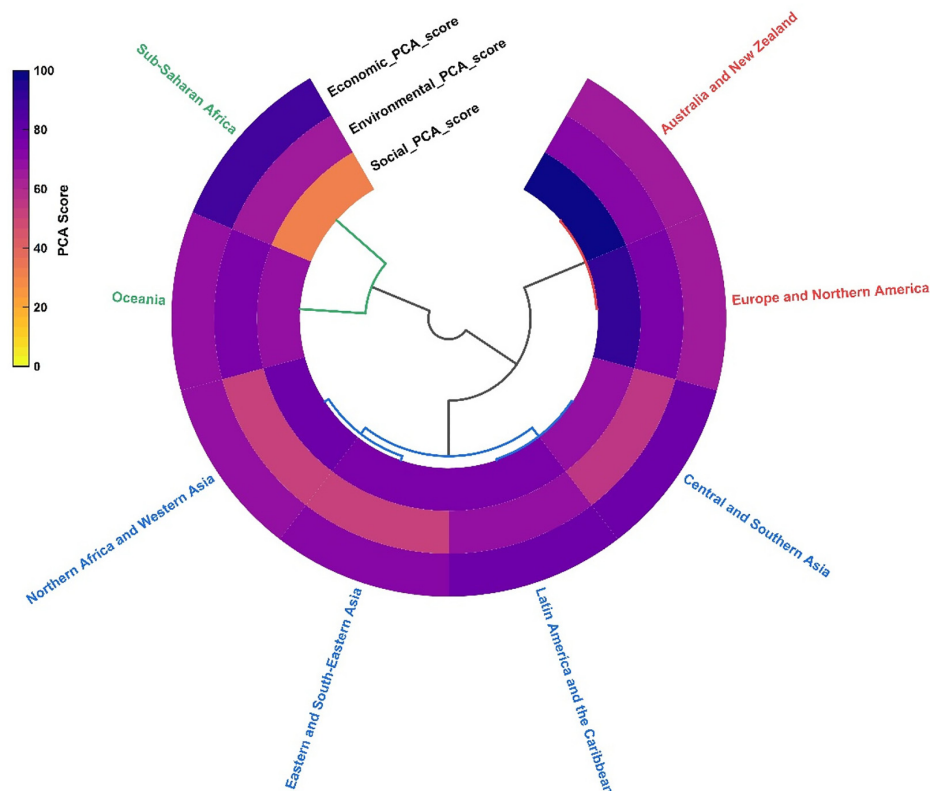


Figure 6. Hierarchical clustering of regions by carbon-intensity performance and sustainability-pillar scores

4. CONCLUSIONS

This study introduces a hybrid framework that combines regression-based direction inference and PCA to identify the factors influencing power sector decarbonization across 170 countries. Normalizing and consolidating 15 social, environmental, and economic indicators into composite sustainability pillars provides a transparent, data-driven approach for evaluating the global CI_e . The results identify the environmental pillar as the key driver of CI_e , with the shares of low-carbon and renewable electricity showing the most substantial negative correlations. In contrast, carbon inefficiency metrics, such as CO_2 per energy unit and CO_2 per unit of GDP, show positive associations. Social and economic factors such as HDI and electricity access directly influence transition readiness rather than emissions.

The analysis also reveals structural disparities across income groups: HICs tend toward low-carbon electricity paths, whereas middle-income countries face the challenges of demand growth and coal dependency. Therefore, reducing emissions in the electricity sector necessitates environmental transformation, the rapid expansion of renewables, and improved carbon productivity. This framework provides a scalable tool for diagnosing decarbonization pathways and supporting equitable parallel-aligned energy policies.

The policy implications differ across income groups. In HICs, decarbonization pathways are primarily associated with accelerating renewable energy deployment, grid optimization, and efficiency improvements. In LICs, LMICs, and UMICs, the framework highlights the importance of enabling conditions, such as access to low-carbon electricity, structural

diversification of generation portfolios, and institutional capacity. The proposed framework can support policymakers by identifying dominant indicator patterns relevant to their development contexts rather than prescribing uniform policy targets.

As this study relied on cross-sectional data from the most recently available year, the findings reflect current structural associations rather than dynamic transition pathways. Structural decarbonization processes evolve, and future longitudinal analyses may reveal changing relationships across countries over time. Future research should further extend this framework using time-series data or scenario-based modeling to capture the evolving decarbonization trajectories.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Maryam Y. Alkhazraj – Mechanical Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; 0009-0003-8475-9865
Email: s202157030@kfupm.edu.sa

Corresponding Author

Mohamed Elsamadony – Research Mentor, Assistant Professor, Mechanical Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; 0000-0001-9431-1157
Email: mohamed.elsamadony@kfupm.edu.sa

ACKNOWLEDGEMENTS

The authors acknowledge the support provided by the Mechanical Engineering Department at King Fahd University of Petroleum & Minerals and thank the JURI editorial team for their guidance and support.

REFERENCES

- (1) Dai, P., Chen, G., Zhou, H., Su, M. & Bao, H. CO₂ mitigation measures of power sector and its integrated optimization in China. *Sci. World J.* **2012**, 907685 (2012).
- (2) van Soest, H. L., den Elzen, M. G. J. & van Vuuren, D. P. Net-zero emission targets for major emitting countries consistent with the Paris Agreement. *Nat. Commun.* **12**, 1–9 (2021).
- (3) Le Quéré, C. *et al.* Drivers of declining CO₂ emissions in 18 developed economies. *Nat. Clim. Chang.* **9**, 213–217 (2019).
- (4) Geels, F. W. Socio-technical transitions to sustainability: a review of criticisms and elaborations of the Multi-Level Perspective. *Curr. Opin. Environ. Sustain.* **39**, 187–201 (2019).
- (5) Neofytou, H., Nikas, A. & Doukas, H. Sustainable energy transition readiness: A multicriteria assessment index. *Renew. Sustain. Energy Rev.* **131**, 109988 (2020).
- (6) Asbahi, A. A. M. H. Al *et al.* Novel approach of Principal Component Analysis method to assess the national energy performance via Energy Trilemma Index. *Energy Reports* **5**, 704–713 (2019).
- (7) Pollesch, N. L. & Dale, V. H. Normalization in sustainability assessment: Methods and implications. *Ecol. Econ.* **130**, 195–208 (2016).
- (8) Xavier, A., Fragoso, R. & Freitas, M. de B. C. Building sustainability composite indicators using a multi-criteria approach. *Eur. J. Oper. Res.* **326**, 326–342 (2025).
- (9) Chen, L., Shuai, C., Chen, X. & Zhao, B. A data-driven framework for assessing global progress towards sustainable development goals. *Sustain. Prod. Consum.* **60**, 217–228 (2025).
- (10) Karbalaei Aghababaei, M., Saifoddin, A., Zahedi, A. & Abdoos, M. Macroeconomic impact of energy transition: A comparative study of developed and developing countries. *Energy Strateg. Rev.* **62**, 101910 (2025).
- (11) Tao, M., Wen, L., Sheng, M. S., Yan, Z. J. & Poletti, S. Dynamics between energy intensity and carbon emissions: What does the clustering effect of labor and capital play? *J. Clean. Prod.* **452**, 142223 (2024).
- (12) Wang, Z. & Jia, X. Analysis of energy consumption structure on CO₂ emission and economic sustainable growth. *Energy Reports* **8**, 1667–1679 (2022).
- (13) Zhou, X., Guo, Q., Wang, Y. & Wang, G. Trade and Embodied CO₂ Emissions: Analysis from a Global Input–Output Perspective. *Int. J. Environ. Res. Public Heal.* **19**, 14605 (2022).
- (14) Shearer, C., Fofrich, R. & Davis, S. J. CO₂ emissions and electricity generation from proposed coal-fired power plants in India Earth's *Future* **5**, 408–416 (2017).
- (15) Foong, A., Pradhan, P., Fror, O. & Kropp, J. Adjusting agricultural emissions for trade matters for climate change mitigation. *Nat. Clim. Chang.* **13**, 3024 (2022).
- (16) Yang, M., Wang, E. & Hou, Y. The relationship between manufacturing growth and CO₂ emissions: Does renewable energy consumption matter? *Energy* **232**, 121032 (2021).
- (17) Ma, Q. *et al.* Impacts of coal burning on ambient PM_{2.5} pollution in China. *Atmos. Chem. Phys.* **17**, 4477–4491 (2017).
- (18) Apergis, N. & Payne, J. E. Per capita carbon dioxide emissions across U. S. states by sector and fossil fuel source: Evidence from club convergence tests. *Energy Econ.* **63**, 365–372 (2017).
- (19) Orkomi, A. A. Historical trends, underlying factors and the 2035 horizon situation of GHG emission in 16 Middle Eastern nations. *Energy Sustain. Dev.* **86**, 101693 (2025).
- (20) Khan, Z., Chatti, W. & Zhu, X. Public energy R & D spending and green energy for sustainable development: COP28 perspective of G7 economies. *Energy* **313**, 133754 (2024).
- (21) Emre, A. & Mehmet, C. The role of government spending on energy efficiency R & D budgets in the green transformation process: insight from the top- five countries. *Environ. Sci. Pollut. Res.* 76472–76484 (2022).
- (22) Zhang, X., Löschel, A., Lewis, J., Zhang, D. & Yan, J. Emissions trading systems for global low carbon energy and economic transformation. *Appl. Energy* **279**, 115858 (2020).
- (23) Dagnachew, A. G., Lucas, P. L., Hof, A. F. & Vuuren, D. P. Van. Trade-offs and synergies between universal electricity access and climate change mitigation in Sub-Saharan Africa. *Energy Policy* **114**, 355–366 (2018).
- (24) Aamaas, B. & Kåre, L. The impact on climate and emissions of clean household cooking energy policies in Tanzania. *Energy Policy* **192**, 114211 (2024).
- (25) Cui, C. *et al.* Energy-related CO₂ emission accounts and datasets for 40 emerging economies in 2010 – 2019. *Earth Syst. Sci. Data* **15**, 1317–1328 (2023).
- (26) Abid, I., Hechmi, S. & Chaabouni, I. Impact of Energy Intensity and CO₂ Emissions on Economic Growth in Gulf Cooperation Council Countries. *Sustainability* **16**, 10266 (2024).
- (27) Deichmann, U., Reuter, A., Vollmer, S. & Zhang, F. The relationship between energy intensity and economic growth: New evidence from a multi-country multi-sectorial dataset. *World Dev.* **124**, 104664 (2019).
- (28) Córdoba, P., Ramos, V., Gómez, E. & Gómez, J. Efficiency of Public Spending on Environmental Protection The Role of.pdf. *Sustain. Dev.* **33**, 545–560 (2025).
- (29) Li, J. *et al.* The Relationship between Energy Consumption , CO₂ Emissions , Economic Growth , and Health Indicators. *Int. J. Environ. Res. Public Heal.* **20**, 2325 (2023).
- (30) Chontanawat, J. ScienceDirect Relationship between energy consumption , CO₂ emission and economic growth in ASEAN: Cointegration and causality model. *Energy Reports* **6**, 660–665 (2020).
- (31) Pata, S. K. & Balcilar, M. Decarbonizing energy: Evaluating fossil fuel displacement by renewables in OECD countries. *Environ. Sci. Pollut. Res.* **31**, 31304–31313 (2024).
- (32) Rahman, A., Murad, S. M. W., Mohsin, A. K. M. & Wang, X. Does renewable energy proactively contribute to mitigating carbon emissions in major fossil fuels consuming countries? *J. Clean. Prod.* **452**, 142113 (2024).
- (33) Hamieh, A. *et al.* Energy Conversion and Management: X Quantification and analysis of CO₂ footprint from industrial facilities in Saudi Arabia. *Energy Convers. Manag.* **16**, 100299 (2022).
- (34) Correia, A. W. *et al.* Effect of Air Pollution Control on Life Expectancy in the United States: An Analysis of 545 U.S. Counties for the Period from 2000 to 2007. *Epidemiology* **24**, (2013).
- (35) By, E. & By, R. A study on the spatial distribution of life expectancy and its air pollution factors in China based on geographically weighted regression. *Front. Public Heal.* **13**, 1565744 (2025).
- (36) Azam, A., Ra, M., Sha, M., Yuan, J. & Salem, S. Human Development Index, ICT, and Renewable Energy-Growth Nexus for Sustainable Development: A Novel PVAR Analysis. *Front. Energy Res.* **9**, 760758 (2021).

Elastic Property Prediction of Stochastic Gyroid TPMS Lattices Using Simulation-based Machine Learning

Abdullah Alduaydi and Ammar Alsheghri*

Cite <https://doi.org/10.64589/juri/218958>

Submitted: November 27, 2025 Revised: January 20, 2026 Accepted: February 03, 2026

ABSTRACT

Triply periodic minimal surface (TPMS) lattices have emerged as promising architectural materials for lightweight structural, thermal, and biomedical applications owing to their smooth curvature, tunable relative density, and favorable manufacturability. However, predicting their elastic behavior remains challenging, especially for stochastic variants in which geometric irregularities introduce nonlinear structure–property relationships. This study developed a systematic machine-learning framework to predict the isotropic elastic constants of stochastic gyroid-based TPMS lattices. A dataset was constructed by varying the unit-cell size, relative density, architecture type (sheet or solid), and surface area, with homogenized elastic properties extracted from the stiffness and compliance matrices based on the isotropic assumption. The Random Forest (RF) and Bayesian Ridge Regression (BRR) models were evaluated using an 80/20 training–test split. For stiffness-related properties, the RF model achieved a coefficient of determination (R^2) value of 0.85 for the elastic modulus E , 0.92 for the shear modulus G , 0.86 for the bulk modulus K , and strong performance for the anisotropy index (AU) ($R^2 = 0.80$), demonstrating its ability to capture nonlinear geometric behavior. The BRR model provided superior accuracy for E , G , K , Poisson’s ratio, and Zener’s ratio and offered interpretable linear coefficients. However, the BRR failed to capture the AU. These two models provide complementary benefits: RF excels in predictive fidelity, whereas BRR enables equation-based design insights. The framework establishes a fast, accurate, and interpretable approach to predict TPMS elasticity and supports future work that integrates experimental validation using additively manufactured polyethylene terephthalate glycol samples and a neural-network-based inverse design.

Keywords: stochastic TPMS lattices, isotropic elasticity, elastic property prediction, random forest regression, bayesian ridge modeling, additive manufacturing

1. INTRODUCTION

Triply periodic minimal surface (TPMS) lattices are architected materials defined by smooth, continuous surfaces that repeat periodically in three dimensions while minimizing the surface area per unit volume^{1,2}. Unlike strut-based lattices, TPMS architectures eliminate sharp nodes and stress concentrations, thereby enabling improved load transfer, enhanced fatigue resistance, and high compatibility with additive manufacturing². Their fully interconnected networks and tunable topology allow concurrent control of stiffness, strength, permeability, and surface area, making them attractive for lightweight aerospace and automotive components, thermal management systems, and biomedical scaffolds that replicate trabecular bone morphology^{3–5}.

Recent research has extended the TPMS design to hybrid and stochastic variants, where controlled geometric perturbations are introduced to tailor the mechanical isotropy and deformation behavior. Chen et al. showed that compositionally graded TPMS lattices improve the stiffness–isotropy performance across multiple loading directions⁵. A related study by Al-Ketan et al.

demonstrated that introducing stochasticity suppressed preferential deformation modes, reduced elastic anisotropy, and promoted a stretching-dominated response, often at the expense of peak stiffness^{3,6}. These findings collectively reveal an inherent tradeoff between maximizing the stiffness and achieving isotropic elastic behavior, which is an essential consideration in architected lattice design.

The prediction of the elastic response of stochastic TPMS lattices across broad geometric design spaces remains challenging. Experimental testing is indispensable for characterizing stiffness and isotropy and for calibrating numerical models; however, exhaustive experimental campaigns become impractical when the topology, relative density, and stochastic parameters vary simultaneously. Consequently, high-fidelity numerical homogenization and finite-element simulations are extensively employed; however, their computational cost scales rapidly with design complexity. Therefore, machine learning (ML) has emerged as a surrogate modeling framework capable of capturing structure–property relationships while remaining grounded in experimentally validated and simulation-based data.

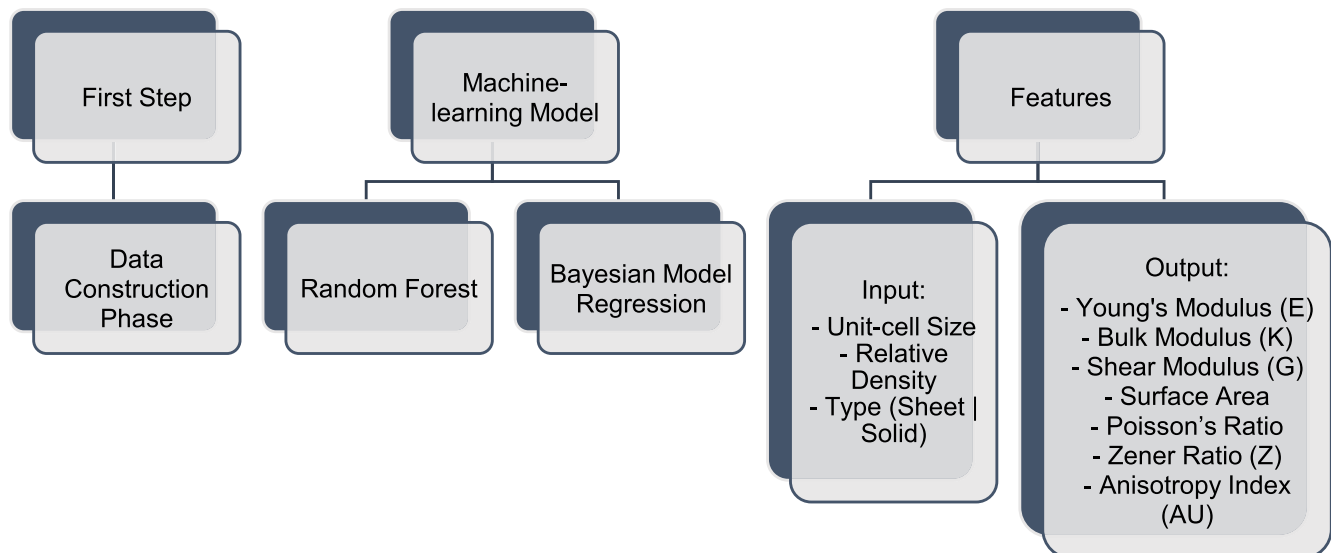


Figure 1. Two-stage methodology for triply periodic minimal surface (TPMS) lattice generation and machine-learning prediction

Recent ML studies have highlighted both the predictive accuracy and interpretability potential. Gongora et al. demonstrated that Gaussian process regression can accurately predict elastic properties across diverse TPMS families⁷. To enhance interpretability, Shapley Additive Explanations—a game-theory-based feature-attribution method that quantifies the contribution of each input variable to a model's prediction—was used to identify the dominant geometric drivers of stiffness and anisotropy⁸.

Phase field-based investigations further support the engineering relevance of stochastic TPMS. Using a Cahn–Hilliard framework, Al-Ketan et al. reported that increasing stochasticity systematically drives the absolute anisotropy index (AU) toward zero, yielding near-isotropic elastic behavior^{3,9}. Complementary experimental studies on additively manufactured stochastic gyroid sheets revealed stretching-dominated deformation, delayed localization, and suppressed shear band formation¹⁰. Together, these results reinforce the stiffness–isotropy tradeoff as a fundamental design constraint for TPMS lattices.

Despite these advances, elastic property prediction for stochastic TPMS remains computationally expensive, and many existing ML approaches rely on large datasets or deep neural networks, which limit interpretability and practicality when data availability is restricted. Therefore, regression-based models that can deliver accurate and physically interpretable predictions using modest sample sizes must be evaluated.

In this study, a curated dataset of stochastic TPMS lattices was computationally generated and used to evaluate two regression-based ML models, Random Forest (RF) and Bayesian Regression (BR), to predict isotropic elastic properties. The predicted outputs include Young's modulus (E), governing axial stiffness; shear modulus (G), describing resistance to shear deformation; bulk modulus (K), quantifying volumetric compressibility; Poisson's ratio (ν), reflecting transverse strain coupling; AU, measuring directional stiffness variation; and the Zener ratio (Z), a classical indicator of elastic isotropy³. Additively manufactured samples fabricated from polyethylene terephthalate glycol (PETG) were referenced to motivate material selection. Experimental

validation and neural-network-based modeling are reserved for future work.

The scope of this study is limited to the computational generation and mechanical evaluation of stochastic TPMS lattices. The synthesized input features included relative density, unit-cell size (UCS), and TPMS realization type (sheet or solid). These parameters were used to train the regression models, which were subsequently employed to predict the aforementioned elastic constants, providing quantitative insight into the accuracy, interpretability, and applicability of classical ML methods for stochastic TPMS design.

2. METHODOLOGY

2.1. Overview of the Methodological Framework.

The methodology adopted in the study followed a structured two-stage workflow. In the first stage, a dataset of stochastic TPMS lattice structures was generated using geometry lattice software, which enabled the parametric control of the TPMS unit-cell geometry, relative density, and lattice realization type. This dataset provides the geometric descriptors and elastic constants required to train the ML models. A standard 80/20 train–test split was used to allocate the samples for training and validation. Figure 1 presents the overall computational workflow for TPMS lattice generation and ML-based elastic property prediction.

2.2. Dataset Construction. ML algorithms require structured, high-quality datasets to identify patterns within complex architectures. Simulation-generated datasets offer major advantages because they allow precise control of the parameters influencing geometric and mechanical behavior while eliminating noise and uncertainties inherent to early experimental testing¹¹. The TPMS lattice geometries were generated using latticing software, where all input parameters were defined explicitly through the lattice design interface of the software. Figure 2 shows a representative screenshot of an analogous input panel from msLattice¹².

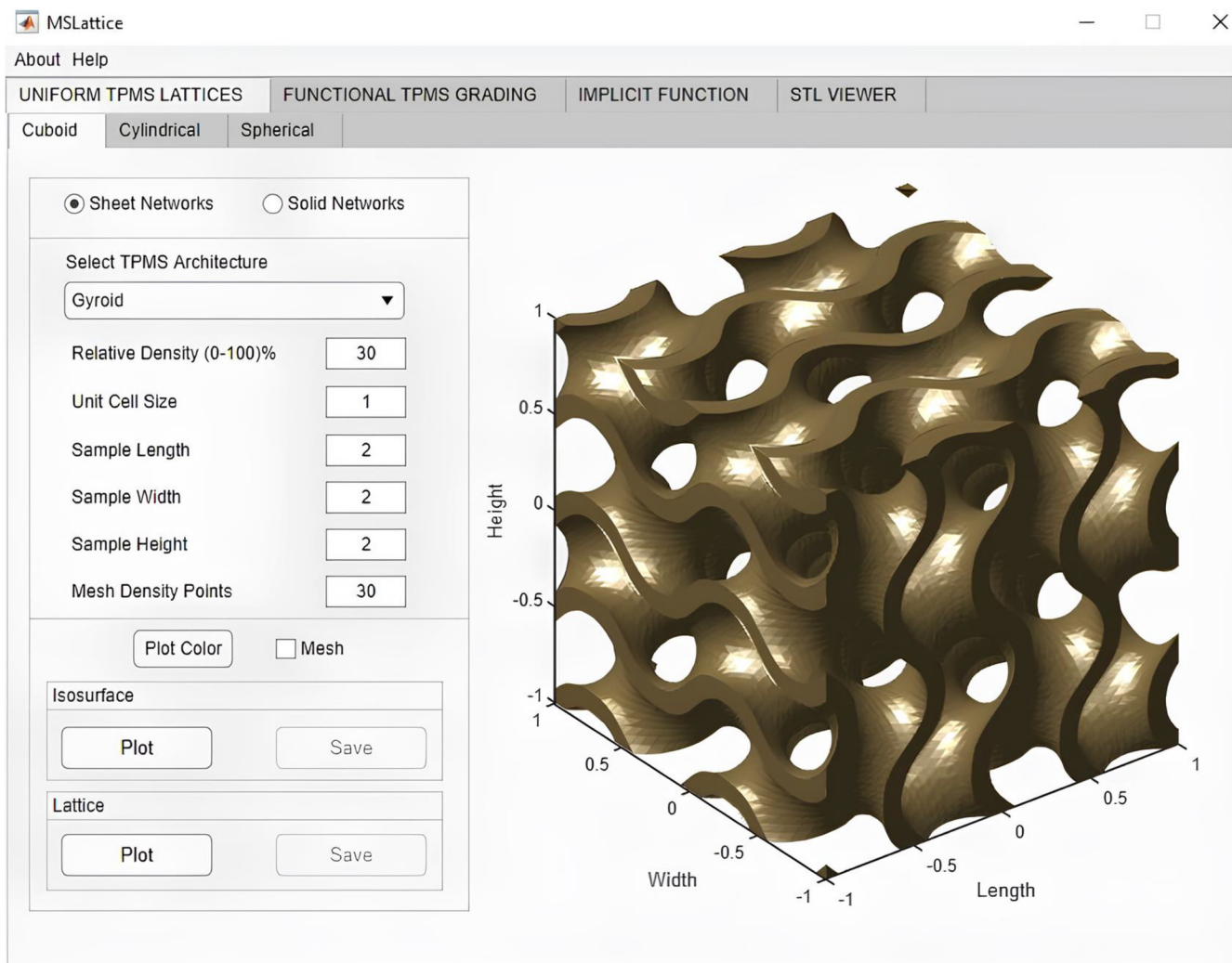


Figure 2. Software interface showing the input panel

2.2.1. Input parameters and computational settings.

Each TPMS lattice is characterized by four primary inputs:

1. UCS
2. Relative density (RD %)
3. TPMS type: Sheet or solid
4. Surface area (automatically computed by the software)

The UCS defines the characteristic geometric length scale of the TPMS lattice and directly influences stiffness scaling, deformation mode, and manufacturability. Larger UCS values typically reduce stiffness at a fixed relative density, whereas smaller UCS values increase structural rigidity but increase manufacturing and meshing costs. Furthermore, RD represents the ratio of the solid material volume to the total unit cell volume and is a dominant factor governing the elastic moduli of cellular materials. The RD values were selected within the commonly reported ranges for mechanically functional TPMS lattices to ensure physically meaningful stiffness predictions^{3,6}. Moreover, sheet-based TPMS lattices consist of thin minimal-surface shells and are commonly used in thermal management and biomedical scaffolds because of their high surface-area-to-volume ratios. Solid TPMS lattices represent volume-filled counterparts and are typically employed in load-bearing and energy-absorbing

applications where higher stiffness is required^{2,6,11}. Finally, surface area is computed directly by the software program and serves as a derived geometric descriptor reflecting geometric complexity and connectivity that indirectly influences the elastic response.

To ensure numerical consistency, mesh density = 100 (dimensionless discretization level) and mechanical density = 80 (solver integration resolution) were fixed across all samples. Smaller values are avoided to prevent discretization-induced numerical artifacts and unstable stiffness predictions.

2.2.2. Data consistency and numerical quality control.

The computational workflow discretizes each TPMS model using a consistent mesh density to ensure numerical comparability among the samples. The exported stiffness represents homogenized elastic responses computed at the unit cell level, avoiding boundary-condition artifacts commonly encountered in manually generated models and providing stable mechanical data for supervised learning. All geometric and mechanical outputs were examined to confirm the absence of numerical instabilities such as negative stiffness values, irregular compliance behavior, or nonphysical Poisson's ratios before inclusion in the ML dataset. Floating surfaces are typical examples of numerical artifacts that can arise in poorly constrained lattice models,

often leading to nonphysical stiffness values. This quality control step ensured that the regression models were trained only on physically meaningful elastic responses.

2.2.3. Output parameters and elasticity computation.

The elastic constants were computed based on two homogenization assumptions:

- Cubic-symmetric lattice assumption
- Isotropic assumption

The cubic-symmetric lattice assumption, where the elastic symmetry reflects the periodic cubic nature of the TPMS unit cell, reduces the number of independent stiffness components and improves the numerical robustness. Furthermore, an isotropic assumption was adopted in this study because previous investigations have shown that stochastic TPMS lattices exhibit near-isotropic elastic behavior, particularly at moderate to high stochasticity levels^{5,8,10}. This assumption enables a compact representation of the mechanical behavior while maintaining an acceptable accuracy for design-level predictions.

Under these assumptions, the elastic properties extracted were Young's modulus (E), shear modulus (G), bulk modulus (K), Poisson's ratio (ν), the absolute anisotropy index (AU), and the Zener ratio (Z).

2.2.4. Example of the data-collection procedure. An example corresponding to the first batch of samples is described to illustrate the dataset construction process. For an initial UCS of 0.20 mm, both solid and sheet TPMS realizations were generated. The UCS represents the characteristic geometric length scale of the TPMS lattice and directly influences stiffness scaling, deformation mechanisms, and manufacturability. The value of 0.20 mm was selected as the lower bound that is within the mesh-resolvable and additively manufacturable limits while enabling sufficient geometric detail for accurate homogenization.

For solid TPMS lattices, the RD was varied discretely across values of 30%, 35%, 50%, 60%, 65%, 70%, 75%, and 80%. These values were selected to span the low-, medium-, and high-density regimes commonly reported in TPMS mechanical studies, allowing systematic observation of density-driven stiffness scaling. For each density level, a distinct solid TPMS lattice was generated, and the corresponding elastic constants, E , G , K , ν , AU , and Z , were computed via numerical homogenization.

For *sheet TPMS lattices*, relative density values were not prescribed directly but emerged from the manipulation of the level-set value. As a result, the achieved RD values were non-integer and included 31.81%, 33.33%, 34.84%, 36.33%, 40.67%, 44.84%, 48.84%, 52.67%, and 56.33%. The digits after the decimal point reflect the geometric constraints imposed by the sheet-thickness-to-UCS ratio, which demonstrates the physically meaningful outcomes of the lattice construction process rather than an arbitrary numerical precision. Each resulting sheet-type lattice exhibits distinct surface area characteristics and elastic responses.

This procedure was repeated for UCS = 0.25 mm and UCS = 0.333 mm, values chosen to represent intermediate and larger geometric length scales frequently examined in TPMS mechanical research³. Collectively, these configurations produced a final dataset of 50 samples, ensuring coverage across the density

Table 1. Parameters used to generate the triply periodic minimal surface (TPMS) dataset (unit cell size (UCS), TPMS type, and corresponding relative densities)

UCS (mm)	Lattice type	Relative density (%)
0.2	Solid	30, 35, 50, 60, 65, 70, 75, 80
	Sheet	31.81, 33.33, 34.84, 36.33, 40.67, 44.84, 48.84, 52.67, 56.33
0.25	Solid	50, 57, 60, 65, 70, 75, 80
	Sheet	41.02, 49.14, 52.95, 56.59, 60.09, 63.42, 66.58, 69.59, 72.43
0.333	Solid	30, 35, 40, 45, 50, 55, 60, 65, 70
	Sheet	33.04, 37.62, 42.04, 50.34, 54.24, 57.96, 61.51, 64.88

regimes, realization types, and length scales. Table 1 summarizes the parameter variation strategies used to construct the dataset.

2.3. Lattice Structure Design. The present study focuses on stochastic gyroid-based TPMS structures selected owing to their well-documented mechanical and manufacturability advantages. Gyroid lattices lack flat surfaces—leading to improved stress distribution—exhibit smooth curvature that is favorable for additive manufacturing, and demonstrate enhanced isotropy when stochastic variations are introduced^{3,10,11}. Phase-field-generated stochastic TPMS variants based on the Cahn–Hilliard formulation have been shown to systematically reduce elastic anisotropy ($AU \rightarrow 0$) and mitigate stress localization, providing strong physical justification for their selection⁹.

Two categories of TPMS unit cell realizations were considered: *solid* and *sheet-based* lattices, as illustrated in Figure 3. The selected UCS values of 0.20, 0.25, and 0.333 mm fall within ranges commonly investigated in TPMS mechanical property and architected-material research, enabling the direct comparison of stiffness evolution with geometric scale^{5,10,11}. Solid lattices were generated over RD ranges of approximately 30–80%, whereas sheet lattices spanned an RD range of 31–72%, reflecting the inherent geometric constraints associated with sheet-based constructions. The parameter-sweeping strategy was designed to capture the nonlinear interactions between density, geometry, and mechanical response across representative TPMS configurations.

3. ML MODELS

ML provides a data-driven framework for identifying complex nonlinear relationships between geometric descriptors and mechanical properties of architectural materials. In the context of TPMS lattices, ML enables an efficient approximation of elastic responses that would otherwise require repeated high-fidelity homogenization simulations. Regression-based ML models are particularly suitable for small-structured datasets (e.g., $n = 50$), where the number of samples substantially exceeds the number of features ($n \gg p$), thereby ensuring statistical stability and limiting overfitting¹³. In such settings, model interpretability and robustness are essential for deriving meaningful engineering insights.

Two regression-based ML models were selected based on their suitability for limited datasets, ability to capture nonlinear relationships, and interpretability: RF Regression and BR.

3.1. RF Regression. RF regression is an ensemble learning method that constructs multiple decision trees using bootstrapped subsets of training data and aggregates their predictions to reduce variance and overfitting^{11,14,15}. The RF framework is well suited for nonlinear mappings, mixed numerical and categorical inputs, and small-to-moderate datasets commonly encountered in computational studies on material.

In this study, four input features were used for each TPMS lattice sample: UCS, RD, surface area (SA), and type. The categorical TPMS type (sheet or solid) was converted into a numerical representation using one-hot encoding (sheet = 1; solid = 0). The resulting input vector is expressed as follows:

$$X_i = [UCS_i, RD_i, SA_i, Solid\ or\ Sheet_i] \quad (1)$$

where the last two terms correspond to the encoded lattice type. The output vector

$$Y_i = [E, G, K, \nu, AU, Z] \quad (2)$$

contains the elastic properties predicted by the model.

Each decision tree produces an independent prediction for a given input, and the final RF prediction is obtained by averaging over all the trees.

$$\hat{y}_i = \frac{1}{k} \sum_{j=1}^k y_{tree\ j} \quad (3)$$

where K denotes the number of trees in an ensemble¹¹. A schematic of the RF architecture is shown in Figure 4 below¹⁶.

The model performance was evaluated using the *mean absolute error*, which measures the average magnitude of the prediction error; *root-mean-square error (RMSE)*, which introduces increased penalties for larger deviations; and the *coefficient of determination (R^2)*, which quantifies the fraction of variance explained by the model^{11,14,15}.

3.2. Bayesian Regression (BR). BR offers a probabilistic framework for regression modeling that includes both available data and prior knowledge of the parameter^{11,17}. The importance of this approach is proven when the uncertainty is critical or when confronted with limited data, which is the optimum choice for the dataset generated from 50 samples. In linear BR, the relationship is captured in a probabilistic manner between the input features, labeled as X , and output features, labeled as Y . For each lattice sample,

$$p(y_i | x_i, w) = \mathcal{N}(w^T x_i, \sigma^2) \quad (4)$$

where w is the vector of regression coefficients, and σ^2 represents observation noise^{11,17}. A prior distribution is assigned to w , and the posterior distribution is obtained by updating this prior using the training dataset (X, y) :

$$p(\hat{y}_i | x_i, X, y) \quad (5)$$

The same input and output variables used in the RF model were employed for BR. A schematic of the BR architecture is shown in Figure 5. This approach provides both point estimates and confidence intervals, allowing for more robust interpretation of the variability in stochastic TPMS structures¹⁷.

4. RESULTS AND DISCUSSION

4.1. Elastic Behavior of the Generated TPMS Lattices.

The effective elastic constants of the generated TPMS lattices were evaluated based on two homogenization assumptions: (i) cubic symmetric lattice behavior and (ii) isotropic behavior. Previous investigations have demonstrated that stochastic TPMS lattices exhibit near-isotropic elastic responses, particularly when fabricated through additive manufacturing, owing to the suppressed directional bias and enhanced geometric randomness. Accordingly, isotropic elastic constants, E , G , K , ν , AU , and Z , were selected as target outputs for the ML models^{8,10,11}.

Across the dataset, increasing the RD led to a systematic increase in E , G , and K , consistent with classical cellular-solid scaling laws. Sheet-based lattices exhibit lower stiffness than

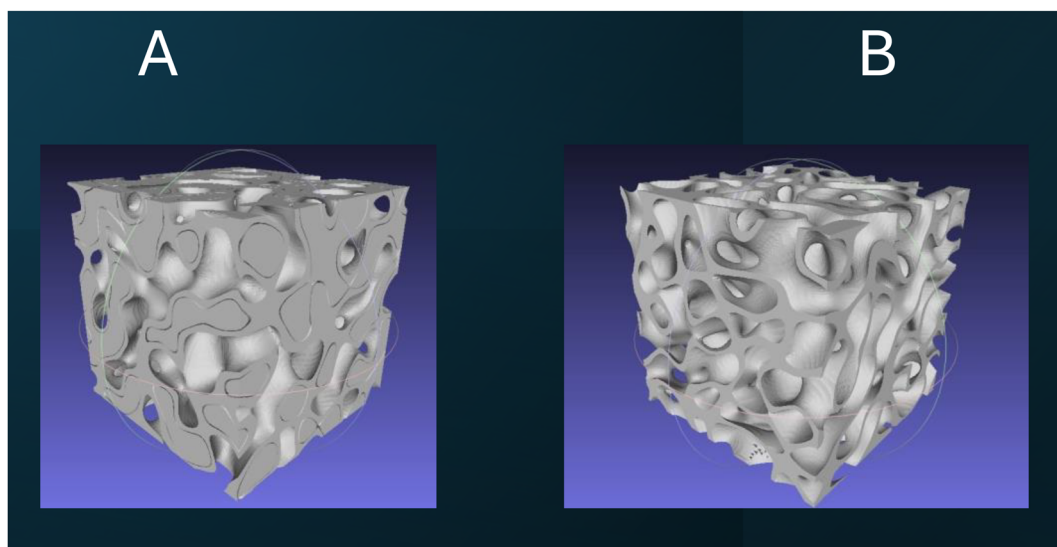


Figure 3. TPMS lattice categories considered in this study: (A) solid and (B) sheet

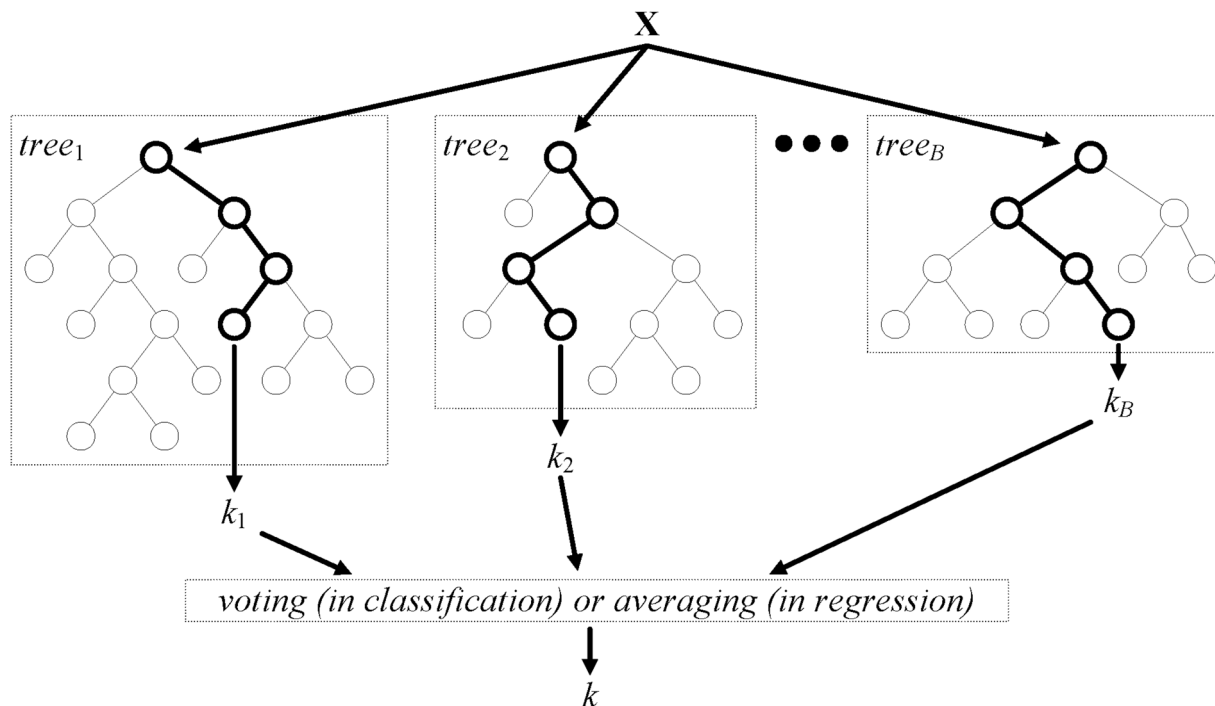


Figure 4. Example of the Random Forest architecture

their solid counterparts at comparable densities, reflecting a reduced load-bearing cross-section but higher surface-area-to-volume ratios. Variations in the UCS further influence the stiffness scaling, with smaller unit cells typically producing higher effective moduli owing to increased geometric constraints. The AU and Z values confirmed reduced elastic anisotropy across most stochastic configurations, supporting the validity of the isotropic assumption for surrogate modeling.

4.2. Elasticity Trends and Parameter Correlations.

This section illustrates the effect of TPMS geometry on the elastic response. Each plot relates the input descriptors, RD (%), UCS, surface area, and architectural type (sheet/solid), to the elastic outputs E, G, K, ν , AU, and Zener ratio.

4.2.1. Absolute AU. As shown in Figure 6, the absolute AU exhibited the strongest and most consistent correlation with the

RD across all three UCS values (0.20, 0.25, and 0.333 mm) for both the sheet and solid architectures. However, its dependence on the surface area was more nuanced. For UCS = 0.333 mm, AU exhibits strong correlations for both architectures, with correlation coefficients of approximately $R^2 \approx 0.87$ for sheet lattices and $R^2 \approx 0.95$ for solid lattices. For the smaller UCS values of 0.20 and 0.25 mm, sheet lattices again yield high correlation values ($R^2 \approx 0.96$ and $R^2 \approx 0.90$), whereas solid lattices display weak surface-area dependence ($R^2 \approx 0.50$ or less). This indicates that solid TPMS structures with small UCS behave as surface-area-independent structures, whereas larger solid cells (0.333 mm) show increased sensitivity to the surface area. This pattern aligns with the findings in the stochastic TPMS literature, which show that a sufficient microstructural volume is required for the AU to meaningfully reflect anisotropy variations.

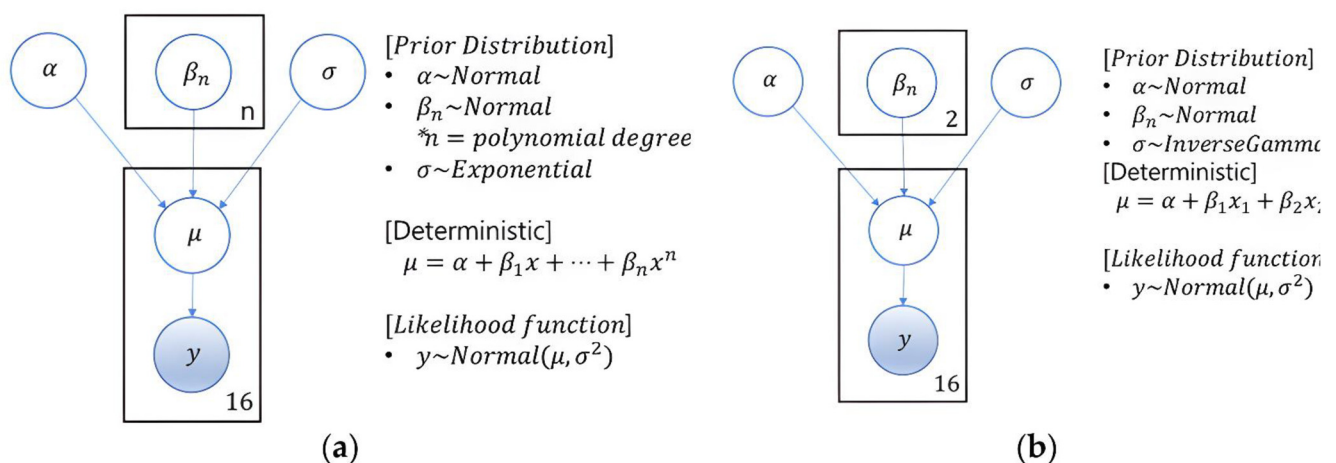


Figure 5. Example of the architecture of the Bayesian regression

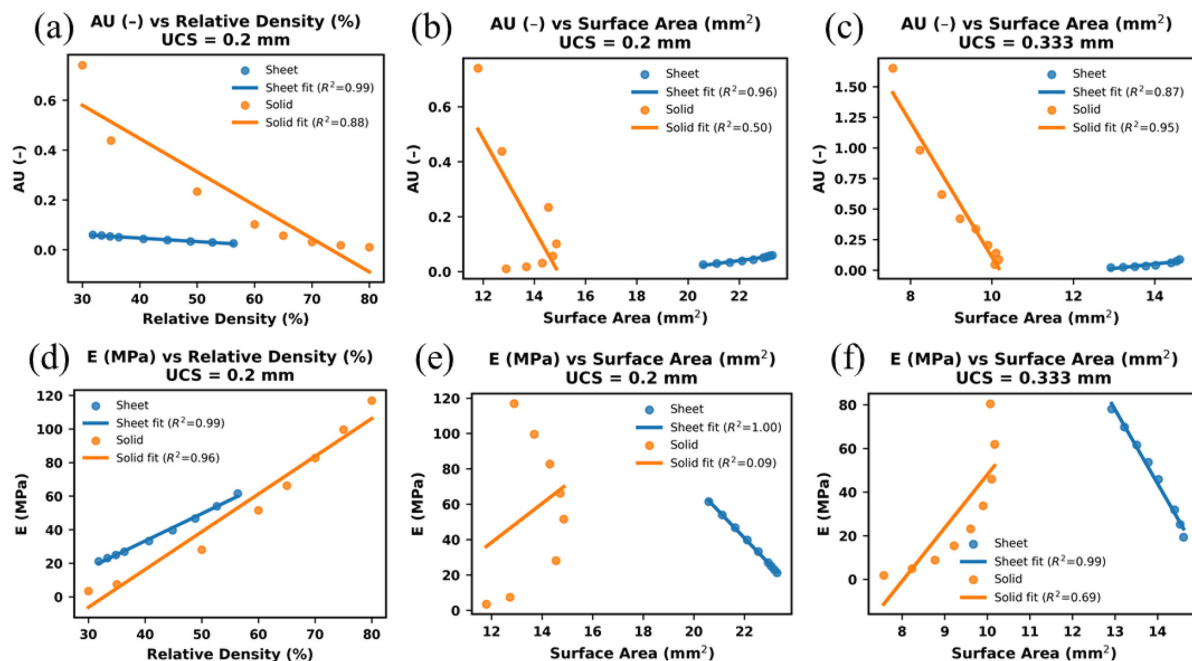


Figure 6. Correlation of absolute anisotropy (AU) and Young's modulus (E) with the geometric input parameters of sheet and solid stochastic TPMS structures: (a) AU versus relative density at UCS = 0.2 mm; (b) AU versus surface area at UCS = 0.2 mm; (c) AU versus surface area at UCS = 0.333 mm; (d) E versus relative density at UCS = 0.2 mm; (e) E versus surface area at UCS = 0.2 mm; and (f) E versus surface area at UCS = 0.333 mm

4.2.2. Elastic modulus (E). As shown in Figure 6, E correlates strongly with the relative density, which is consistent with classical cellular-solid scaling³. However, the effect of surface area varies according to the architecture. Sheet TPMS displays a strong correlation between E and surface area ($R^2 \approx 0.90$). By contrast, solid TPMS (UCS = 0.20 mm) show nearly no correlation between E and surface area and those at UCS = 0.333 mm show a moderate correlation ($R^2 \approx 0.69$). These observations suggest that sheet-type architectures are predominantly surface-area-dominated, whereas solid architectures derive stiffness primarily from the relative density, especially at finer UCS values.

4.2.3. Shear and bulk modulus (G and K). As shown in Figure 7, both G and K exhibit trends similar to that of E. They show a strong dependence on the RD for all architectures. Surface area plays a significant role in sheet lattices, where it strongly influences G and K. By contrast, the solid lattices showed minimal sensitivity to the surface area, especially at the UCS value of 0.20 mm. These patterns are consistent with the physical understanding that the stiffness in sheet-type TPMS is governed by the wall thickness and curvature, whereas the solid TPMS stiffness is primarily associated with the strut volume and density.

4.2.4. Poisson's ratio (ν). As shown in Figure 8, ν demonstrates a clear architectural asymmetry. Sheet-type lattices exhibit a strong correlation with RD, indicating that ν increases or decreases predictably with density variations. Conversely, solid-type lattices exhibit near-zero correlation between ν and RD, suggesting that their lateral deformation behavior is largely decoupled from density. This means that a solid TPMS yields a ν value that remains relatively stable regardless of the density, whereas sheet lattices provide greater tunability at the expense of stronger geometry-property coupling. Surface-area correlations

were reversed at a very small UCS value (0.20 mm), highlighting microstructural scale effects.

4.2.5. Zener ratio (Z). As shown in Figure 8, the variation of Z exhibits a strong architectural dependence and notable variability at different UCS values. Moreover, as shown in Figure 9, when UCS = 0.20 mm, both sheet and solid lattices exhibit weak correlations (solid ≈ 0.24). When UCS = 0.25 mm, sheet lattices dominate the correlation ($R^2 \approx 0.68$), whereas at UCS = 0.333 mm, solid lattices exhibit the strongest correlation ($R^2 \approx 0.84$). These inconsistencies arise because the value of Z is derived from the elastic tensors of ideal cubic materials^{18,19}. When applied to finite, stochastic, and noncubic cellular structures, size effects and microstructural heterogeneity introduce variability until a sufficiently large representative volume is reached.

4.3. ML Predictions. Two supervised learning approaches were implemented to predict the elastic response of stochastic TPMS lattices: the RF Regressor and BR Regression (BRR) models. The input feature set consisted of RD, UCS, surface area, and categorical architecture type, encoded as sheet = 0 and solid = 1. The six predicted elastic properties were E, G, K, ν , AU, and Z. Model performance was quantified using the RMSE and R^2 values, allowing a direct comparison between the two learning approaches.

4.4. RF Performance. Table 2 summarizes the performance of the RF model. The RF model achieved R^2 values of 0.85 for E, 0.92 for G, and 0.86 for K, demonstrating an excellent predictive capability for stiffness-related properties. It also performed strongly for the anisotropy index AU, achieving an R^2 value of 0.80, which was notably higher than that of the Bayesian model for AU. The model provided reasonably good predictions for Z ($R^2 \approx 0.70$). However, the ν exhibited weak predictive performance ($R^2 = 0.26$), indicating that it is only weakly influenced

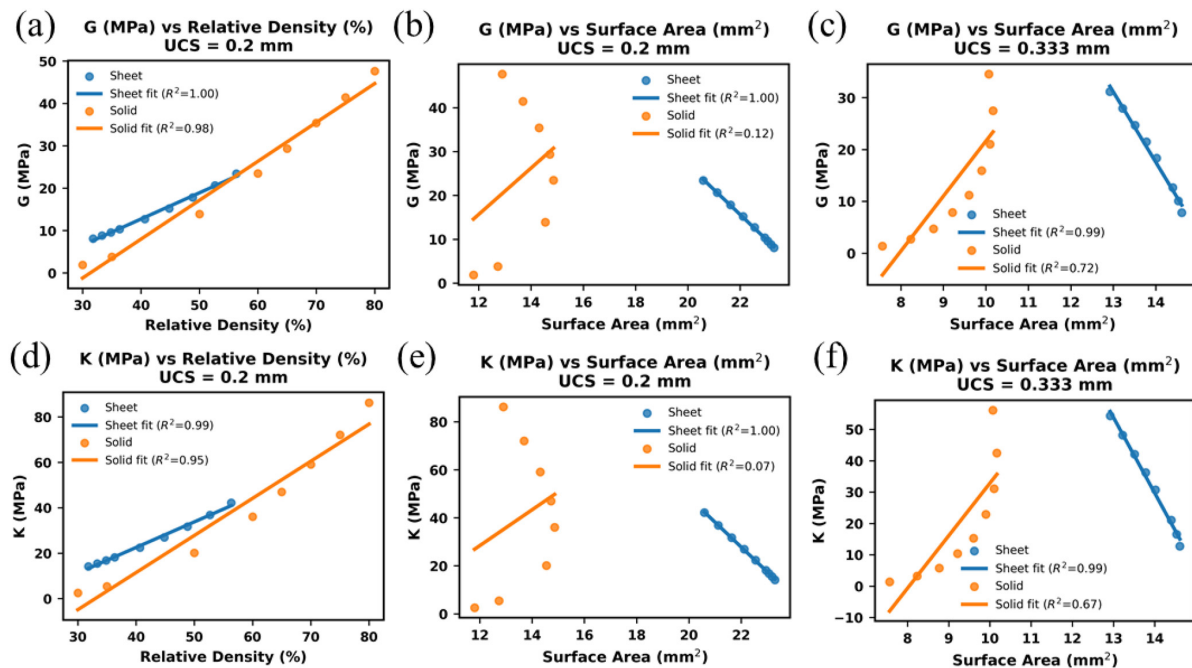


Figure 7. Correlation of shear modulus (G) and bulk modulus (K) with the geometric input parameters of sheet and solid stochastic TPMS structures: (a) G versus relative density at UCS = 0.2 mm; (b) G versus surface area at UCS = 0.2 mm; (c) G versus surface area at UCS = 0.333 mm; (d) K versus relative density at UCS = 0.2 mm; (e) K versus surface area at UCS = 0.2 mm; and (f) K versus surface area at UCS = 0.333 mm

by the current geometric descriptors—consistent with the earlier elasticity analysis. Overall, RF performed best when high predictive fidelity was required, particularly for AU and stiffness-related properties.

4.5. Bayesian Ridge Regression (BRR) Performance.

Table 3 presents the BRR results. The BRR model outperformed the RF model for E, G, and K, achieving slightly lower RMSE values and R^2 scores of 0.88, 0.92, and 0.86, respectively. This model also demonstrated substantially better performance for

ν , with R^2 improving from 0.26 (RF) to 0.57. The Bayesian model delivered the highest accuracy for the Zener ratio, reaching $R^2 \approx 0.81$. However, it performed poorly for AU, yielding a negative R^2 value (-0.40), indicating that the anisotropy metric exhibited nonlinear relationships that were not well captured by a linear model. These results highlight that BR is advantageous when interpretability and linear relationships dominate, whereas RF is preferred when capturing nonlinear geometry–property interactions, particularly for anisotropy characteristics.

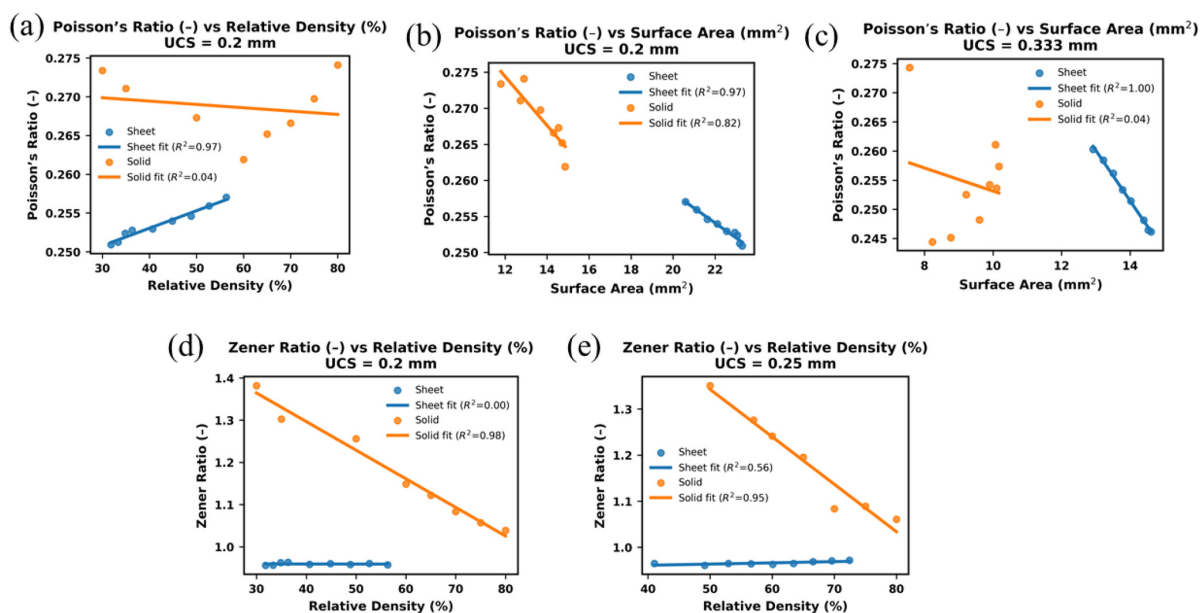


Figure 8. Correlation of Poisson's ratio and Zener ratio with the geometric input parameters of sheet and solid stochastic TPMS structures: (a) Poisson's ratio versus relative density at UCS = 0.2 mm; (b) Poisson's ratio versus surface area at UCS = 0.2 mm; (c) Poisson's ratio versus surface area at UCS = 0.333 mm; (d) Zener ratio versus relative density at UCS = 0.2 mm; and (e) Zener ratio versus relative density at UCS = 0.25 mm

Table 2. Random Forest performance metrics

Variable	Root-mean-square error (RMSE)	Coefficient of determination R^2
Young's modulus (E)	8.39	0.85
Shear modulus (G)	2.16	0.92
Bulk modulus (K)	5.76	0.86
Anisotropy index (AU)	0.038	0.80
Poisson's ratio (ν)	0.00745	0.26
Zener ratio (Z)	0.0726	0.70

Table 3. Bayesian Ridge performance metrics

Variable	RMSE	R^2
E	7.33	0.88
G	2.13	0.92
K	5.64	0.86
AU	0.101	-0.40
ν	0.00571	0.57
Z	0.0578	0.81

4.6. Interpretable Bayesian Coefficients. The BRR model was introduced to complement the RF model by providing an *explicit analytical relation* between geometry and mechanical response. For each output y , the Bayesian model takes the following form:

$$y \approx \beta_0 + \beta_{RD} \cdot RD + \beta_{UCS} \cdot UCS + \beta_{SA} \cdot SA + \beta_{Solid} \cdot I(Solid), \quad (6)$$

where $I(Solid) = 1$ for solid architectures and 0 for *sheet* architectures, respectively. This formulation allows for a direct physical interpretation of each coefficient.

4.6.1. Elastic modulus (E). The relation of E indicates that increasing RD and surface area increases stiffness, whereas a larger UCS slightly reduces it. The negative coefficient of $I(Solid)$ suggests that at fixed RD/UCS/SA settings, sheet architectures tend to be stiffer than solid ones (because solid = 1 introduces a negative shift).

$$E \approx -75.6 + 2.12 RD - 0.104 UCS + 1.07 SA - 5.72 I(Solid) \quad (7)$$

4.6.2. Shear modulus (G). As shown in Eq. 8, the trends of the shear modulus are consistent with E; RD and SA have positive contributions, UCS has a small negative effect, and sheet-type lattices again show a slightly higher shear stiffness compared with solids.

$$G \approx -30.8 + 0.863 RD - 0.0062 UCS + 0.395 SA - 0.335 I(Solid) \quad (8)$$

4.6.3. Bulk modulus (K). As indicated by Eq. 9, the effects of RD and SA on the volumetric stiffness are strong and positive, whereas the UCS and the solid indicator are both slightly opposite, again implying that finer and more sheet-like lattices are more efficient in bearing volumetric loads for a fixed RD.

$$K \approx -58.8 + 1.53 RD - 0.0189 UCS + 0.896 SA - 1.15 I(Solid) \quad (9)$$

4.6.4. Poisson's ratio (ν). As indicated by Eq. 10, the extremely small magnitudes of the RD and SA coefficients suggest that the ν is primarily controlled by the UCS and topology, with the UCS exerting a more pronounced negative influence; that is, smaller cells tend to yield higher ν values, which is also observed in the elasticity results in the previous section.

$$\nu \approx 0.310 + 1.78 \times 10^{-4} RD - 0.134 UCS - 0.00165 SA - 8.59 \times 10^{-4} I(Solid) \quad (10)$$

4.6.5. Zener ratio (Z). The Z value decreased at increasing RD, UCS, and SA values, as indicated by Eq. 11, whereas the solid architectures ($I(Solid) = 1$) exhibited a modest increase in the Z values relative to the sheet structures. The relatively high R^2 value (≈ 0.81) indicates that this linear relationship captures a substantial part of the anisotropic stiffness distribution.

$$Z \approx 1.65 - 0.00625 RD - 0.0521 UCS - 0.0198 SA + 0.159 I(Solid) \quad (11)$$

These equations provide a compact, design-oriented mapping from geometry to mechanical response, enabling sensitivity analysis and parametric exploration without requiring full simulations or experiments.

4.7. Implications for TPMS Mechanical Behavior and Design. The mechanical trends observed in the stochastic TPMS dataset have direct implications for engineering design and functional material optimization. The strong dependence of the stiffness-related properties (E, G, and K) on the relative

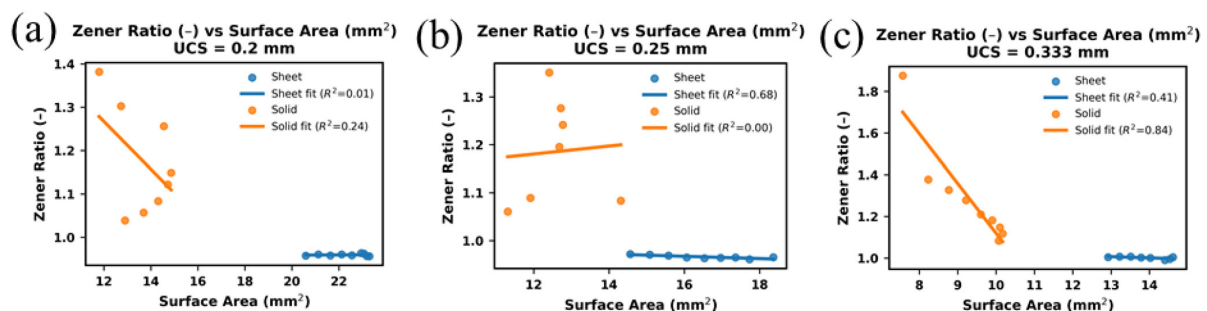


Figure 9. Correlation of Zener ratio with surface area for sheet and solid stochastic TPMS structures at different unit cell sizes: (a) Zener ratio versus surface area at UCS = 0.2 mm; (b) Zener ratio versus surface area at UCS = 0.25 mm; and (c) Zener ratio versus surface area at UCS = 0.333 mm

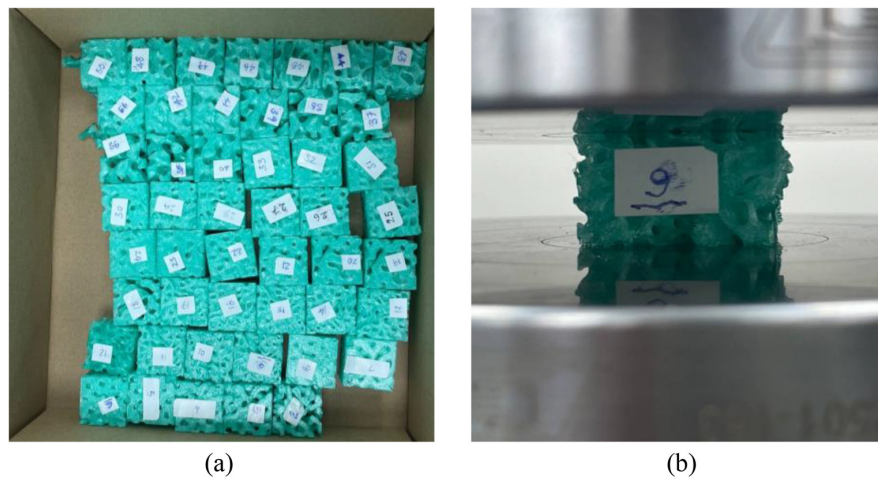


Figure 10. (a) Additively manufactured stochastic TPMS samples fabricated from polyethylene terephthalate glycol (PETG); and (b) a representative compression test specimen

density reinforces the potential of TPMS lattices for lightweight structural applications where density-driven optimization is essential, such as in aerospace components and energy-absorbing structures¹. The pronounced influence of the surface area on the sheet architecture indicates that the wall curvature and surface topology can be strategically adjusted to tune the deformation modes and enhance energy dissipation. By contrast, the relative insensitivity of solid architectures to variations in the surface area highlights their suitability for consistent load-bearing performance across a wide density range. Furthermore, the variability of the anisotropy measures, particularly the UCS-dependent behavior observed for the AU and Z values, suggests that TPMS lattices can be engineered to transition between isotropic and directionally biased mechanical responses, a trend also observed experimentally in stochastic gyroid systems^{8,9,11}. These observations collectively demonstrate that stochastic TPMS geometries offer a broad and tunable design space in which geometric descriptors can be tailored to satisfy application-specific mechanical requirements.

Despite the promising predictive performance achieved, several limitations should be acknowledged. First, the dataset consists of 50 simulation-derived samples, which, while sufficient for regression-based modeling, remains relatively small for capturing the full stochastic variability of TPMS geometries. Second, all elastic constants were obtained under the isotropic homogenization assumption, which may mask localized anisotropic effects at smaller representative volumes. Additionally, only linear elastic behavior was considered; nonlinear deformation, buckling, and failure mechanisms were not modeled. These limitations suggest that while the developed framework provides reliable elastic predictions within the defined design space, further validation with expanded datasets and physical testing is necessary to generalize the conclusions to broader TPMS applications.

4.8. Future Work. Figure 10 shows representative 3D-printed PETG samples intended for experimental validation in a subsequent study; however, they were not used in the analysis presented in this study. These samples will be mechanically tested in future studies to directly compare the

simulation-derived elastic constants with the experimentally measured values, enabling the assessment of printing-induced deviations, anisotropy effects, and material nonlinearities.

In addition to experimental validation, the ML framework developed in this study can be extended in several directions. Larger datasets generated from high-resolution phase-field simulations can be used to improve the predictive accuracy and explore a broader design space for stochastic TPMS structures. Neural-network-based models, including fully connected deep networks and physics-informed architectures, will also be explored to capture nonlinear interactions that are not well represented by linear models such as BR, particularly for certain properties such as the AU. Additionally, an inverse-design pipeline can be developed to predict the optimal TPMS geometries from the target mechanical properties, enabling accelerated design workflows for biomedical implants, structural lattices, and thermal-mechanical components.

Overall, our future work will focus on bridging computational predictions, experimental validation, and inverse design to create a complete end-to-end framework for stochastic TPMS lattice engineering.

5. CONCLUSIONS

This study investigated the elastic behavior of stochastic TPMS lattices and evaluated the predictive capabilities of two ML models, RF Regression and BRR, using geometric descriptors as inputs. By systematically varying the UCS, RD, architecture type, and surface area, a diverse dataset of 50 stochastic gyroid-based structures was created, enabling a detailed examination of the geometric-mechanical relationships under the isotropic elasticity assumption.

Analysis of the homogenized elastic properties revealed clear structure-property trends. Relative density was consistently the strongest predictor of the stiffness-related outputs (E , G , and K), whereas surface area played a dominant role in sheet-type lattices and exerted only a minor influence on fine-scale solid lattices. The Poisson's ratio exhibited architecture-dependent sensitivity, remained relatively stable in solid TPMS but varied considerably

in sheet lattices. Z exhibited substantial variability across UCSs, reflecting the difficulty in applying cubic anisotropy metrics to finite stochastic cellular structures.

The ML predictions reinforced these observations. The RF model demonstrated high accuracy for stiffness metrics and AU, achieving R^2 values in the range of 0.80–0.92 across most outputs. The BR model provided competitive or superior performance for E, G, K, ν , and Z and simultaneously yielded interpretable linear equations that exposed the relative influences of all the geometric parameters. BR was performed only for the AU, confirming that the anisotropic behavior is intrinsically nonlinear in this design space. Together, the two models form a complementary framework: RF excels in accuracy and handling nonlinearities, whereas the BR model offers interpretable, equation-based insights valuable for parametric design and optimization.

Overall, the findings of this study demonstrate that ML provides an efficient and accurate alternative to full homogenization simulations for predicting the mechanical behavior of stochastic TPMS lattices. The integration of geometric descriptors enables both predictive capabilities and mechanistic understanding, thereby supporting accelerated design workflows for architectural materials.

Future studies will extend this framework to include physical validation. PETG samples produced via additive manufacturing underwent compression testing, enabling a direct comparison between the simulated and experimental elasticities. Additionally, the ML pipeline will be expanded using neural network architectures and larger phase-field-generated datasets to explore inverse design capabilities and predict lattice geometries directly from the desired mechanical performance. The next step will be to combine computational design, physical validation, and automated TPMS synthesis.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Abdullah Alduaydi – Mechanical Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0009-0004-0211-0139
Email: s202170370@kfupm.edu.sa

Corresponding Author

Ammar Alsheghri – Research Mentor, Assistant Professor, Mechanical Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0000-0001-5651-7103
Email: ammar. sheghri@kfupm.edu.sa

ACKNOWLEDGEMENTS

The authors acknowledge the support provided by the Mechanical Engineering Department and the Interdisciplinary Research Center (IRC) for Biosystems and Machines at King Fahd University of Petroleum & Minerals (KFUPM). The authors also extend their appreciation to Dr. Mohd Tauheed, a post-doctoral fellow at the IRC for Biosystems and Machines, and to Ermias Wubete, a PhD student in the Mechanical Engi-

neering Department at KFUPM, for their support in this work.

REFERENCES

- (1) Yeranee, K. & Rao, Y. A Review of Recent Investigations on Flow and Heat Transfer Enhancement in Cooling Channels Embedded with Triply Periodic Minimal Surfaces (TPMS). *Energies* **15**, 8994 (2022).
- (2) Koh, T. Y. & Sutradhar, A. Untethered selectively actuated microwave 4D printing through ferromagnetic PLA. *Additive Manufacturing* **56**, 102866 (2022).
- (3) Al-Ketan, O., Lee, D.-W. & Abu Al-Rub, R. K. Mechanical properties of additively-manufactured sheet-based gyroidal stochastic cellular materials. *Additive Manufacturing* **48**, 102418 (2021).
- (4) Maevskaia, E., Guerrero, J., Ghayor, C., Bhattacharya, I. & Weber, F. E. Triply Periodic Minimal Surface-Based Scaffolds for Bone Tissue Engineering: A Mechanical, *In Vitro* and *In Vivo* Study. *Tissue Engineering Part A* **29**, 507–517 (2023).
- (5) Chen, R. *et al.* Ultra-stiff and quasi-elastic-isotropic triply periodic minimal surface structures designed by deep learning. *Materials & Design* **244**, 113107 (2024).
- (6) Ren, J. *et al.* Fe-Mn-Al-C high-entropy steels with superior mechanical properties at 4.2 K. *Materials & Design* **228**, 111840 (2023).
- (7) Gongora, A. E. *et al.* Accelerating the design of lattice structures using machine learning. *Sci Rep* **14**, 13703 (2024).
- (8) Ejeh, C. J., Barsoum, I. & Abu Al-Rub, R. K. Impact behavior of periodic, stochastic, and anisotropic minimal surface-lattice sandwich structures. *International Journal of Mechanical Sciences* **276**, 109359 (2024).
- (9) Gómez, H., Calo, V. M., Bazilevs, Y. & Hughes, T. J. R. Isogeometric analysis of the Cahn–Hilliard phase-field model. *Computer Methods in Applied Mechanics and Engineering* **197**, 4333–4352 (2008).
- (10) Fu, J. *et al.* Isotropic design and mechanical characterization of TPMS-based hollow cellular structures. *Composite Structures* **279**, 114818 (2022).
- (11) Hussain, S., Lee, C., Tsang, Y. P. & Waqar, S. A machine learning-based recommendation framework for material extrusion fabricated triply periodic minimal surface lattice structures. Preprint at <https://doi.org/10.21203/rs.3.rs-3723296/v1> (2024).
- (12) Al-Ketan, O. & Abu Al-Rub, R. K. MSLattice: A free software for generating uniform and graded lattices based on triply periodic minimal surfaces. *Mat Design & Process Comms* **3**, (2021).
- (13) Harrell, F. E. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. (Springer International Publishing, Cham, 2015). doi:10.1007/978-3-319-19425-7.
- (14) Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).
- (15) Tin Kam Ho. Random decision forests. in *Proceedings of 3rd International Conference on Document Analysis and Recognition* vol. 1 278–282 (IEEE Comput. Soc. Press, Montreal, Que., Canada, 1995).
- (16) Verikas, A., Vaciukynas, E., Gelzinis, A., Parker, J. & Olsson, M. Electromyographic Patterns during Golf Swing: Activation Sequence Profiling and Prediction of Shot Effectiveness. *Sensors* **16**, 592 (2016).
- (17) Yu, J. & Oh, H. AI-Based Degradation Index from the Microstructure Image and Life Prediction Models Based on Bayesian Inference. *Sustainability* **15**, 7298 (2023).
- (18) Lethbridge, Z. A. D., Walton, R. I., Marmier, A. S. H., Smith, C. W. & Evans, K. E. Elastic anisotropy and extreme Poisson's ratios in single crystals. *Acta Materialia* **58**, 6444–6451 (2010).
- (19) Kube, C. M. Elastic anisotropy of crystals. *AIP Advances* **6**, 095209 (2016).

Influence of Ba-Site Substitution on the Structural Properties of Tl-2223 Superconductor

Hawra Tuhaifa and Khalil A. Ziq*

Cite <https://doi.org/10.64589/juri/219229>

Submitted: February 12, 2025 Revised: February 03, 2026 Accepted: March 12, 2026

ABSTRACT

Thallium-based high-temperature superconductors, exhibit a strong structural sensitivity to cation substitutions. Although most previous studies focused on partial Ba-site substitutions, the effect of complete Ba-site replacement remains insufficiently explored. In this study, the atomic size effect on the structural properties of Tl-2223 was investigated by fully substituting Ba^{2+} with smaller cations, namely, Ca^{2+} and Mg^{2+} , which introduced negative chemical pressure into the lattice. Polycrystalline $\text{Tl}_2\text{Ba}_2\text{Ca}_2\text{Cu}_3\text{O}_{10}$, $\text{Tl}_2\text{Ca}_4\text{Cu}_3\text{O}_{10}$, and $\text{Tl}_2\text{Mg}_2\text{Ca}_2\text{Cu}_3\text{O}_{10}$ samples were synthesized using the solid-state reaction method and characterized using X-ray diffraction (XRD) and X-ray photoelectron spectroscopy (XPS). The XRD results revealed systematic peak shifts toward higher diffraction angles accompanied by lattice distortion, increased microstrain, and higher dislocation density with decreasing ionic radius. These findings highlight the strong influence of atomic size on the structural framework of Tl-2223 and provide a basis for future studies on its superconducting properties.

Keywords: high-temperature superconductors, atomic substitution, crystal structure, X-ray diffraction, X-ray photoelectron spectroscopy, chemical pressure

1. INTRODUCTION

Superconductivity is defined two basic properties, zero electrical resistance, and the Meissner effect below some critical transition temperature.

From a temperature-based perspective, superconductors can be categorized as low-temperature superconductors (LTS) and high-temperature superconductors (HTS). Among HTS materials, thallium (Tl)-based superconductors have attracted significant attention owing to their relatively high transition temperatures, with the Tl-2223 phase exhibiting the highest critical temperature (approximately 122 K) within the Tl-based superconducting family¹. Owing to its high transition temperature and layered crystal structure, Tl-2223 has been extensively studied for possible energy-related applications, particularly where reduced electrical loss is desired. Atomic substitution is widely adopted for tuning the internal structures and various superconducting properties. Cation substitution, including partial substitution at the Ba site, where Ba^{2+} is typically replaced in limited concentrations^{2–4} has been extensively explored for Tl-2223 superconductors. However, complete substitution at the Ba site has been rarely reported, particularly using smaller cations such as Ca^{2+} and Mg^{2+} . Consequently, the structural consequences of full Ba-site substitution and the associated chemical pressure effects remain insufficiently understood.

In this study, we investigated the atomic size effect resulting from complete Ba-site substitution in the Tl-2223 superconductor. Specifically, the Ba^{2+} ion (ionic radius: ~1.42

Å) was completely replaced by smaller cations from the same group in the periodic table, namely Ca^{2+} (~1.12 Å) and Mg^{2+} (~0.89 Å) to introduce negative chemical pressure on the crystal lattice. This approach examines the influence of full ionic replacement on the structural properties of the Tl-2223, particularly the lattice parameters, strain, and microstructural features.

Although superconducting properties were not directly measured in this study, the understanding of structural modifications is crucial for correlating chemical substitution with superconducting performance. Therefore, this study aims to provide a systematic structural comparison of Ba-site-substituted Tl-2223 samples, which can facilitate further investigations involving superconducting property measurements.

2. EXPERIMENTAL WORK

Stoichiometric ratios of high-purity (>99.9%) Tl_2O_3 , BaO_2 , CaO , and CuO powders were thoroughly ground using an agate mortar and pestle to ensure homogeneous mixing and then pressed into dense cylindrical pellets ($R \sim 1\text{ cm}$). The pellet was wrapped in double layers of silver foil⁵ and sealed inside a quartz tube partially filled with Ar gas to minimize thallium volatilization. A 10–20% excess of Tl_2O_3 was used to compensate for thallium loss during high-temperature processing⁶. The sealed quartz tube was heated at a rate of 10 K/min to 725 K and subsequently at 2 K/min to 1078 K, held at this temperature

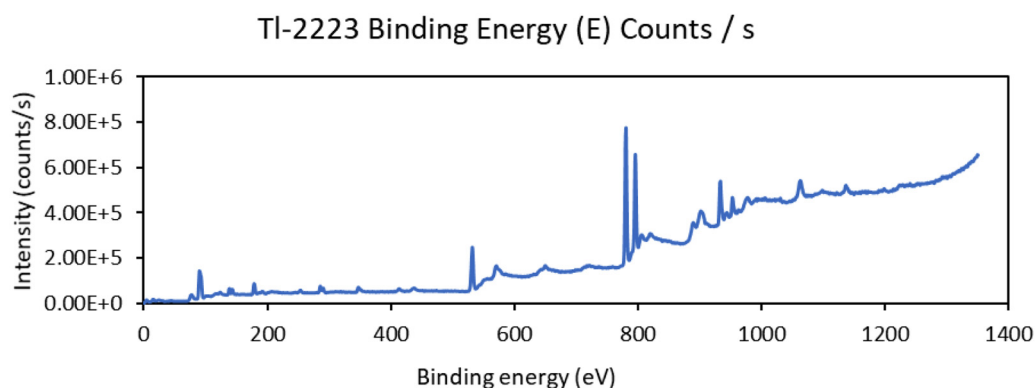


Figure 1. XPS wide scan survey for the pure Tl-2223 high temperature superconductor

for 120 min, and then allowed to cool naturally to room temperature⁷. The substituted samples were synthesized following the same procedure with adjusted stoichiometric compositions; $\text{Tl}_2\text{Ca}_4\text{Cu}_3\text{O}_{10}$, and $\text{Tl}_2\text{Mg}_2\text{Ca}_2\text{Cu}_3\text{O}$. The phase composition and crystal structure were analyzed using Bruker D2 Phaser X-ray diffractometer with Cu $K\alpha$ radiation ($\lambda = 1.54184 \text{ \AA}$). Diffraction patterns were recorded over 2θ range (10 to 80 degrees) to capture the principal reflections of the Tl-2223 phase. The lattice parameters were calculated using Bragg's law, while the crystallite size was estimated using the Scherrer equation. X-ray photoelectron spectroscopy (XPS) spectrum were obtained using a Thermo-Scientific Escalab-250-Xi XPS spectrometer. Survey and high-resolution spectra were acquired under high-vacuum conditions. Prior to detailed core-level analysis, the sample surfaces were etched using Ar-ion sputtering to reduce surface contamination⁸. The binding energy scale was calibrated according to standard XPS procedures. All measurements were conducted at room temperature to ensure consistency among the samples.

3. RESULTS AND DISCUSSIONS

3.1. X-ray Photoelectron Spectroscopy (XPS) Analysis.

3.1.1. $\text{Tl}_2\text{Ba}_2\text{Ca}_2\text{Cu}_3\text{O}_{10}$ (Standard sample). XPS was performed on the Tl-2223 high-temperature superconductor to identify various elements present in the prepared samples. The data are presented in Fig. 1 over a wide binding energy (BE) range (0–1350 eV) and intervals of 1 eV. The corresponding binding energies for all elements used are given in Table 1. The high intensity of all peaks indicates a strong signal from all the

elements. The Tl peaks strongly indicate the minimal loss of Tl, which is the most volatile element in the sample.

The intense Ba 3d peaks at approximately 780 eV and approximately 795 eV confirm the strong presence of Ba in the Tl-2223 superconductor. The Cu 2p doublet forms well-resolved peaks at approximately 933 eV and approximately 952 eV, clearly indicating the presence of oxidized Cu. The O 1s peak at approximately 531 eV is strong and well-resolved, whereas the C 1s signal near 285 eV is weak, indicating surface contamination. The carbon peak was further reduced after two minutes of Ar-ion etching. Furthermore, the 890–1005 eV region corresponds to the Ba and Cu Auger transitions and higher-BE core levels. The general features of the survey scan confirmed the expected elemental composition of the Tl-2223 high-temperature superconducting phase. Moreover, the scan did not indicate the presence of any additional impurities (other than C), especially from the Ag foil, which was used to wrap the sample to reduce Tl evaporation. Carbon appears to be a surface contaminant.

The atomic percentages of each element, determined XPS, are shown in Table 2. After normalizing the metal fraction (Tl + Ba + Ca + Cu = 41.08%), the atomic percentage of Tl is 6.06% (2.49/41.08). However, based on the formula unit $\text{Tl}_2\text{Ba}_2\text{Ca}_2\text{Cu}_3\text{O}_{10}$, the expected Tl fraction is 22.22% (2 Tl atoms / 9 total atoms). Therefore, the detected Tl content is approximately 27% of the expected value.

The Tl 4d spectrum (370–420 eV) exhibits a well-defined spin-orbit doublet, as shown in Fig. 2. Background subtraction was not required because the linear background affected only the intensity of the peak and did not affect the BE. The main low-BE peak is centered at $385.1 \pm 0.1 \text{ eV}$, which corresponds to Tl 4d_{5/2}, while the higher-BE component at $406.2 \pm 0.1 \text{ eV}$ corresponds to Tl 4d_{3/2}. The difference between these BEs is the spin-orbit splitting of approximately 21.1 eV, which is consistent with the

Table 1. Core level peaks for all elements in the sample

Element and core level	Energy range
Tl 4f	118–124 eV
Tl 4d	311–353 eV
Cu 2p	930–960 eV
O 1s	528–532 eV
C 1s (surface contamination)	284–286 eV
Sr 3d	approximately 130–138 eV
Ca 2p	approximately 346–355 eV

Table 2. Core level peaks for all elements in the sample

Element (peaks combined)	Atomic %
Tl (all Tl peaks combined)	2.49
Ba (all Ba peaks combined)	22.05
Ca (all Ca peaks combined)	8.47
Cu (all Cu peaks combined)	8.07
O (all O peaks combined)	58.90

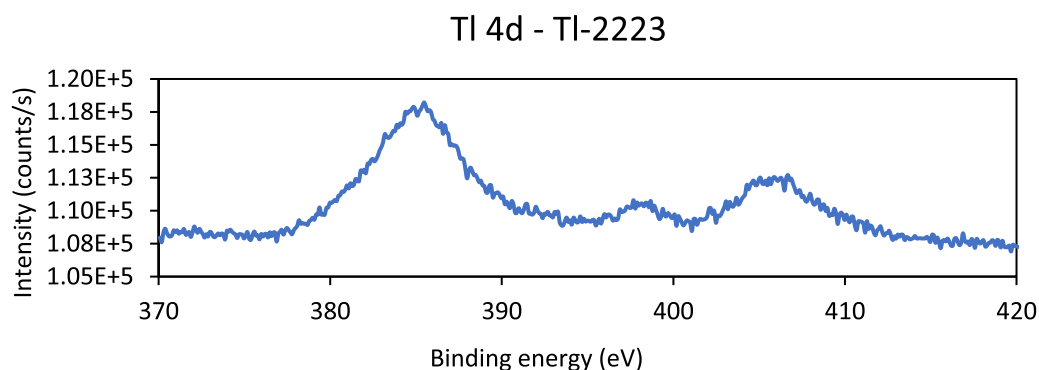


Figure 2. High resolution scan of the Tl-4d peak of the pure Tl-2223 high temperature superconductor

typical values reported for Tl in oxidized environments. The areas under the peaks were numerically integrated after background subtraction. The ratio of the two areas ($4d_{3/2}/(4d_{5/2}) \approx 0.45$, which is reasonably close to the expected statistical ratio given the overlap with the tails and residual background.

The presence of two peaks within the range (398–402 eV) suggests a shake-up or multiplet feature, rather than a separate chemical state. This further indicates that Tl is in a high oxidation state (Tl^{3+}) in the crystalline lattice, with no clear evidence of additional Tl chemical species or separate impurity phases in this energy region.

Ionic substitution at the Ba site in Tl-2223 superconductors changed electronic structure and oxidation states of various elements, which were reflected in the XPS spectrum of oxygen. Fig. 3 shows the O 1s peak with a linear background. The O 1s spectrum spans 520–540 eV, with a peak maximum at 530.8 ± 0.1 eV and full width at half maximum (FWHM) of approximately 2.26 eV. The peak is asymmetric, with the main peak and shoulder near 528 eV and near 530 eV, respectively. The strong narrow peak (approximately 2 eV) in width is characteristic of O^{2+} in the lattice structure and indicates that oxygen coordination in Tl–O, Ba–O, and Cu–O is typical of Tl-based high temperature superconductors.

The integrated areas under the peaks yield information on oxygen binding in the Tl-2223 superconductor. The ratio of the areas under both peaks was approximately 18%. Thus, a residual amount of surface-related oxygen or oxygen defects contributed to the O 1s envelope, even after etching the surface for 2 min with 3 Kev Ar-ions.

Table 3. Atomic percentages of the Ca-substituted sample

Element (peaks combined)	Atomic %
Tl (all Tl peaks combined)	1.35
Ca (all Ca peaks combined)	21.44
Cu (all Cu peaks combined)	23.44
O (all O peaks combined)	53.77

3.1.2. Substituted samples. The detected Tl contents in the Ca- and Mg-substituted samples were 1.35% and 1.61%, respectively, as shown in Tables 3 and 4. After normalizing the metal fractions of both samples ($Tl + Ca + Cu = 46.23\%$) for the Ca-substituted sample and ($Tl + Mg + Ca + Cu = 50\%$) for the Mg-substituted sample, the Tl atomic percentages were found to be 2.92% ($1.35/46.23$) and 3.22% ($1.61/50$), respectively. Thus, the detected Tl contents in the Ca- and Mg-substituted samples correspond to approximately 13.14% and 14.49% of their expected stoichiometric values, respectively.

The main O 1s peak occurs at approximately 531.3 eV with FWHM of approximately 3.47 eV, as shown in Fig. 4. This is typical of O^{2+} bound to mixed cations (Tl, Mg, Cu, Ca), which can be assigned to lattice oxygen in the crystalline structure of Tl-Mg-2223 high temperature superconductor, primarily in CuO₂ planes and Tl/Mg–O/Ca–O layers. Moreover, the broad peak (approximately 4.5 eV) indicates that oxygen may occupy several sites in the structure. The observed asymmetry in the O 1s peak is primarily attributed to the shoulder at a high BE. The ratio between the two areas was approximately 8%, indicating a minor contribution from oxygen adsorbed at the surface. After etching,

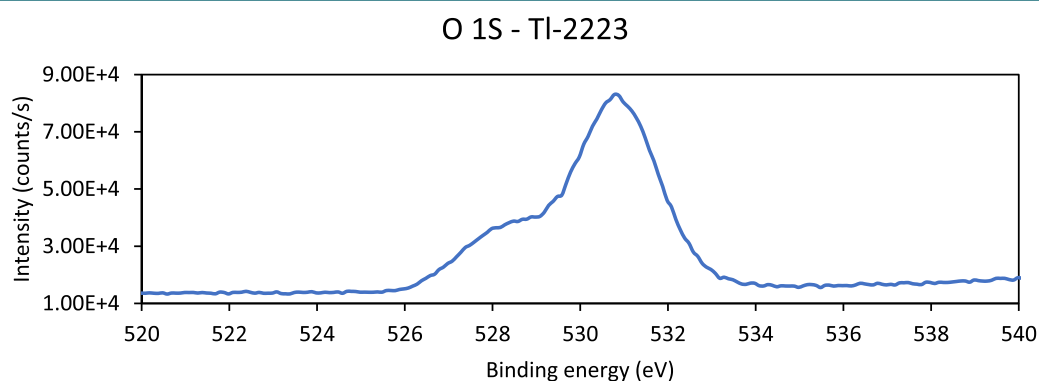


Figure 3. High resolution scan of the O 1s peak of the pure Tl-2223 high temperature superconductor

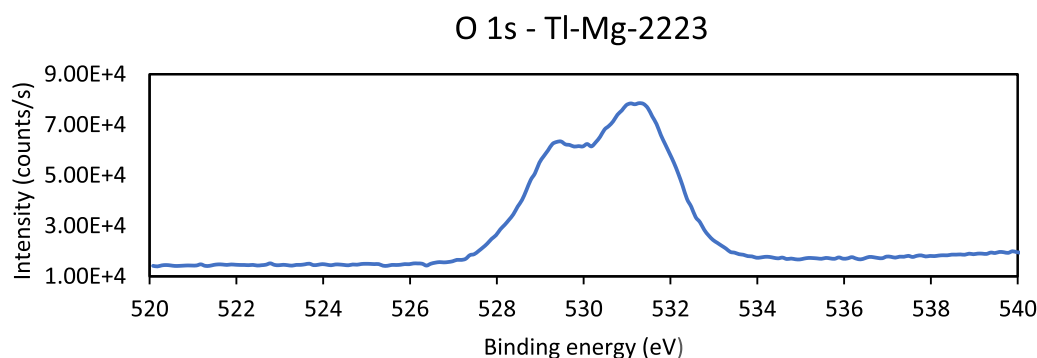


Figure 4. O1s peak in the Tl–Mg–2223 high-temperature superconductor

the observed low intensity indicates minimal oxygen contamination, defects, and oxygen non-stoichiometry in the Tl–Mg–2223 sample.

As shown in Fig. 5, the main O 1s peak is at approximately 528.9 eV with an FWHM of approximately 3.36 eV. This BE was slightly **lower** than the observed O 1S peak in the Tl–Ba/Mg sample, as discussed previously. The shift toward lower BE suggests that the O^{2+} anions in Tl–Ca–2223 experience stronger local screening and a higher electron density. This is consistent with the idea that Ca^{2+} layers are more directly coupled to the CuO_2 planes and do not introduce heavy alkaline-earth cations with large polarizabilities. Moreover, the broad FWHM is comparable to that of the Mg-substituted sample and reflects mixed oxygen environments and coordination (CuO_2 , Ca–O, Tl–O). The contribution to the high-energy peak was larger than that observed for the Ba-, Sr-, and Mg-substituted Tl-2223 samples. The ratio of the area of the main peak to the shoulder was approximately 70%, indicating that approximately 30% of the oxygen peak intensity originated from the high BE component. As the spectrum was obtained after Ar-ion etching, this high oxygen percentage

is unlikely to result from surface contamination. However, the low-BE (approximately **528.9 eV**) is attributed to O^{2-} ions in the lattice structure and dominated by CuO_2 -plane oxygen and Ca–O coordination. The higher BE (approximately 531 eV) may be due to oxygen defects or different chemical environments at the oxygen sites. This indicates that more etching is required to reduce the surface oxidation state and oxygen adsorption on the surface. Alternatively, these surface states may arise because CaO/ $CaCO_3$ tends to form as a secondary surface phase in the Ca-rich sample and contributes strongly to the observed high-BE O 1s.

Therefore, a comparison of the O 1S peaks in the Tl–Ba, Tl–Ca, and Tl–Mg–2223 XPS spectra showed that all the peaks exhibited main peaks and shoulders. The ratio of the areas under these peaks was related to the secondary oxygen phase, which occurred primarily because of surface contamination. The Tl–Mg–2223 sample revealed the cleanest O 1s peak, with only approximately 8% high-BE contribution (surface contamination). However, its main O 1s peak strongly resembled that of Ba, but with a broadened FWHM.

Table 4. Atomic percentages of the Mg-substituted sample

Element (peaks combined)	Atomic %
Tl (all Tl peaks combined)	1.61
Mg (all Mg peaks combined)	13.40
Ca (all Ca peaks combined)	17.92
Cu (all Cu peaks combined)	17.07
O (all O peaks combined)	51.65

3.1.3. Interpretation. Although XPS indicates a deviation from stoichiometry in all three samples owing to the reduced Tl atomic percentage, the XRD results suggest otherwise. This discrepancy may arise from the fundamental difference between surface-sensitive XPS and bulk-sensitive XRD, indicating that Tl loss is likely a surface-limited phenomenon that is potentially caused by oxidation or Tl volatility during synthesis and annealing.

3.2. XRD analysis.

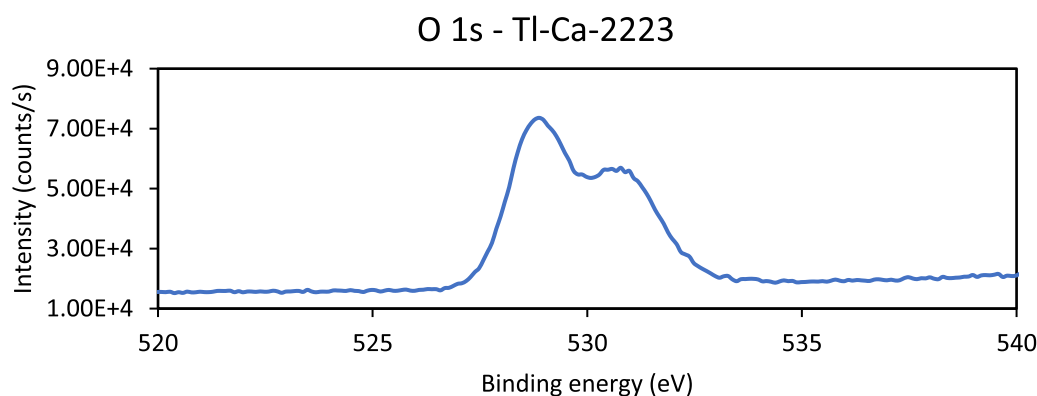


Figure 5. O1s peak in the Tl–Ca–2223 high-temperature superconductor

Table 5. Positions of peaks with interplanar spacing values

Peak No.	Ti-(2223)		Ti-Ca-(2223)		Ti-Mg-(2223)	
	2 θ (°)	d-(Å)	2 θ (°)	d-(Å)	2 θ (°)	d-(Å)
1	28.15394	3.16957	32.87177	2.72466	33.65339	2.66314
2	32.00918	2.79608	33.70044	2.65953	35.85105	2.50477
3	33.30085	2.69052	43.69815	2.07146	39.09591	2.30403
4	47.5574	1.91198	38.1832	2.35698	43.20174	2.0941
5	58.23464	1.58430	36.02202	2.49328	48.9946	1.8592
6	55.91145	1.64449	39.24621	2.29556	62.56407	1.48468
7	50.58918	1.80428	32.87041	2.72476	61.8543	1.499997

3.2.1. Lattice parameters. The interplanar spacings (d-spacings) of each sample were also calculated (Table 5). The Miller indices (hkl) were assigned based on crystallographic data available in the Materials Project database⁹ and applied to the two substituted samples to illustrate the structural effect of cation substitution. The lattice parameters were computed using Bragg's law as follows.

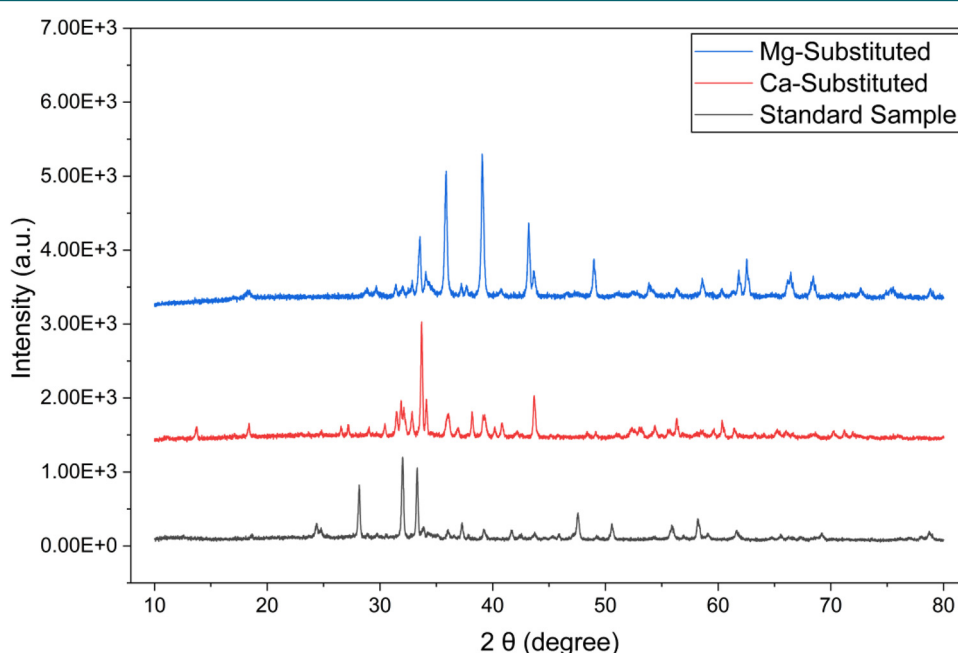
The obtained lattice parameters of the standard sample ($a = 3.85$ and $c = 36.7$ Å) are in good agreement with published data¹, indicating a well-formed crystalline Ti-2223 structure. For the Ca-substituted sample, we obtained $a=3.79$ and $c=37.07$ Å, exhibiting a decrease of 1.56% and an increase of 1.01% in the a - and c -parameters, respectively. For the Mg-substituted sample, we obtained, $a=3.76$ and $c=36.95$ Å, which were 2.34% smaller and 0.68% higher, respectively, than the standard Ti-2223 values. These data indicate that the “ a ” lattice-parameter consistently decreases with the ionic radius, indicating a directly proportional relationship between it and the atomic size. However, no clear trend was observed between the ionic radius and c -parameter.

3.2.2. Crystallite size. The crystallite size (D) of each sample was calculated using the Scherrer equation. The standard

sample exhibited an average crystallite size of 34.726 nm, whereas the Ca- and Mg-substituted samples showed values of 26.504 and 27.202 nm, respectively. The results demonstrated a general reduction in the crystallite size with decreasing ionic radius. However, the trend was not strictly linear, as the Mg-substituted sample possessed a slightly larger crystallite size than the Ca-substituted sample, despite its smaller ionic radius.

3.2.3. Microstrain. The ionic substitutions often lead to induced stress-strain in the crystals. The Williamson-Hall method has been used to evaluate the induced microstrain and the crystallite size in all prepared samples. The standard sample exhibited the lowest average strain (2.865), whereas the Ca-substituted sample exhibited a significant increase (9.266). The Mg-substituted sample exhibited an intermediate strain value of 5.047. Considering the direct relationship between strain and stress, the Ca-substituted sample experienced the highest internal stress, leading to the largest negative chemical pressure, which may have enhanced the transition temperature.

3.2.4. Dislocation density. The dislocation density of each sample was determined using the reciprocal of the square

**Figure 6.** XRD spectra of all samples

of the crystallite size. The standard sample exhibited an average dislocation density of 0.891 nm^{-2} . This value increased to 4.172 and 11.731 nm^{-2} in the Mg- and Ca-substituted samples, respectively, which is expected to enhance flux pinning and improve magnetic levitation performance in the superconducting state. Although the dislocation density did not vary linearly with the ionic radius, the results indicated a general inverse trend between the ionic size and dislocation density.

3.2.5. Peak shifts. According to Bragg's law, peaks shift to higher 2θ values as the d-spacing decreases, which is clearly observed in Fig. 6. The Miller index (112) peak was used for comparison. The standard sample exhibits a peak at 33.301° . In the Mg-substituted sample, the (112) peak shifted to 33.653° , representing an increase of 1.06%. Despite having a larger ionic radius than that of Mg, the Ca-substituted sample showed a further shift to 33.700° , corresponding to an increase of 1.2%.

Overall, the results suggest that the peak positions shift toward higher 2θ values with decreasing ionic radius, despite the slight non-linear behavior.

4. LIMITATIONS

This study focused on the structural characterization of Ba-site substitution in Tl-2223 superconductors. Superconducting properties such as critical temperature (T_c) and critical current density (J_c) were not measured. In addition, because XPS is a surface-sensitive technique, the observed Tl deficiency may primarily represent surface effects rather than the bulk composition. Further studies are required to correlate these structural changes with superconducting performance.

5. CONCLUSIONS

This study examined the effect of atomic-size substitution on Tl-2223 superconductors using XRD and XPS analyses. Although the XPS results showed significant surface Tl deficiency, XRD confirmed the preservation of the Tl-2223 phase, suggesting that Tl loss was likely restricted to the surface due to volatility or oxidation.

Substituting Ba^{2+} with smaller cations (Ca^{2+} and Mg^{2+}) produced noticeable structural modifications, including a consistent reduction in the a lattice parameter and a slight increase in the c lattice parameter. The crystallite size generally decreased with decreasing ionic radius, whereas both the microstrain and dislocation density increased, especially in the Ca-substituted sample, indicating a higher internal stress and enhanced lattice distortion. Peak shifts toward higher 2θ values further confirmed the reduction in interplanar spacing upon substitution.

Overall, atomic size substitution clearly altered the structural properties of Tl-2223, and these structural changes may influence the superconducting behavior. Future work should focus on measuring the superconducting properties (such as T_c , J_c , and magnetic performance) to fully correlate the chemical pressure effects with superconductivity.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Hawra Tuhaifa – Department of Materials Science and Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0009-0001-5566-6112
Email: s202263400@kfupm.edu.sa

Corresponding Author

Khalil A. Ziq – Research Mentor, Professor, Department of Physics, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0000-0002-5061-7004
Email: kaziq@kfupm.edu.sa

ACKNOWLEDGMENTS

The authors acknowledge the support provided by the Department of Materials Science and Engineering, and the Department of Physics at KFUPM. The authors also thank the technical staff surface science laboratory team for their assistance with specimen preparation and testing.

REFERENCES

- (1) Hassan, B., Alnakhilani, A. & Abdulhafiz, M. Investigating the effects of high magnetic fields on the phase stability of $\text{Tl}_2\text{Ba}_2\text{Ca}_2\text{Cu}_3\text{O}_{10-\delta}$ superconductor. *J. Phys. Sci.* 35, 33–44 (2024).
- (2) Naseem, K., Khan, N. A. & Safeer, S. H. Effect of magnesium doping to reduce the charge reservoir layer in $\text{Cu}_{0.5}\text{Tl}_{0.5}(\text{Ba}_{2-x}\text{Mg}_x)\text{Ca}_2\text{Cu}_3\text{O}_y$ ($x = 0, 0.15, 0.25, 0.35$) superconductors. *J. Electron. Mater.* 50, 2164–2170 (2021).
- (3) Ali, A. H., Ali, A. K. D. & Jasim, K. A. Study the partial substitution for Sr at the Ba on the properties of $\text{Tl}_{1.6}\text{Hg}_{0.4}\text{Ba}_{2-y}\text{Sr}_y\text{Ca}_2\text{Cu}_3\text{O}_{10\delta 5}$ superconductors. *IOP Conf. Ser. Mater. Sci. Eng.* 871, 012079 (2020).
- (4) Al-Rawi, S. S., Razig, K. H. & Baqi, A. S. Effect of partial substitution of Ag and Sr on the structural properties of $\text{Tl}_{2-x}\text{Ag}_x\text{Ba}_{2-y}\text{Sr}_y\text{Ca}_2\text{Cu}_3\text{O}_{10+\delta}$ high-temperature superconductor. *J. Coll. Basic Educ.* 19, 685–693 (2022).
- (5) Alnakhilani, A., Hassan, B., Ahlam, M. A. & Abdulhafiz, M. Effect of preparation and annealing temperature on the properties of (Hg,Tl) -2223 superconductor. *J. Phys. Sci.* 30, 71–79 (2019).
- (6) R.E. Gladyshevskii, Ph. Galez, K. Lebbou, J. Allemand, R. Abraham, M. Couach, R. Flükiger, J.-L. Jorda, M. Th. Cohen-Adad. Structural characterization and superconducting properties of $(\text{Tl}_{0.5}\text{Pb}_{0.5})(\text{Sr}_{2-x}\text{Ba}_x)\text{Ca}_2\text{Cu}_3\text{O}_{9-\delta}$. *Physica C* 267, 93–105 (1996).
- (7) Ph. Galez, Th. Hopfinger, J.L. Soubeyroux, M. Lomello-Tafin, Ch. Opagiste, Ch. Bertrand, J.L. Jorda. The reaction pathway for the formation of Tl-2223. *Physica C* 372–376, 1137–1140 (2002).
- (8) D.N.G. Krishna, J. Philip, “Review on surface-characterization applications of X-ray photoelectron spectroscopy (XPS): recent developments and challenges”, *Appl. Surf. Sci. Adv.*, 12 (2022), Article 100332, 10.1016/j.apsadv.2022.100332.
- (9) Materials Project Database, Lawrence Berkeley National Laboratory. Available at: <https://materialsproject.org>.



***Journal of Undergraduate
Research International***