KFUPM
Journal of Undergraduate Research International

Diamond Open Access

Article *Computer Science*

KF-2025: Deep Learning-Based Classification of Canteen Food Trays Using Custom-Collected Dataset

Saad Obada Alshami and Ali Nasir*

Cite https://doi.org/10.64589/juri/214613

Submitted: May 31, 2025 **Revised:** November 06, 2025 **Accepted:** November 23, 2025

ABSTRACT

University canteens serve large student populations within short time windows, necessitating efficient and automated food service solutions. This study presents KF-2025, the first real-world, tray-based food image dataset collected directly from the King Fahd University of Petroleum and Minerals (KFUPM) students' canteen to facilitate automated self-checkout and smart cafeteria applications. The dataset comprises 3,222 manually annotated images across 26 food classes, derived from over 1,000 tray recordings under real canteen conditions. Four transfer learning models, i.e., ResNet-50, ResNet-152, MobileNetV2, and EfficientNet-B0 were trained and evaluated. The best performance was achieved by ResNet-152 using the Adam optimizer, attaining 97.25% accuracy, 97.62% precision, 93.78% recall, and 94.53% F1-score. The high performance across all models confirms the robustness and practical utility of KF-2025 for real-world deployment, including on resource-constrained devices. Beyond its immediate application to canteen automation, KF-2025 provides a foundation for advancing context-aware food recognition, intelligent campus systems, and data-driven nutritional monitoring, thereby contributing to the broader development of smart and sustainable institutional environments.

Keywords: food recognition, deep learning, image classification, ResNet, efficientNet, transfer learning, custom dataset, smart canteen

1. INTRODUCTION

As Artificial Intelligence and deep learning evolve rapidly, several domains of daily life are becoming subjects of research. For example, industrial productivity [1], transportation [2], and banking [3] have already undergone transformations due to deep learning. One promising area for deep learning application is food consumption, which is an essential part of daily human activities. Accordingly, food classification powered by deep learning models has become a prominent research area to allow the detection and classification of food types, offering practical solutions to improve daily lives. For instance, food classification can help reduce waiting times in restaurants and canteens by enabling self-checkout, and can help estimate caloric intake.

However, food classification faces several challenges, including overlapping food categories [4], complex appearance of food items (where even identical ingredients may vary in shape and color) [5], and differences between the environments in which food images are captured and those in which they are recognized [6]. Such factors can significantly hinder model accuracy and classification performance.

University canteens are among the most complex and challenging environments for food recognition, where food appearance in terms of shape, size, and color is inconsistent. The same dish can contain several ingredients, and identical food classes may differ significantly from one canteen to another, limiting the utility of other food datasets. In addition, many food classes are missing from available datasets. At King Fahd University of Petroleum and Minerals (KFUPM), thousands of students visit the canteen daily, particularly during breaks, leading to long waiting times. Therefore, automating the checkout process could enhance students' experiences by saving time and improving convenience.

To address this need, this study aims to construct the first real-time tray-based food image dataset, KF-2025, collected from KFUPM's student canteen. This dataset covers most of the food classes presented daily in the cafeteria. Furthermore, this study provides experimental results when training and testing multiple AI classification models on this dataset after preprocessing and annotating, contributing to a complete food recognition pipeline that can be deployed in the future for automated self-checkout systems.

2. RELATED WORK

In the field of food recognition, advancing artificial intelligence (AI) models has garnered attention due to their potential impact on daily life. A typical pipeline for developing food recognition systems involves constructing or utilizing existing food datasets, preprocessing and annotating images, and training models on those datasets to assess the performance. Food classification is

Published under an exclusive license by KFUPM

Figure 1. Plov from CAFD (left) and Rice from UEC-FoodPix (right). Despite both being rice-based, their appearance differs greatly, illustrating the domain mismatch across cuisines

highly influenced by the quality of the dataset and the techniques used to annotate it.

Some well-known food recognition datasets include Food-101[7], which contains 101,000 images across 101 food categories; UEC-FoodPix Complete[8], which contains 10,000 images; and UNIMIB2016[9]. These datasets were collected under varying conditions. Food-101 was captured in a controlled laboratory environment, whereas UNIMIB2016 was captured in a real-world university canteen, reflecting different environmental settings. Additionally, food datasets vary in terms of food classes and regional cuisine, such as the CAFD dataset[10], which focuses on Central Asian food, and the UEC-FoodPix dataset, which focuses on Eastern Asian food. Although both datasets represent Asian food, they are substantially different; they share no common food classes and similar base classes differ significantly in appearance, as illustrated in Figure 1. For example, while both "Rice" (from UEC-FoodPix) and "Plov" (from CAFD) are rice-based dishes, their presentation, accompaniments, and overall appearance differ significantly. This highlights that using food datasets from different cuisines can result in major mismatches when applied to a local context.

Two notable image datasets have been proposed for Middle Eastern food: the MEFood and EMFID platform databases. MEFood is a large-scale dataset covering 38 categories, with images collected from search engines, such as Google, and social media platforms, such as Instagram, to increase representativeness[11]. The EMFID provides a comprehensive collection of food

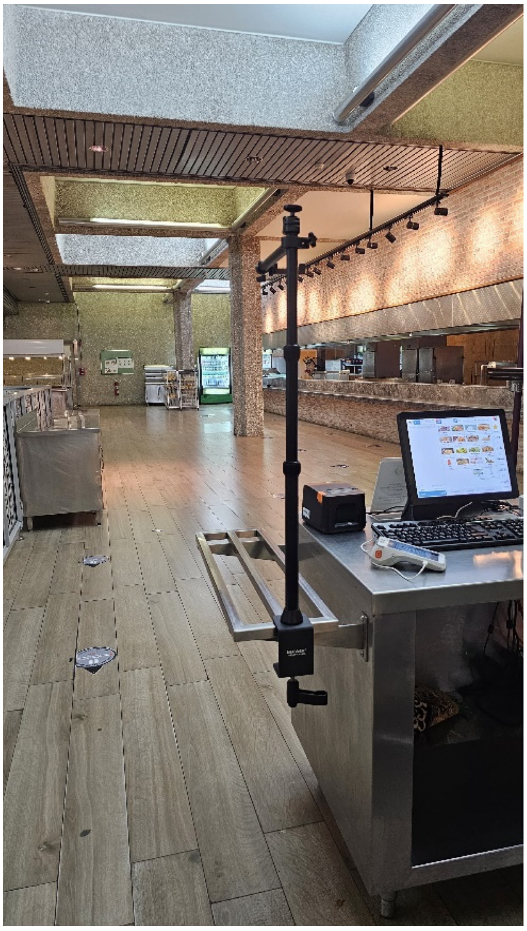

Figure 2. Camera setup used to capture tray images

Figure 3. Sample tray image with multiple food items

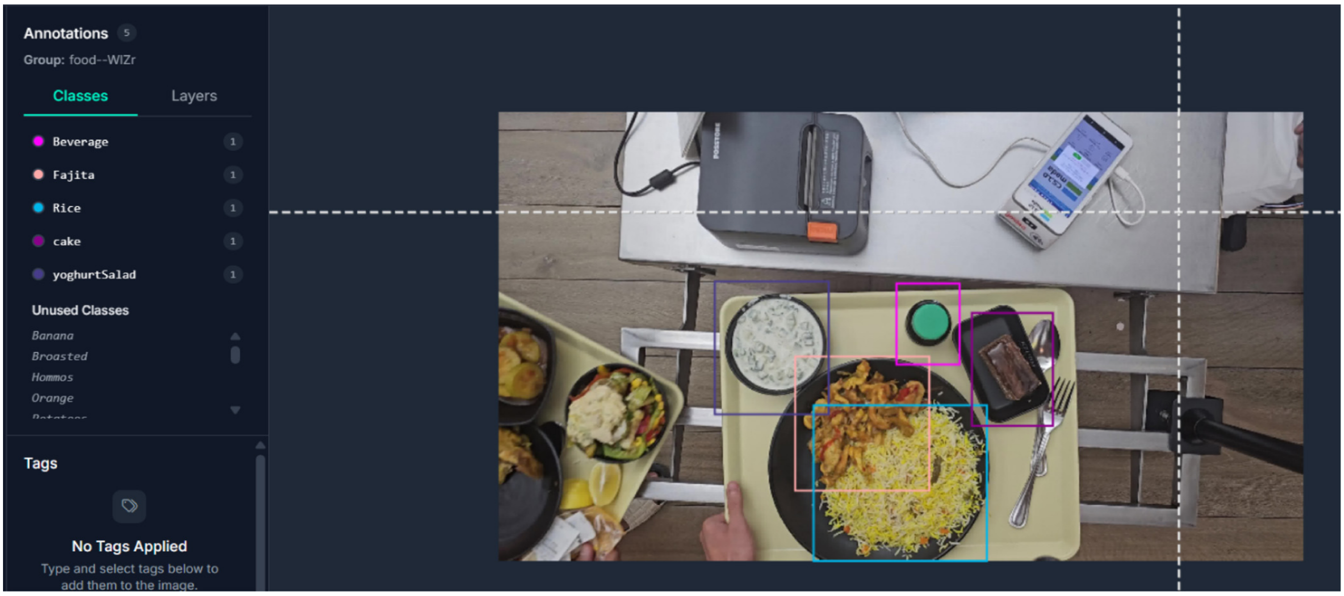

Figure 4. Bounding boxes used via Roboflow to define the boundaries of each food class

samples from Eastern Mediterranean countries for educational and research purposes [12]. While both datasets include a variety of regional food classes, their images were not captured in real-life environments such as canteens, which reduces their suitability for deployment in practical tray-based applications.

In addition to the dataset content, the annotation strategy plays a crucial role in the performance of food recognition systems. Several techniques have been adopted to annotate food images, each offering different levels of detail and computational requirements. For instance, in Food-101, images are grouped by class, with each image placed in a folder that corresponds to its category. Although this method is simple, it limits the granularity of information in each image. In contrast, semantic segmentation has been applied in some studies [9], where the food content in an image is isolated by separating it from the background. This method allows for more detailed labeling but is computationally expensive. In addition, the use of bounding boxes to isolate individual food items in trays, where food items were marked within trays using the Roboflow platform [10], enables the recognition of multiple food items in a single image, an essential feature of tray-based food classification.

Furthermore, manual annotation has been shown to outperform automatic and AI-assisted annotations in terms of accuracy. For instance, in the UEC-FoodPix Complete dataset, the model trained on fully manual annotations achieved higher accuracy (0.668) than the model trained with automatic labeling (0.560) [8]. This demonstrates the importance of accurate and reliable annotations, especially in complex environments, such as

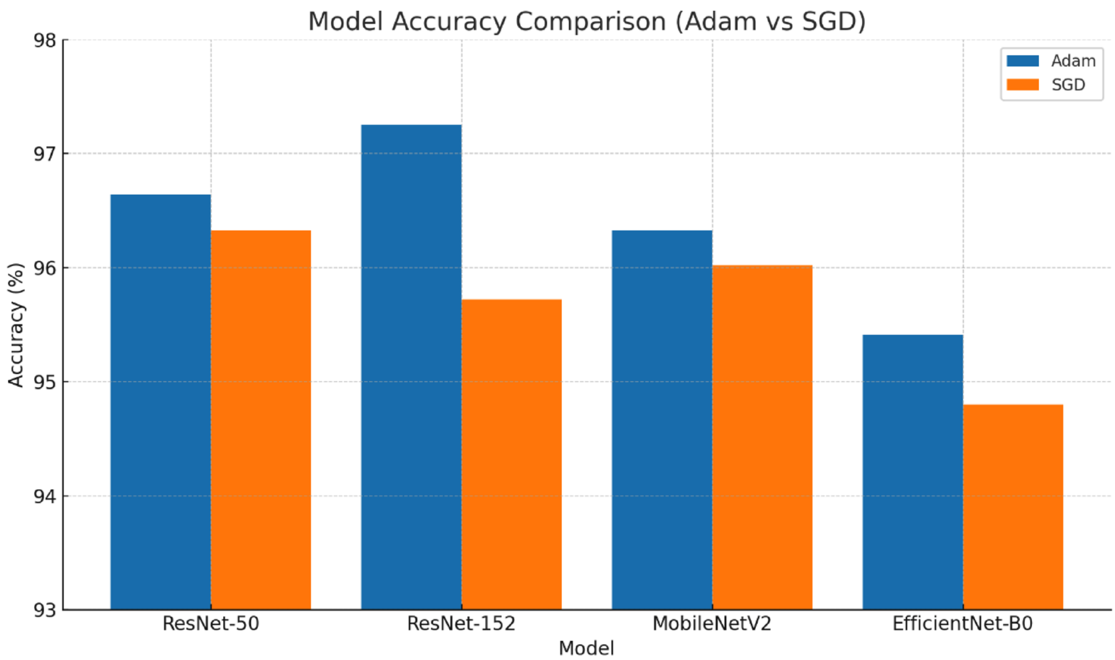

Figure 5. Model accuracies on KF-2025 using Adam and SGD

Table 1. *Evaluation metrics of the models trained on the KF-2025 dataset using SGD and Adam optimizers*

Model	Optimizer	Learning Rate	Accuracy	Precision	Recall	F1-score
ResNet-50	Adam	0.001	96.64%	95.47%	93.74%	93.33%
ResNet-50	SGD	0.001	96.33%	92.35%	91.02%	89.74%
ResNet-152	Adam	0.001	97.25%	97.62%	93.78%	94.53%
ResNet-152	SGD	0.001	95.72%	94.45%	90.05%	89.76%
MobileNetV2	Adam	0.001	96.33%	96.01%	91.20%	92.09%
MobileNetV2	SGD	0.001	96.02%	90.25%	90.16%	89.37%
EfficientNet-B0	Adam	0.001	95.41%	92.40%	91.13%	89.74%
EfficientNet-B0	SGD	0.001	94.80%	95.43%	88.34%	89.08%

university canteens, where food items often overlap and vary in appearance.

Finally, the choice of models and training techniques significantly affects the performance of food recognition systems. For example, early studies that used the Food-101 dataset achieved an accuracy of approximately 50.76%[7]. However, later research utilizing more advanced models and transfer learning techniques achieved accuracies of 80%[13] and 93.26 %, respectively[14]. More recent studies have employed pretrained architectures, such as Residual Networks (ResNet) and EfficientNet, further enhancing food recognition accuracy, reaching 97.54%[15]. These improvements highlight the continuous advancements in the field and the importance of model selection for achieving higher performance.

Earlier studies contained a greater variety of models reporting high metric results. For instance, DenseNet-246 achieved a top-1 accuracy on ImageNet[16]. ResNeXt-101 achieved an accuracy of 79.6 on the ImageNet further[17]. The vision transformer achieved 88.6% accuracy when pretrained on a JFT-300M[18]. However, these models are computationally heavy compared to other networks such as ResNet and MobileNet, making them less optimal for online applications when computational ability is constrained.

3. METHODOLOGY

3.1. Data Collection.

In the KFUPM student restaurant, there are four checkout points where students pay for meals. At one of these points, a phone (Samsung Galaxy S23 Ultra) was mounted on a stand to capture videos (4K resolution, 30 FPS) of the trays from a top-down perspective, ensuring that all items on the tray were visible and clearly detectable. The camera setup is shown in Figure 2.

Using the FFmpeg tool, the recorded videos were decomposed into individual frames. After discarding frames in which the tray contents were obscured by people or suffered from noise and motion blur, 1,039 usable images were extracted. Given that each tray typically includes multiple food components (e.g., rice, dessert, and beverages), as illustrated in Figure 3, it was essential to annotate each item individually to enable more precise downstream processing.

3.2. Annotation.

The next step was to upload the selected images to Roboflow, a platform that provides all the necessary tools for building and deploying computer vision models, including annotation and segmentation. Many KF-2025 images contain more than one type of food on a single plate, making food classification based on whole dishes impractical. Therefore, bounding boxes were used to isolate each food item. In this step, 711 images were manually annotated into 28 food classes using bounding boxes, as illustrated in Figure 4.

3.3. Preprocessing.

The annotated 711 images were randomly split into 563 training, 74 validation, and 74 test images, comprising approximately 80% for training, 10% for validation, and 10% for testing. Subsequently, the annotations were exported in YOLOv8 format. Using the bounding box coordinates, each food item was cropped based on its class labels. For example, all cropped regions labeled as “cake” were grouped under the “cake” class.

Owing to the varying availability of food items in the canteen, there was an inherent imbalance in the dataset. Commonly served items such as beverages, rice, and yoghurt had the most images, with the beverage class containing the most cropped images at 458. In contrast, rarely served items had fewer images, with burgers having the fewest at 6. Due to the extremely low number of samples in the ‘burger’ and ‘coleslaw’ classes, they were removed from the dataset. This resulted in a final dataset of 26 classes and 3,222 cropped images, including training, validation, and test sets.

3.4. Trained Models.

Three transfer-learning models were selected to represent different levels of architectural complexity: MobileNetV2 and EfficientNet-B0 (lightweight), ResNet-50 (moderate), and ResNet-152 (complex).

3.4.1. ResNet (50 and 152).

ResNets are convolutional neural networks (CNNs) pretrained on over a million images and are capable of classifying them into 1,000 object categories. ResNet-50 contains 50 layers (48 convolutional layers, one maximum pooling layer, and one average pooling layer)[19] while ResNet-152 consists of 152 layers[20].

3.4.2. MobileNetV2.

A lightweight CNN with 28 layers introduced by Google in 2017 was designed for mobile and embedded devices[21]. It has significantly fewer parameters than other CNNs while maintaining competitive accuracy[22].

3.4.3. EfficientNet-B0.

EfficientNet is a CNN developed using a multi-objective neural architecture search to optimize both accuracy and computational efficiency (FLOPS). The EfficientNet family consists of eight models (B0–B7), where the number of parameters increases with the model complexity[23].

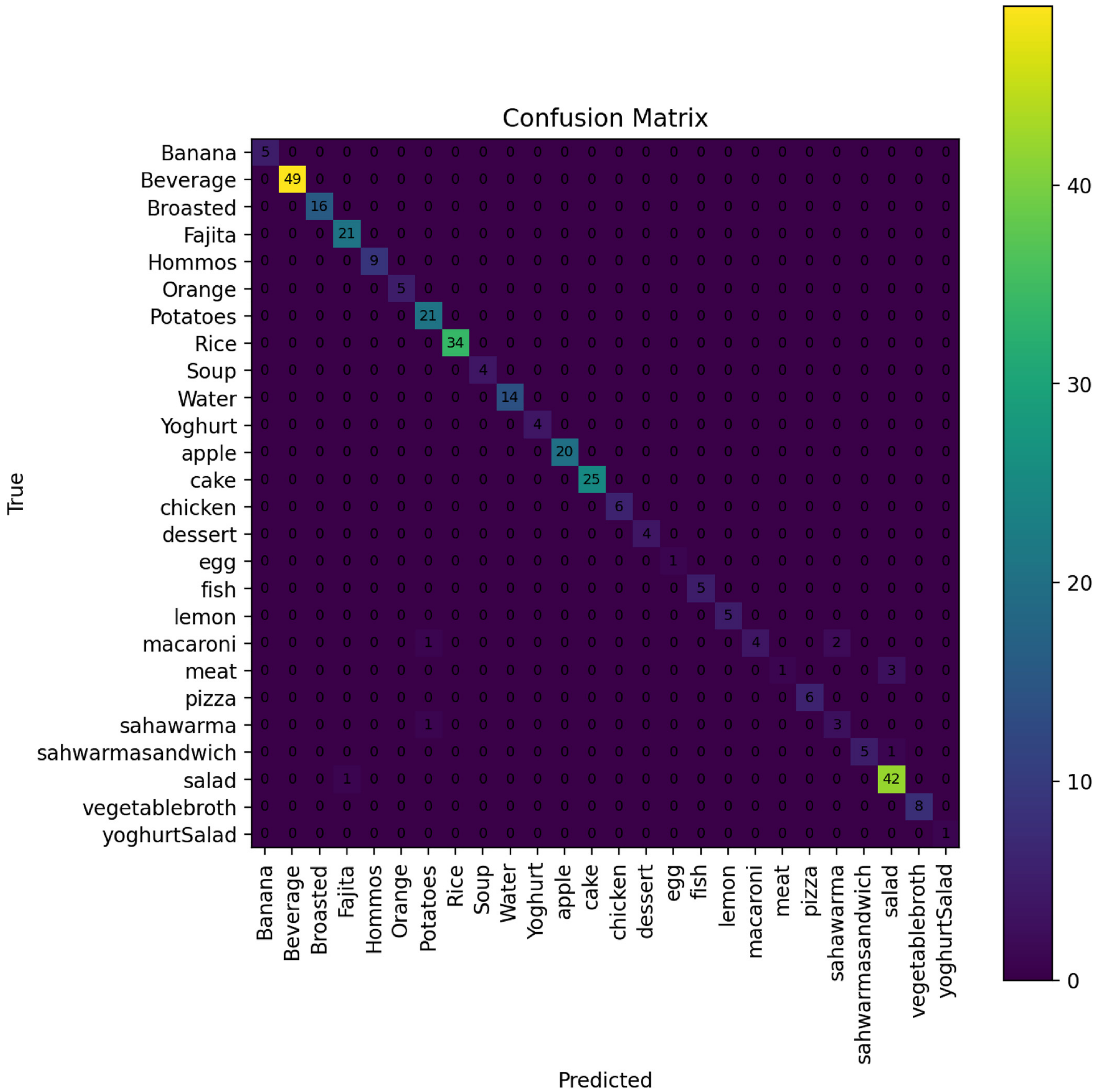

Figure 6. Confusion matrix of ResNet-152 with the Adam optimizer (the best-performing model). The majority of test samples were correctly classified, with some exceptions noted in the Salad and Shawarma classes

3.5. Training Setup. Setting the learning rate to 0.001 and batch size to 32, SGD and Adam optimizers were used to train the models on the KF-2025 dataset. Adam was selected for its fast convergence and minimal tuning needs[24], which are suitable for smaller datasets, whereas SGD was included for its strong generalization with sufficient tuning[25]. The models were trained using 15 epochs of fine-tuning, and the entire model was trained. The trained models were then evaluated using accuracy, precision, recall, and F1-score to assess both model performance and the overall quality of the dataset.

Although the accuracy metric reflects the overall accuracy of the model in recognizing food classes, it can be misleading in imbalanced datasets, where certain classes dominate. Therefore, precision and recall were also considered to ensure that the model was not biased toward the more frequent classes while neglecting the rare ones. Finally, the F1-score provides a balanced measure by combining precision and recall, thereby offering a more inclusive view of the model's performance across all classes.

4. RESULTS

The experimental results are shown in Table 1 and Figure 5.

In addition to the test metric results, it was important to examine the confusion matrices of the experiments to identify the classes that were more frequently misclassified. The confusion matrices for the three selected experiments representing the different models and optimizers are provided in Figures 6 to 9.

Given that the KF-2025 dataset was newly introduced, it is important to compare its results with established food classification benchmarks such as Food-101. To ensure fairness, the same models and optimizers were trained on both datasets for five epochs, with consistent hyperparameters. Table 2 presents the best-performing model-optimizer combinations for each dataset based on accuracy, precision, recall, and F1-score.

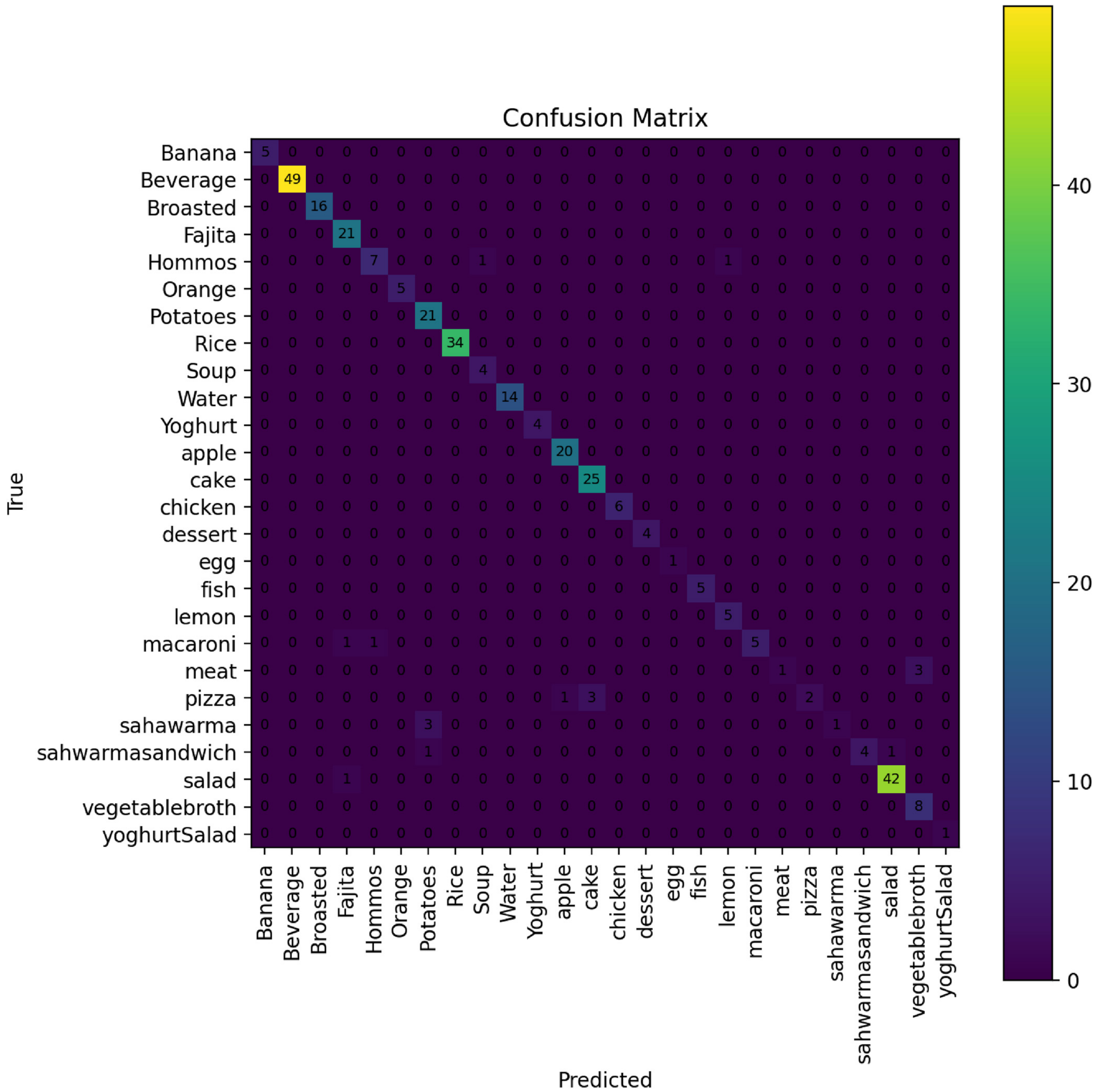

Figure 7. Confusion matrix of MobileNetV2 using the SGD optimizer. The matrix shows notable misclassification of Shawarma as Potatoes multiple times

5. DISCUSSION

As shown in Table 1, for the same optimizer type, ResNet architectures consistently outperformed the other models. This can be attributed to the use of residual blocks with skip connections, which enables deeper networks to learn more complex and abstract features effectively[26]. Among the models tested, ResNet-152 achieved the highest performance when paired with the Adam optimizer, likely owing to its increased depth and greater number of trainable parameters compared with ResNet-50. However, ResNet models are more computationally intensive than lightweight architectures such as MobileNetV2 and EfficientNetB0, which limits their deployment on devices with limited processing power. Nonetheless, the fact that the lowest accuracy recorded across all models was 94.80% suggests that even lightweight models can achieve a strong performance when efficiency is a priority.

Regarding the optimizer choice, the models trained with Adam generally achieved a higher accuracy than those trained with

Table 2. *Performance comparison between Food-101 and KF-2025 datasets using their best-performing model and optimizer configurations over five training epochs*

Dataset	Model	Optimizer	LR	Accuracy	Precision	Recall	F1-score
Food-101	ResNet-152	SGD	0.001	80.05%	80.50%	80.05%	80.08%
KF-2025	MobileNetV2	Adam	0.001	94.46%	92.03%	92.03%	92.21%

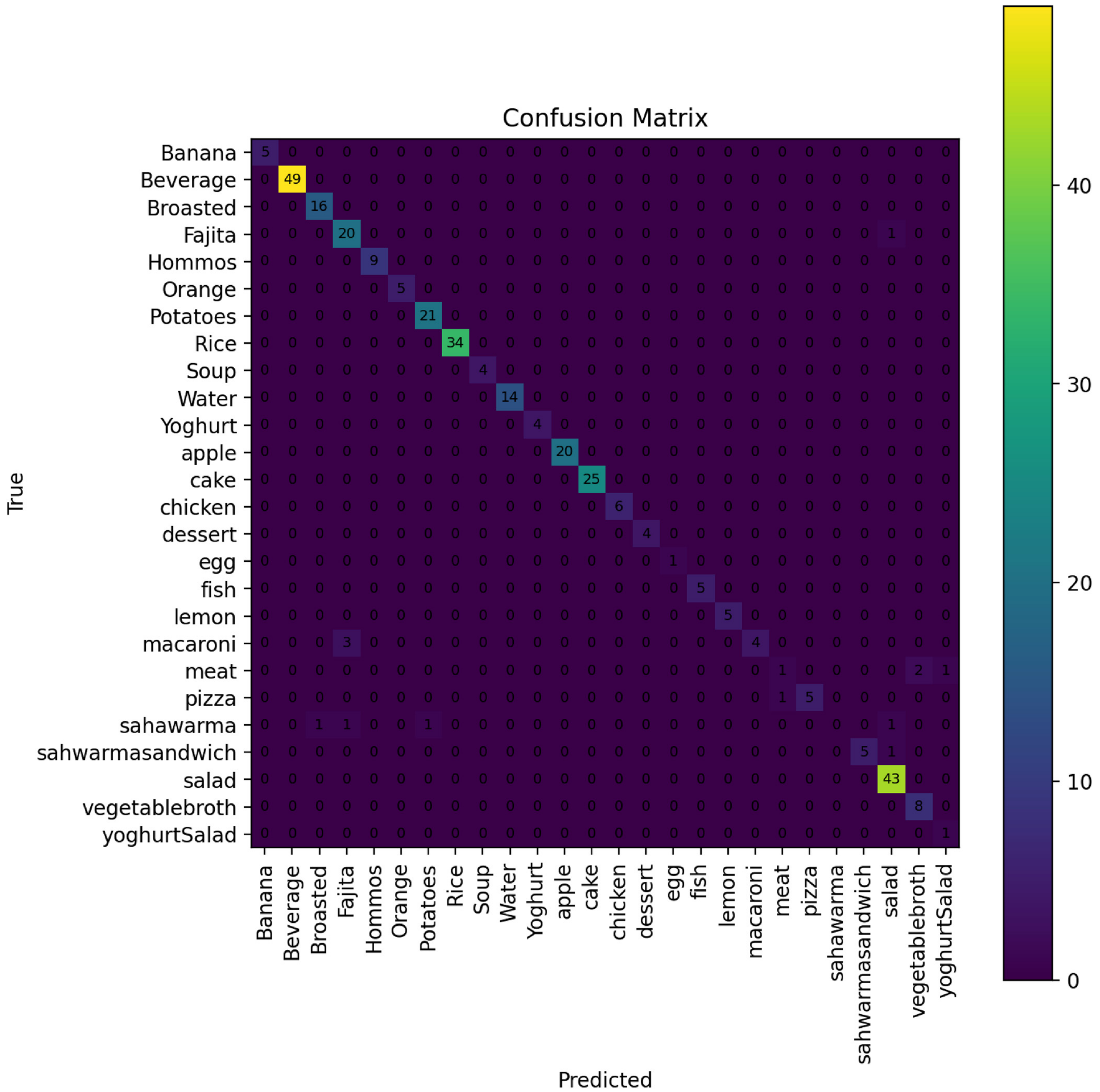

Figure 8. Confusion matrix of EfficientNet-B0 with the Adam optimizer. In addition to misclassification of the Shawarma class, the matrix shows multiple instances of Macaroni being misclassified as Fajita

SGD. This may be attributed to Adam's adaptive learning mechanism and hyperparameters, which typically require minimal tuning[24], making it well suited for scenarios involving relatively small training datasets, as in this study.

Although the precision, recall, and F1-scores were generally high, their lower values compared with the accuracy suggest that the models may favor the majority classes, which is expected given the class imbalance in the dataset. Reducing this imbalance by ensuring more uniform class distributions could help improve these metrics and bring them closer to overall accuracy.

As shown in the confusion matrices (Figures 6, 7, 8, and 9), most classes were correctly identified. However, some misclassifications occurred, likely owing to similarities in ingredients and the visual appearance of certain food items served in the canteen. For instance, in Figure 7, the model misclassified the "Shawarma" class as "Potatoes" three times. This confusion is likely due to the fact that shawarma is often served with fries in the same dish, leading to overlapping visual features. Figure 8 illustrates this point, showing how both classes appear visually similar when served together.

Additionally, as shown in the confusion matrices, particularly in Figure 9, and in the misclassified samples illustrated in Figure 11, images labeled as "Macaroni" were frequently predicted as belonging to other classes. This issue arises from the notable intra-class variability within the Macaroni category, in which plates prepared with different sauces and ingredients are grouped under the same label. Introducing more fine-grained class definitions, for example, by separating macaroni dishes according to sauce type or main ingredients, could help the model capture distinctive features more effectively and improve the classification accuracy.

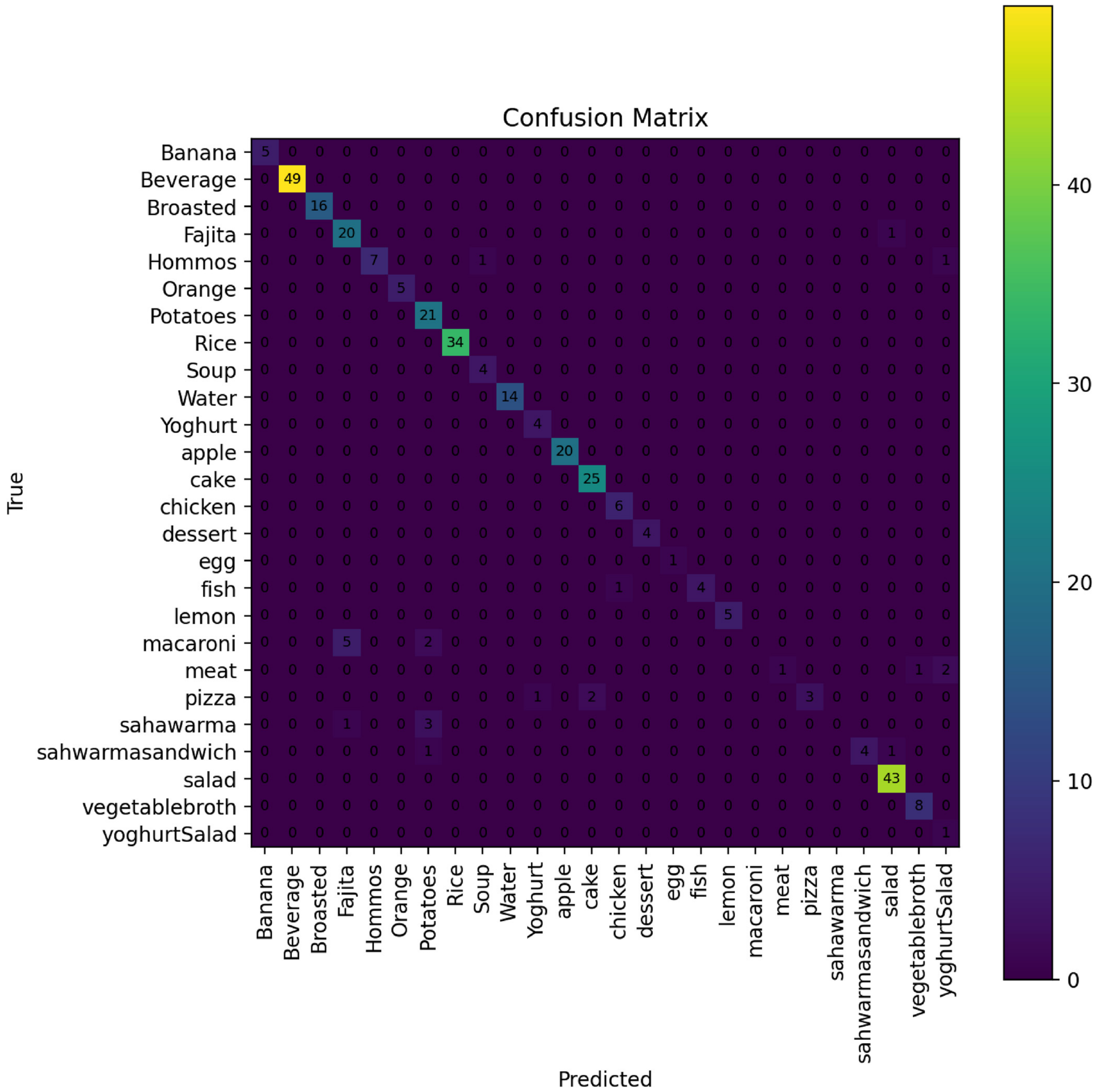

Figure 9. Confusion matrix of ResNet-50 with the SGD optimizer. The confusion matrix confirms the misclassification problem with the Macaroni and Shawarma classes, given the relatively high number of errors

Table 2 presents a performance comparison between Food-101 and KF-2025 under the same training conditions and models. The superior performance of KF-2025 can be attributed to several factors. First, KF-2025 includes fewer classes, which reduces overall classification complexity. Second, Food-101 contains images collected from diverse internet sources, introducing high intraclass variability in visual appearance. Therefore, effective feature extraction for Food-101 requires extensive training.

Furthermore, the close match between accuracy and F1-score in Food-101 suggests a relatively balanced class distribution, whereas the KF-2025 dataset suffers from a class imbalance. This imbalance likely contributes to the performance gap between the accuracy and F1-score in KF-2025. Thus, increasing the number of samples in underrepresented classes may improve model fairness and yield more consistent metrics, similar to the results observed for Food-101.

Despite its small size and imbalance, the strong results of the KF-2025 highlight its practical effectiveness for real-world canteen classification tasks, especially in settings with a limited number of food categories.

5.1. Limitations and Future Work. Despite the promising results, the current dataset includes only 26 food classes, which do not fully represent the complete variety of meals served in student restaurants. Some food types were excluded due to the limited number of available samples, which may have affected the overall inclusiveness of the dataset. Additionally, the dataset suffers from class imbalance, where certain food categories are overrepresented compared to others. This imbalance may have influenced the model's performance, particularly in underrepresented classes. Furthermore, techniques that typically improve model generalization, such as data augmentation, were not applied in this study. To address these limitations, future work will involve

Figure 10. Examples showing visual similarity between Shawarma (left) and Potatoes (right), which may contribute to misclassification

collecting additional images across different days of the week to cover full menu variations, using polygon-based annotations for more accurate labeling, applying data augmentation techniques to enhance model robustness, and balancing class distributions during data collection.

A key focus of future work should be the deployment of trained models and associated challenges. While ResNet models can be deployed on embedded platforms, they typically require computationally advanced devices, such as Jetson Xavier, to achieve adequate performance[27]. In actual canteen settings, deploying multiple units at different checkout points can significantly increase costs, and installation may be constrained by limited space. In contrast, lightweight architectures, such as MobileNetV2 and EfficientNet-B0, can be deployed on smaller and more cost-effective devices, such as Raspberry Pi[28], although they generally report slightly lower accuracy. Given the importance of efficient and real-time recognition processes, high-accuracy heavy models with powerful hardware versus lightweight models with affordable devices should be evaluated to determine the most practical setup for canteen environments.

6. CONCLUSIONS

This study presented KF-2025, the first real-world, tray-based food image dataset collected directly from the students' canteen at the King Fahd University of Petroleum and Minerals. KF-2025 captures real, context-specific images from the KFUPM student restaurant, targeting automation use cases such as self-checkout and nutritional monitoring. The dataset was constructed by recording high-resolution videos at the checkout points, extracting thousands of frames, and manually annotating food items

Figure 11. Examples of Macaroni images misclassified due to high intra-class variability, where items with different shapes, colors, and features are grouped under the same class

using bounding boxes in RoboFlow. After preprocessing, 3,222 cropped food images across 26 classes were prepared for training, validation, and testing.

To evaluate the suitability of the dataset for real-world deployment, four pre-trained models, ResNet-50, ResNet-152, MobileNetV2, and EfficientNet-B0, were trained using both the Adam and SGD optimizers. The best performance was achieved by ResNet-152 with Adam, reaching 97.25% accuracy, 97.62% precision, 93.78% recall, and 94.53% F1-score. Across all the models, Adam consistently outperformed SGD, highlighting its robustness for relatively small and imbalanced datasets. The strong performance across multiple architectures reflects the clarity of the collected images, quality of the annotations, and overall reliability of KF-2025. Despite its limited size compared with large-scale public datasets, KF-2025 proved effective for food classification tasks in a real university environment.

This dataset provides a valuable foundation for future research and applications, including real-time food recognition, automated checkout systems, and nutrition monitoring. Further improvements, such as increasing the sample size to address class imbalance, applying data augmentation, and testing on resource-constrained devices, can enhance its practicality and impact for smart canteen system applications.

7. ETHICAL AND PRIVACY CONSIDERATIONS

Because video recording took place at a sensitive location, namely the checkout point of the university canteen, specific ethical and privacy concerns were considered. These included potential exposure to students' faces and any visible banking information. The following measures were implemented to ensure full compliance with privacy and ethical standards:

1. The camera was positioned to capture only the food trays from a top-down angle, ensuring that the students' faces were not included in any recordings.
2. All extracted frames used for the dataset were reviewed to confirm that no credit cards, phones, or personal banking details appeared in any image.
3. The data collection process was carried out with approval from the KFUPM Food Services, and the dataset remained private and confidential, shared only for academic evaluation purposes, and not publicly released.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Saad Obada Alshami – *Department of Control and Instrumentation Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;*
0009-0007-5809-1215
Email: s202182770@kfupm.edu.sa

Corresponding Author

Ali Nasir – *Research Mentor, Assistant Professor, Department of Control and Instrumentation Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;*
0000-0002-7789-428X
Email: ali.nasir@kfupm.edu.sa

ACKNOWLEDGEMENTS

This research was conducted with the support of the SDAIA-KFUPM Joint Research Center for Artificial Intelligence, which provided access to practical equipment and high-specification remote workstations for conducting the experiments. The Department of Control and Instrumentation Engineering and the KFUPM Undergraduate Research Office (URO) provided additional guidance and support.

DATA AVAILABILITY STATEMENT

The dataset will be accessible upon reasonable request from the authors.

REFERENCES

(1) Brunelli, L., Masiero, C., Tosato, D., Beghi, A. & Susto, G. A. Deep learning-based production forecasting in manufacturing: A packaging equipment case study. *Proc. Manuf.* **38**, 248–255 (2019).

(2) Abosaq, H. A. et al. Unusual driver behavior detection in videos using deep learning models. *Sensors* **23**, 311 (2022).

(3) Clements, J. M., Xu, D., Yousefi, N. & Efimov, D. Sequential deep learning for credit risk monitoring with tabular financial data. *arXiv preprint* arXiv:2012.15330 (2020).

(4) Deng, L. et al. Mixed dish recognition with contextual relation and domain alignment. *IEEE Trans. Multimed.* **24**, 2034–2045 (2021).

(5) Wu, X., Fu, X., Liu, Y., Lim, E.-P., Hoi, S. C. H. & Sun, Q. A large-scale benchmark for food image segmentation. *arXiv preprint* arXiv:2105.05409 (2021).

(6) Lu, Y. et al. A multi-task learning approach for meal assessment. *arXiv preprint* arXiv:1806.10343 (2018).

(7) Bossard, L., Guillaumin, M. & Van Gool, L. Food-101 – Mining discriminative components with random forests. *Lect. Notes Comput. Sci.* **8694**, 446–461 (2014).

(8) K. Okamoto and K. Yanai, "UEC-FoodPIX Complete: a Large-Scale Food Image Segmentation Dataset," in Lecture notes in computer science, 2021, pp. 647–659. doi: 10.1007/978-3-030-68821-9_51.

(9) Ciocca, M., Napoletano, P. & Schettini, R. Food recognition: A new dataset, experiments, and results. *IEEE J. Biomed. Health Inform.* **21**, 588–598 (2017).

(10) A. Karabay, A. Bolatov, H. Varol, and M.-Y. Chan, "A central Asian Food dataset for personalized dietary interventions," Nutrients, vol. 15, no. 7, p. 1728, Mar. 2023, doi: 10.3390/nu15071728.

(11) M. Y. Ansari and M. Qaraqe, "MEFood: a Large-Scale Representative benchmark of quotidian foods for the Middle East," IEEE Access, vol. 11, pp. 4589–4601, Jan. 2023, doi: 10.1109/access.2023.3234519.

(12) "EMFID." https://emfid.org/frontend/web/index.php?r=site/goals

(13) V, G., Vutkur, P. & V, P. Food classification using transfer learning technique. *Glob. Transit. Proc.* **3**, 225–229 (2022).

(14) Tripathi, P. Transfer learning on deep neural network: A case study on Food-101 food classifier. Int. J. Eng. Appl. Sci. Technol. 5, (2021).

(15) Philip, A. Optimizing Food-101 classification with transfer learning: A fine-tuning approach using EfficientNetB0. *Int. J. Intell. Inf. Syst.* **13**, 59–77 (2024).

(16) G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," arXiv (Cornell University), Jan. 2016, doi: 10.48550/arxiv.1608.06993.

(17) S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," arXiv (Cornell University), Jan. 2016, doi: 10.48550/arxiv.1611.05431.

(18) A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv (Cornell University), Jan. 2020, doi: 10.48550/arxiv.2010.11929.

(19) C. K. K. Reddy et al., "A fine-tuned vision transformer based enhanced multi-class brain tumor classification using MRI scan imagery," Frontiers in Oncology, vol. 14, Jul. 2024, doi: 10.3389/fonc.2024.1400341.

(20) "Abnormality Detection in Mammography using Deep Convolutional Neural Networks," IEEE Conference Publication — IEEE Xplore, Jun. 01, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8438639

(21) Y. Li, J. Xue, K. Wang, M. Zhang, and Z. Li, "Surface Defect Detection of Fresh-Cut Cauliflowers Based on Convolutional Neural Network with Transfer Learning," Foods, vol. 11, no. 18, p. 2915, Sep. 2022, doi: 10.3390/foods11182915

(22) A. G. Howard et al., "MobileNets: efficient convolutional neural networks for mobile vision applications," arXiv (Cornell University), Jan. 2017, doi: 10.48550/arxiv.1704.04861.

(23) M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," arXiv (Cornell University), Jan. 2019, doi: 10.48550/arxiv.1905.11946

(24) D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," arXiv (Cornell University), Jan. 2014, doi: 10.48550/arxiv.1412.6980.

(25) P. Zhou, J. Feng, C. Ma, C. Xiong, S. C.-H. Hoi, and W. E, "Towards theoretically understanding why SGD generalizes better than ADAM in deep learning," arXiv (Cornell University), Jan. 2020, doi: 10.48550/arxiv.2010.05627.

(26) Singh, P. Detection of Gender in Crowds Using ResNet Model. J. Electrical Systems, 389–408 (2024).

(27) Y. Han, Z. Nan, S. Zhou, and Z. Niu, "DVFS-Aware DNN Inference on GPUs: Latency modeling and Performance analysis," arXiv (Cornell University), Feb. 2025, doi: 10.48550/arxiv.2502.06295.

(28) R.-Y. Ju, T.-Y. Lin, J.-H. Jian, and J.-S. Chiang, "Efficient convolutional neural networks on Raspberry Pi for image classification," arXiv (Cornell University), Jan. 2022, doi: 10.48550/arxiv.2204.00943.