

Vision-Based Convolutional Neural Network System for Automated Recognition of Infant Complementary Foods

Nani Purwati^{a,b,1,*}, R. Rizal Isnanto^{a,c,2}, Martha Irene Kartasurya^{d,3}, Andino Maseleno^{e,4}

^a Doctoral Program of Information Systems, Postgraduate School, Universitas Diponegoro, Semarang, Indonesia

^b Department of Information Systems, Faculty of Engineering and Informatics, Universitas Bina Sarana Informatika, Indonesia

^c Department of Computer Engineering, Faculty of Engineering, Universitas Diponegoro, Semarang, Indonesia

^d Department of Public Health Nutrition, Faculty of Public Health, Universitas Diponegoro, Semarang, Indonesia

^e Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Thailand

¹ nani.npi@bsi.ac.id; ² rizal_isnanto@yahoo.com; ³ marthakartasurya@live.undip.ac.id; ⁴ andino.m@chula.ac.th

* Corresponding Author

ARTICLE INFO

ABSTRACT

Article history

Received November 22, 2025

Revised January 25, 2026

Accepted February 14, 2026

Keywords

Convolutional Neural

Network;

ComFoodID21;

Deep Learning;

Food Recognition;

Infant Complementary Food

Automatic introduction of complementary foods plays an important role in supporting objective and efficient nutritional monitoring of young children. However, the visual characteristics of complementary foods, which tend to be homogeneous in terms of color, texture, and shape pose significant challenges for computer vision-based recognition systems. This study aims to develop and evaluate a deep learning-based system for automatic recognition of complementary foods. This study proposes an image classification system for complementary foods based on Convolutional Neural Network (CNN) using ComFoodID21, a specialized dataset designed to represent the visual characteristics of complementary foods. The main contributions of this research include the development of a CNN-based recognition system for homogeneous visual domains, the compilation of ComFoodID21 as an initial benchmark dataset, and the evaluation of multiple CNN architectures in this domain. Three CNN architectures, namely EfficientNetB0, ResNet50, and MobileNetV2, were initialized using ImageNet pre-trained weights and fine-tuned on the ComFoodID21 dataset. Performance was evaluated using accuracy, precision, recall, F1-score, and training time analysis to assess computational efficiency. Experimental results show that ResNet50 provides the best trade-off between accuracy and training efficiency, achieving 98.28% accuracy with faster convergence despite a slightly longer per-epoch computation time. EfficientNetB0 attains comparable accuracy but requires more epochs to converge, while MobileNetV2 yields lower accuracy and slower convergence in the homogeneous visual domain. The proposed system demonstrates potential for application in smart nutrition monitoring and decision-support systems. The ComFoodID21 dataset is available to the research community and can be accessed at <https://tinyurl.com/datasetcomfoodid21>.

© 2025 The Authors.

Published by Association for Scientific Computing Electrical and Engineering.

This is an open-access article under the [CC-BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



1. Introduction

Complementary feeding is a critical stage in child nutrition for infants aged 6-24 months, marking the transition from exclusive breastfeeding to a more diverse diet [1]-[3]. During this period, complementary foods are essential to meet nutritional requirements that breast milk alone can no longer fulfill, thereby supporting optimal growth, development, and health outcomes [4]-[7]. Accurate and objective monitoring of complementary feeding practices is therefore essential to ensure adequate nutrition in early childhood [8]-[10].

However, current monitoring of complementary food consumption and nutritional composition remains largely manual and subjective [11], [12], leading to inconsistencies and potential nutritional disparities, particularly in regions with diverse feeding practices and limited resources [5], [13]. Variability in the nutritional composition of complementary foods further reduces the reliability of manual assessments [14]-[16], increasing the risk of nutritional imbalance in early childhood [17], [18]. These limitations highlight the need for automated and objective monitoring approaches.

Computer vision-based food recognition systems, supported by artificial intelligence (AI), have emerged as a promising solution for improving dietary assessment and food monitoring [19], [20]. By automating food identification and classification, AI-based systems can reduce human error and improve efficiency in nutrition tracking [21]-[26]. Convolutional Neural Networks (CNNs) have been widely adopted for food image recognition due to their strong feature extraction capabilities and high classification accuracy [27]-[31]. For example, Inception-v4 achieved 95.2% accuracy on Indonesian food [32], while YOLOv5 and Faster R-CNN also showed promising results in detection and recognition [33].

Despite these advances, most existing food recognition studies and benchmark datasets primarily focus on adult or general foods [34]-[38]. Popular datasets such as Food-101 [39], UEC FOOD 256 [33], and Nutrition5k [40], along with other large-scale datasets including UEC FOOD 100/256 [33], Food2K [37], MyFoodRepo-273 [41], MedGRFood [42], and THFOOD-100 [43], contain food images with rich variations in color, texture, shape, and presentation. In contrast, complementary foods typically exhibit highly homogeneous visual characteristics, such as soft textures (e.g., porridge or puree), limited color variation dominated by pale or neutral tones, and minimal structural edges or distinctive shapes [44]-[46]. As a result, the distribution of visual features in complementary food images is more compact and overlapping across classes, making inter-class discrimination substantially more challenging than in adult food datasets such as Food-101.

This visual homogeneity presents a major challenge for CNN-based classification, as CNNs often rely on texture contrast, color variation, and shape cues to differentiate between food categories [47], [48]. Although transfer learning and data augmentation techniques are commonly employed to improve model robustness [32], [49]-[52], these approaches are not always sufficient for domains characterized by high intra-class similarity and low inter-class variability, such as complementary foods [53], [54]. In addition, differences in background, lighting, and perspective also complicate the recognition task [49], [55]. The absence of domain-specific datasets further limits systematic evaluation and benchmarking of recognition models in this domain.

Recent advances, including ensemble learning and knowledge distillation combining CNNs and Vision Transformers, have shown potential in improving classification accuracy and feature extraction [53], [56]-[61]. Hierarchical classification, food localization, and new architectures such as ASTFF-Net have also been introduced to handle multi-object scenes and complex lighting [49], [54]. However, no research has specifically examined CNN-based recognition for complementary foods. The lack of domain-specific datasets and systematic evaluation highlights a significant research gap, motivating this study to explore CNN-based recognition as an initial benchmark for complementary food image analysis.

To address this gap, this research proposes a vision-based complementary food recognition system using Convolutional Neural Networks. The proposed system is designed to automatically classify complementary food images with high accuracy, supporting more objective and efficient child

nutrition monitoring. To enable this, a new dataset named ComFoodID21 is developed, specifically representing complementary foods with homogeneous visual characteristics that are not captured in existing adult food datasets.

The novelty of this research lies in the construction of a complementary food-specific dataset and the systematic evaluation of CNN architectures under conditions of high visual homogeneity. The main contributions of this research are as follows:

- (1) The design and development of a CNN-based complementary food recognition system tailored to visually homogeneous domains;
- (2) The construction of the ComFoodID21 dataset as an initial benchmark for deep learning-based complementary food recognition; and
- (3) An empirical evaluation of multiple CNN architectures to analyze their performance and computational efficiency in this challenging domain.

2. Method

This study proposes a deep learning-based image classification framework to recognize complementary foods for breastfed infants that have homogeneous visual characteristics. The approach used adopts transfer learning to compare the performance of three Convolutional Neural Network (CNN) architectures, namely MobileNetV2, EfficientNetB0, and ResNet50, under the same experimental settings. The research methodology flow includes the stages of data collection, pre-processing and instance-based dataset partitioning, data augmentation, CNN model training, and performance evaluation. The overall methodology flow diagram is shown in Fig. 1.

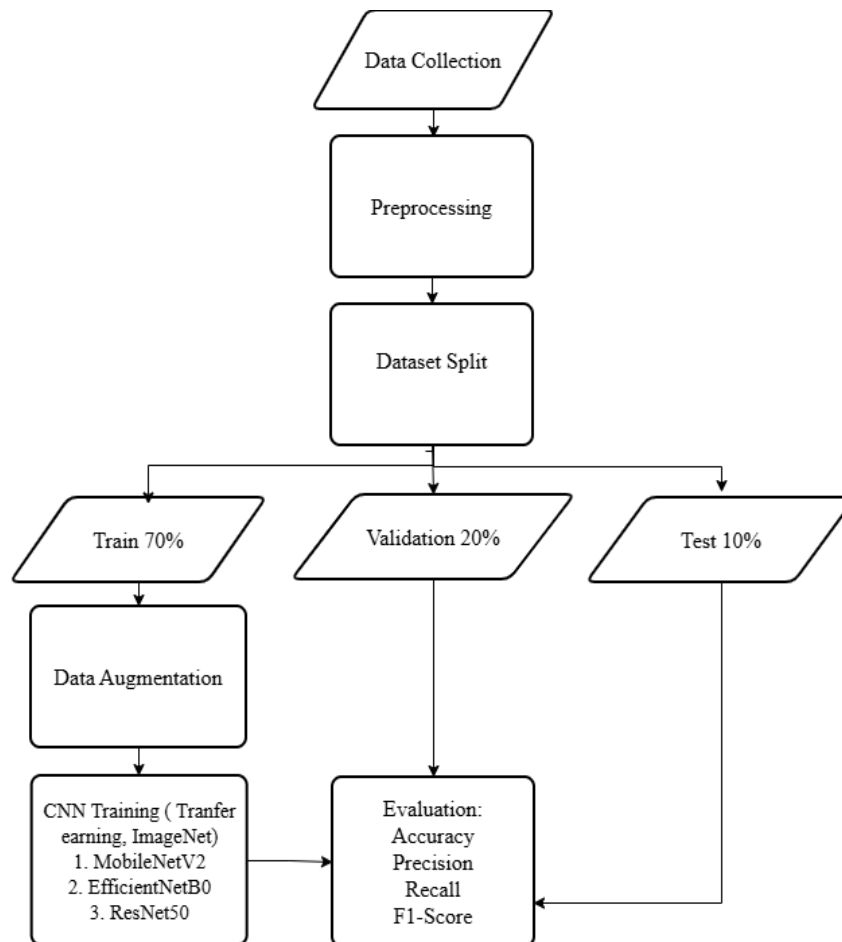


Fig. 1. The proposed method

2.1. Data Collection and Preparation

The dataset used in this study, ComFoodID21, was specifically developed to represent complementary feeding menus for infants aged 6–24 months. The compilation of the dataset began with the design of complementary feeding menus based on the nutritional and health standards of the Indonesian Ministry of Health and the Complementary Feeding for Toddlers book [62], [63], as well as through validation by nutritionists and the Nutrisurvey nutrition database application. This dataset consists of 1,378 images of complementary foods grouped into 21 menu classes, with a total of 71 food categories reflecting the variety of complementary foods typical of Indonesia. Fig. 2 shows an example of the ComFoodID21 dataset.

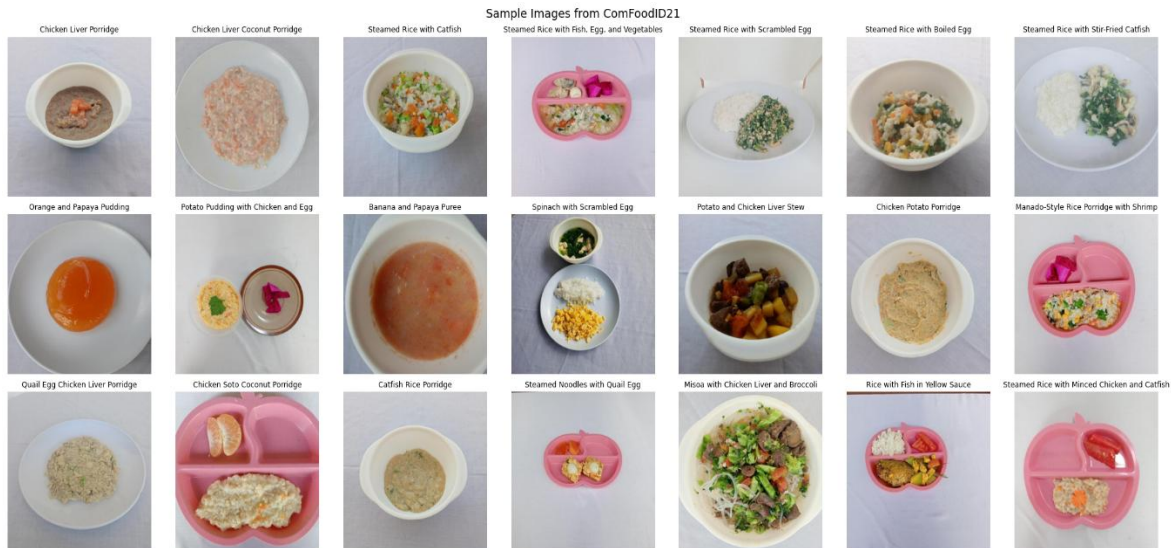


Fig. 2. Example of the ComFoodID21 Dataset

2.2. Pre-Processing and Dataset Division

The images that have been collected then undergo pre-processing, including resizing according to the model input size, normalising pixel values, and adjusting the data format. The dataset was annotated using Roboflow in classification mode, so that each image was labelled with a class and organised into a class-based directory structure for CNN training.

The dataset was divided into 70% training data, 20% validation data, and 10% test data. To prevent data leakage, the dataset was separated at the instance level (plate level), rather than at the image level. All images originating from a single plate of food-including variations in angle and distance of image capture-were placed in only one subset (training, validation, or testing). Class distribution was verified to ensure data balance between subsets. Fig. 3 shows the class-wise distribution graph of the dataset.

2.3. Data Augmentation

Data augmentation was applied only to the training set to enhance model generalization and reduce overfitting. The augmentation techniques included random small-angle rotations, horizontal flipping, zooming, and slight brightness variations. Validation and test datasets were kept unchanged to ensure unbiased performance evaluation. Fig. 4 is an example of augmentation results.

2.4. CNN Architectures and Training Strategy

Three CNN architectures based on transfer learning were employed. MobileNetV2 and EfficientNetB0 were selected for their computational efficiency and lightweight design, while ResNet50 was used as a deeper baseline model to assess feature extraction capability in visually homogeneous domains.

All models were initialized with ImageNet pre-trained weights. Training was conducted using the Adam optimizer with categorical cross-entropy as the loss function. Early stopping and learning rate reduction strategies were implemented to mitigate overfitting. Each model was trained using the same dataset configuration to ensure fair and consistent comparison.

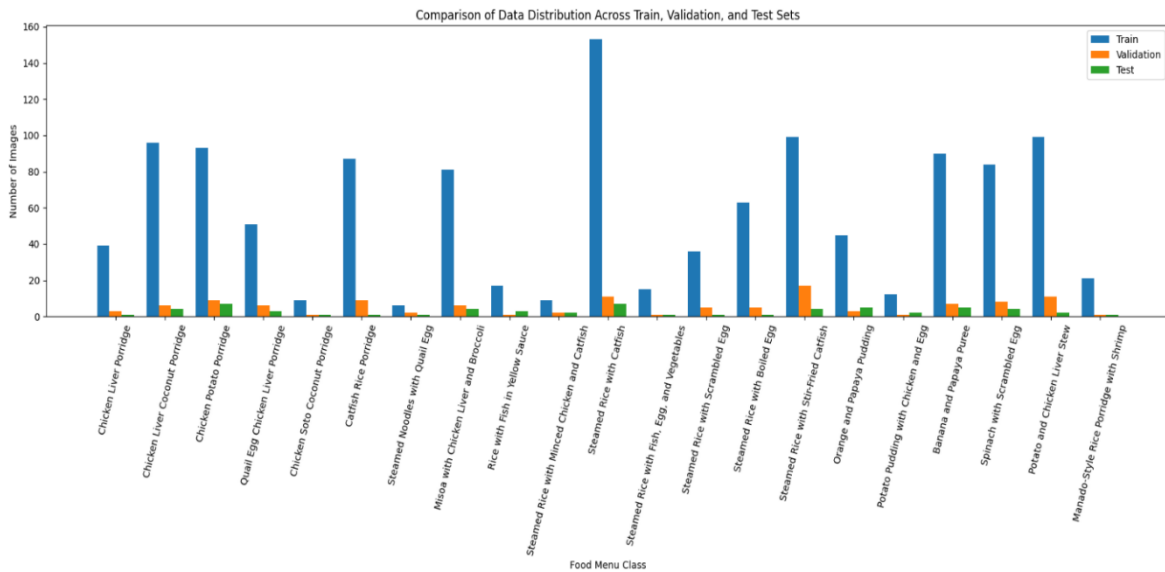


Fig. 3. Dataset distribution per class ComFoodID21 dataset

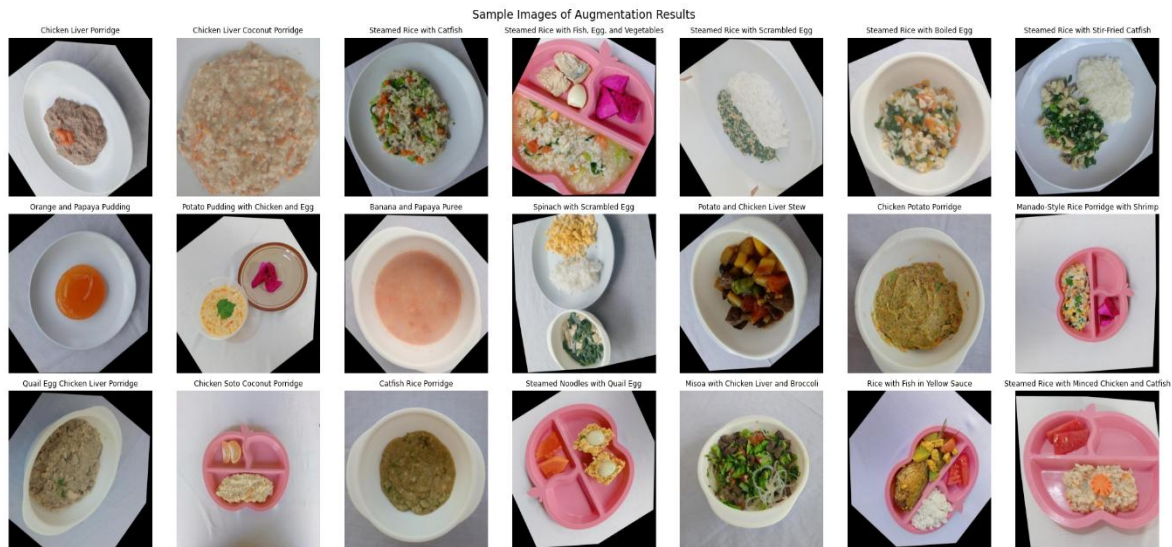


Fig. 4. Sample images from augmentation results

2.5. Evaluation Matrix

Model performance was evaluated using accuracy, precision, recall, and F1-score metrics. These metrics provide a comprehensive assessment of classification performance, particularly under potential class imbalance conditions. Final evaluation was performed on the held-out test set, which was not used during training or validation.

3. Results and Discussion

This section presents the experimental results and discussion of the proposed computer vision-based complementary food recognition system. Experiments were conducted to evaluate the performance of three Convolutional Neural Network (CNN) architectures, namely MobileNetV2,

EfficientNetB0, and ResNet50, in performing automatic classification of complementary food images using the ComFoodID21 dataset.

3.1. EfficientNetB0 Model Results

Based on Fig. 5, the training process over 30 epochs shows a consistent increase in accuracy accompanied by a steady decrease in loss. During the early training phase (epochs 0–5), both training and validation accuracy increase sharply, indicating effective initial feature learning. After approximately the 10th epoch, the validation accuracy stabilizes around 0.95 and continues to improve gradually, reaching 98.28% in the final evaluation, as summarized in Table 1.

The training and validation loss curves exhibit a smooth and monotonically decreasing trend, approaching convergence at low loss values in the range of 0.15-0.2. The close alignment between training and validation curves suggests no significant overfitting, indicating that the model generalizes well to unseen data. Overall, these results demonstrate that EfficientNetB0 provides a balanced trade-off between accuracy and computational efficiency when applied to highly homogeneous complementary food images, where discriminative features are primarily derived from subtle texture and color variations.

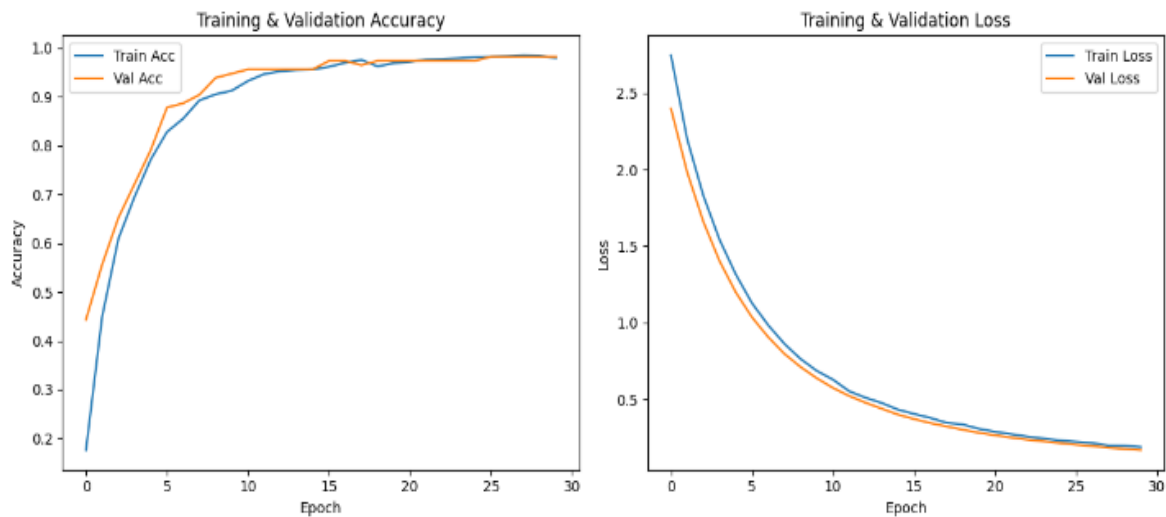


Fig. 5. EfficientNetB0-Training & validation accuracy and loss

Based on the confusion matrix in Fig. 6, the EfficientNetB0 model shows excellent classification performance on most menu classes in the ComFoodID21 dataset, characterized by a predominance of correct predictions on the main diagonal. Many classes, such as Chicken Potato Porridge, Misoa with Chicken Liver and Broccoli, and Steamed Rice with Minced Chicken and Catfish, were classified accurately without errors. This indicates that the model is capable of extracting relevant visual features when the differences in texture or ingredient composition are sufficiently clear.

However, some prediction errors still occur in classes with high visual similarity, especially soft-textured and similarly colored foods. For example, some Chicken Liver Coconut Porridge samples were mispredicted as Steamed Rice with Scrambled Egg, which was likely due to their similar bright color and soft texture. Similar errors were also found in rice and egg-based classes, as well as in classes with a limited number of test samples, where a single error had a proportionally significant impact. This pattern shows that classification errors are mainly influenced by visual homogeneity between classes, which is a major challenge in baby food recognition.

3.2. MobileNetV2 Model Results

The MobileNetV2 model achieved an overall accuracy of 93.10% with a precision value of 89.68%, recall of 90.87%, and an F1-score of 88.21%. Although its performance was lower than that of EfficientNetB0 and ResNet50, these results show that lightweight architectures are still capable of

reliably classifying side dish images in domains with high visual homogeneity. As shown in Fig. 7, the training and validation accuracy curves increase sharply in the early epochs and begin to stabilize after approximately the 10th epoch, while the loss curve shows a consistent decline with no noticeable difference between the training and validation data, indicating a stable learning process without signs of overfitting. However, slower convergence and lower accuracy compared to deeper architectures indicate that MobileNetV2's design, which focuses on computational efficiency, may limit its ability to extract discriminative features in food images with very similar textures and colors.

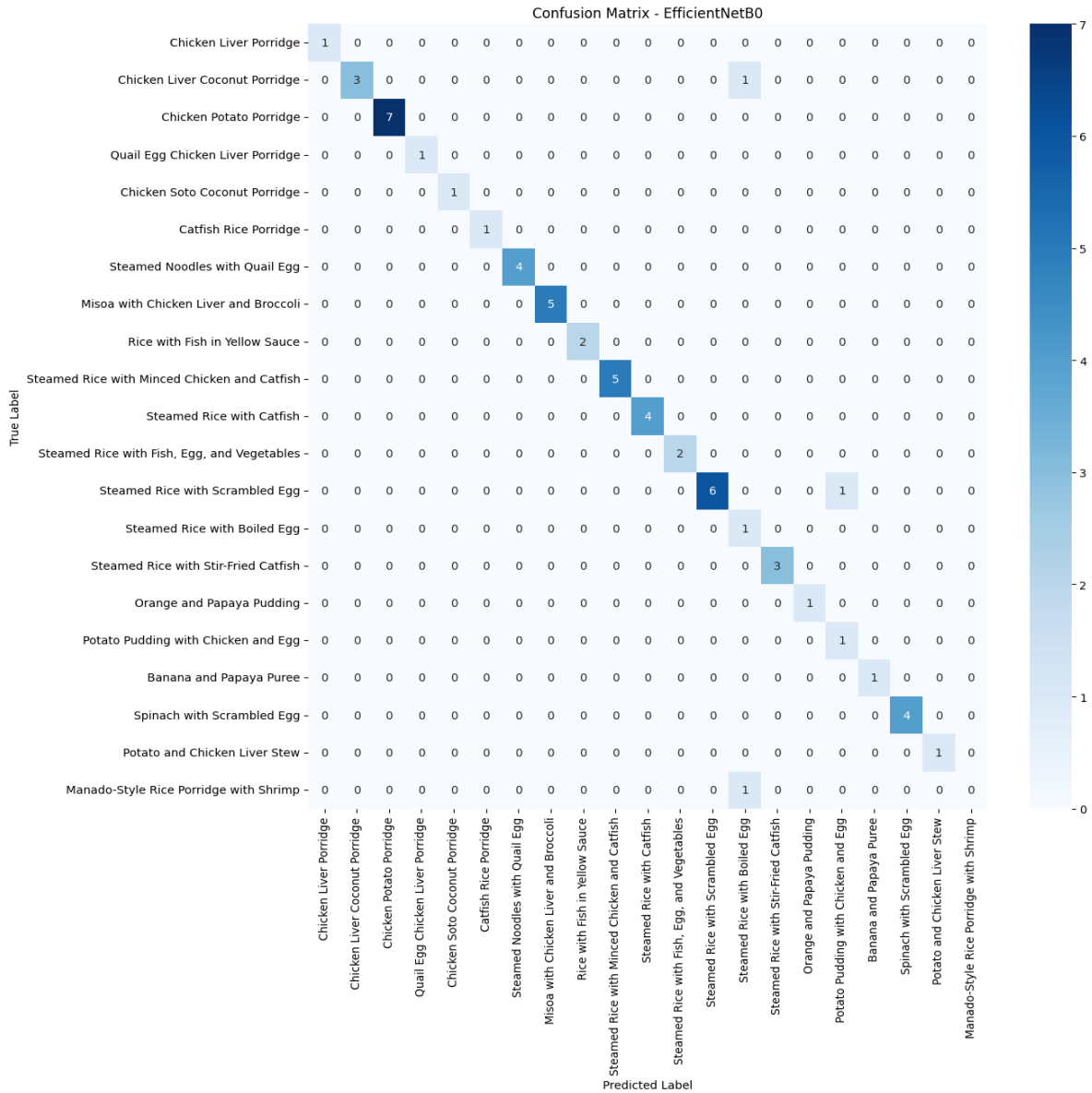


Fig. 6. Confusion Matrix- EfficientNetB0

The confusion matrix for MobileNetV2 in Fig. 8 shows that most samples in each class were classified correctly, as indicated by the dominance of values on the main diagonal. Classes with relatively unique visual characteristics, such as Chicken Potato Porridge, Steamed Rice with Scrambled Egg, and Spinach with Scrambled Egg, had high prediction accuracy and experienced almost no classification errors. However, some errors still occur in classes with high color and texture similarity, especially in porridge-based and soft rice-based menus. Misclassification generally occurs between menus that are both brightly colored and soft in texture, indicating the limitations of MobileNetV2 in extracting subtle discriminative features in domains with high visual homogeneity.

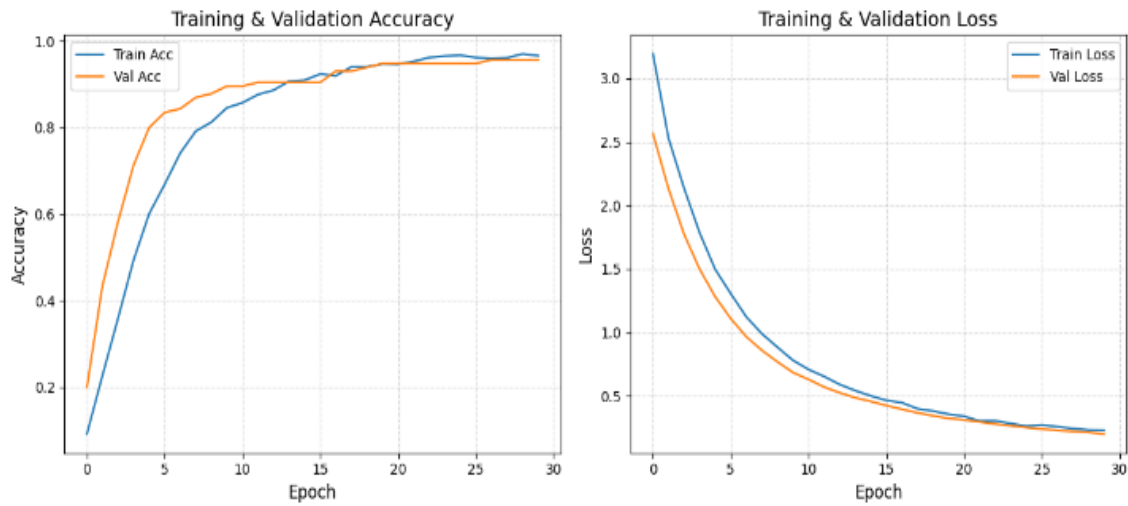


Fig. 7. MobileNetV2-Training & validation accuracy and loss

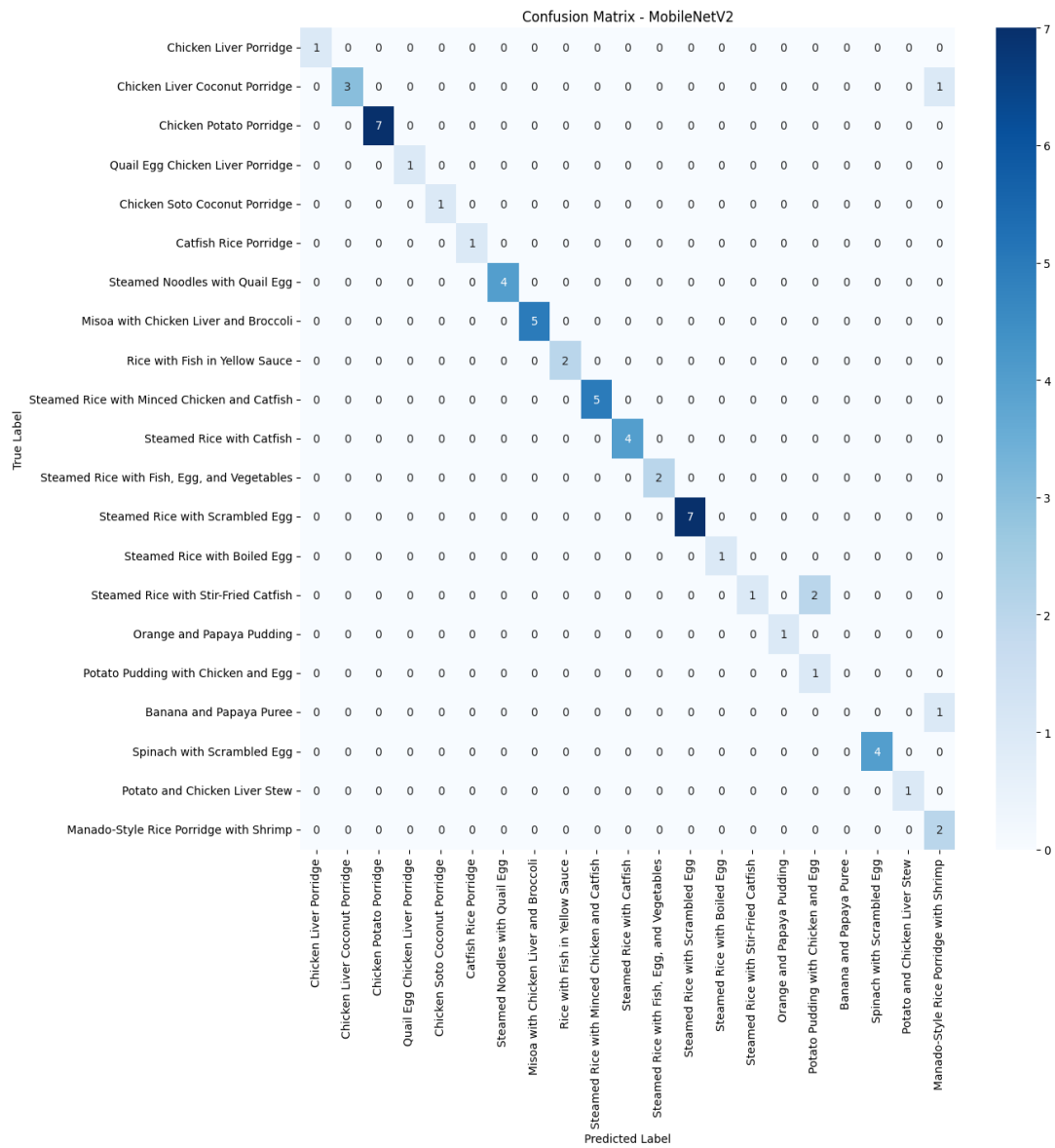


Fig. 8. MobileNetV2-Confusion matrix

3.3. Results of the ResNet50 Model

The training results using the ResNet50 architecture show excellent and stable performance in the classification process of complementary food images. Based on the Training & Validation Accuracy graph in Fig. 9, the model experienced a rapid increase in accuracy in the first 10 epochs, and stabilized around 98-99% in the 20th epoch until the end of the training. The difference between training accuracy and validation accuracy is very small, indicating that the model does not suffer from overfitting and is able to generalize well to the validation data. This is reinforced by the loss graph, which decreases consistently until it converges below 0.1, indicating an efficient and stable optimization process during training.

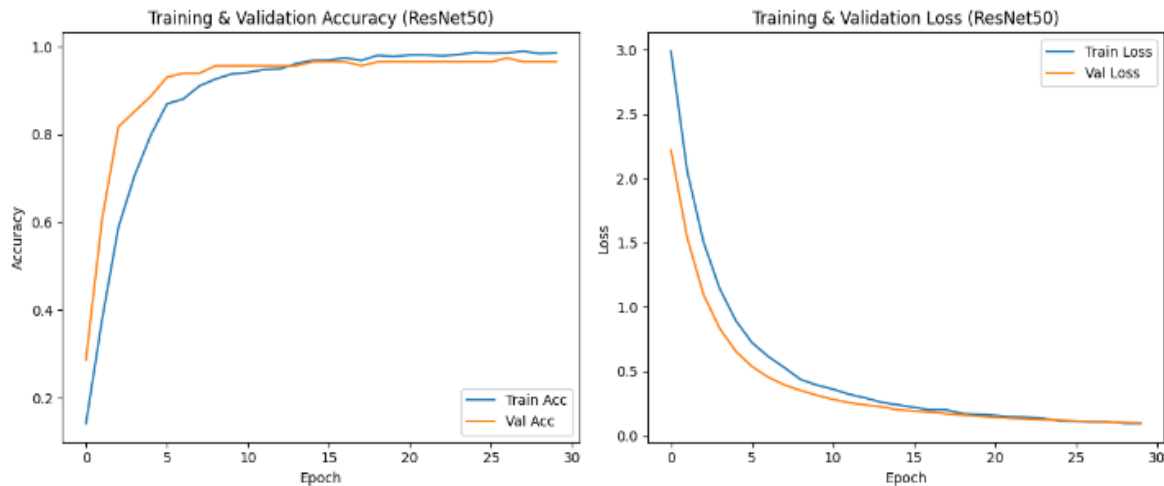


Fig. 9. ResNet50-Training & validation accuracy and loss

The confusion matrix visualization in Fig. 10 shows that the ResNet50 model is able to classify most side dish classes with high accuracy, as indicated by the dominance of high values on the diagonal of the matrix. Several classes, such as Chicken Potato Porridge, Steamed Rice with Scrambled Egg, Steamed Noodles with Quail Egg, and Spinach with Scrambled Egg, are classified with near-perfect accuracy. Prediction errors are relatively limited and generally occur in classes with very similar visual characteristics, particularly in rice-based dishes or porridge with a soft texture and no visually noticeable differences in protein composition. This pattern is consistent with the high level of visual homogeneity in the ComFoodID21 dataset and does not indicate any systematic bias in the model. Quantitatively, ResNet50 achieved an accuracy of 98.28%, precision of 99.14%, recall of 98.28%, and F1-score of 98.28%, confirming the superiority of this model over other CNN architectures.

3.4. Discussion

Experimental results show that the three Convolutional Neural Network (CNN) architectures EfficientNetB0, MobileNetV2, and ResNet50 are able to perform classification of complementary food images with high accuracy, but with different performance characteristics and computational efficiency. Table 1 presents a comparative analysis of classification performance and training efficiency across the three CNN architectures. While EfficientNetB0 achieved the highest validation accuracy (98.28%) with the fastest average computation time per epoch (~6.7 s), ResNet50 demonstrated the fastest convergence, reaching stable validation performance within approximately 12-15 epochs. In contrast, MobileNetV2 required a substantially larger number of epochs to converge (approximately 25-30 epochs), despite its lightweight architecture.

Although MobileNetV2 is designed to be computationally efficient, its slower convergence on the ComFoodID21 dataset suggests that lightweight depthwise separable convolutions may struggle to learn discriminative features in visually homogeneous food images. ResNet50, despite having a slightly higher average time per epoch, benefited from its residual learning mechanism, enabling faster

optimization and more stable feature learning. These results indicate that convergence behavior and architectural learning capacity play a more critical role than per-epoch computational cost when classifying complementary foods with low texture and color variability.

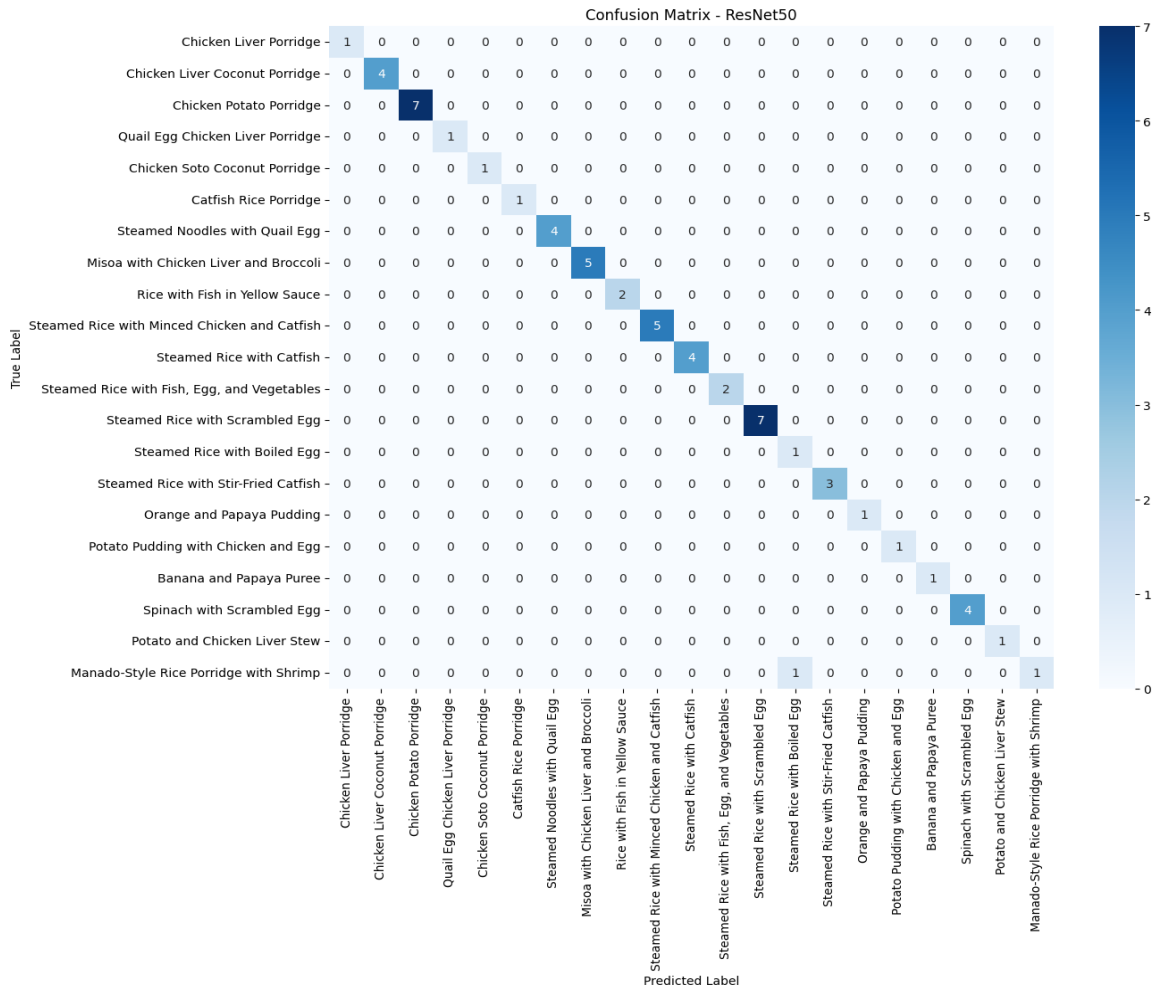


Fig. 10. ResNet50-Confusion matrix

Table 1. Performance and Training Efficiency Comparison of Three CNN Architectures on the ComFoodID21 Dataset

Model	Epoch	Average Per Epoch (seconds)	Epochs to Convergence	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Scientific Observation
EfficientNetB0	30	~6.7	~18–20	98.28	97.62	97.62	96.83	Fast per epoch; balanced accuracy and convergence
MobileNetV2	30	~8.3	~25–30	93.10	89.68	90.87	88.21	Lightweight but slower convergence on homogeneous textures
ResNet50	30	~8.9	~12–15	98.28	99.14	98.28	98.28	Deeper residual learning enables faster convergence

To further verify that the model focuses on semantically relevant regions rather than background artifacts, Grad-CAM visualizations were generated using the ResNet50 model. As shown in Fig. 11, the highest activation regions consistently correspond to the food content, particularly areas with distinctive texture and color composition, while minimal attention is paid to the plate boundaries or

background surface. This observation confirms that the proposed model learns meaningful visual cues from complementary food images and mitigates background bias, which is a common issue in small-scale food image datasets.

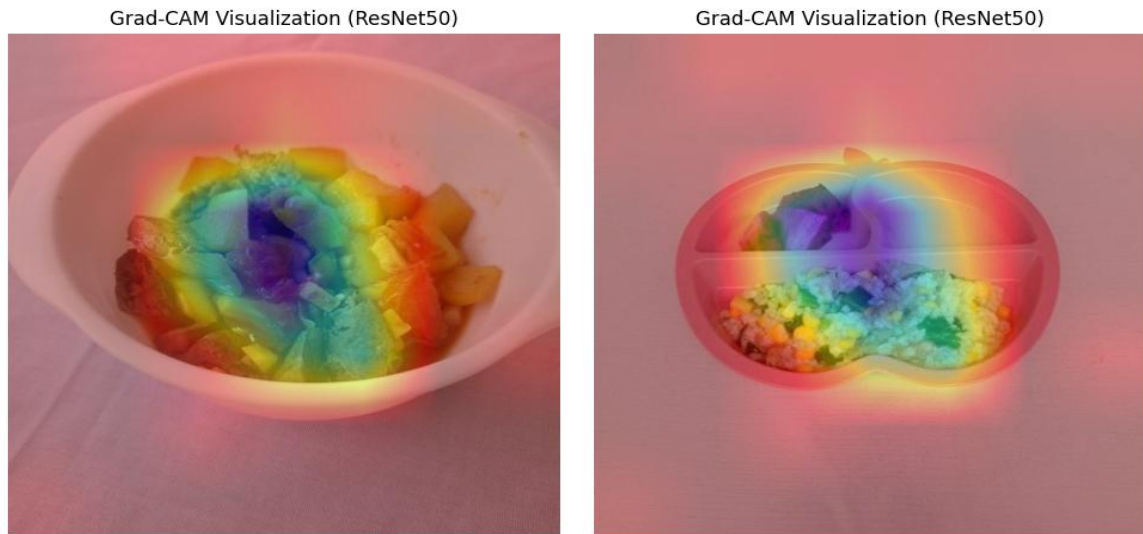


Fig. 11. Sample of Grad-CAM visualization results

Table 2 presents a comparison of this study with several previous studies relevant to the field of deep learning-based food image classification. The comparison focuses on the type of dataset, methodological approach, and the best model used in each study. The comparative studies were selected to represent various food classification scenarios, ranging from adult food, general food, to regional food. Through this comparison, it can be seen that previous studies generally used datasets with relatively more diverse visual variations, while this study focused on images of baby food, which have more uniform visual characteristics. Thus, this table serves to position the proposed study in the context of existing literature and clarify the differences in domains and challenges addressed.

Table 2. Comparison with previous studies

Study	Dataset	Method	Best Model
[42]	MedGRFood	EfficientNetB2	Top-1: 83.4%, Top-5: 97.8%
[64]	Food-11	Transfer Learning + Ensemble Learning (VGG19, ResNet50, MobileNetV2, AlexNet)	Ensemble Learning: 96.8%
[65]	Indian food	ResNet50 & InceptionV3	InceptionV3: 96.24%
Our study	ComFoodID21	EfficientNetB2, MobileNetV2, ResNet50	ResNet50: 98.28%

Furthermore, [66] reported that MobileNetV2, although computationally efficient, suffers from limitations in generalization to objects with similar shapes and colors, which is consistent with the performance gap observed between MobileNetV2 and ResNet50 in Table 1. In addition, previous research [67] demonstrated that transfer learning-based deep architectures such as ResNet50, outperform lightweight models in capturing complex food semantics.

4. Conclusion

This study proposes and evaluates a computer vision-based side dish recognition system using a Convolutional Neural Network (CNN) architecture, utilising the ComFoodID21 dataset, which has a high level of visual homogeneity in terms of colour, texture, and shape. The experimental results show that the CNN model is still capable of learning discriminative visual representations even though the differences between classes are relatively subtle. Among the three architectures tested, ResNet50

showed the best performance with an accuracy of 98.28% and faster convergence compared to EfficientNetB0 and MobileNetV2, confirming the superiority of deeper residual architectures in the highly homogeneous food image domain.

Scientifically, this research contributes an initial benchmark for the classification of food images and empirical analysis of the behaviour of various CNN architectures in domains with high visual homogeneity. Practically, the proposed system has the potential to be applied in smart nutrition monitoring applications, such as automatic food consumption recording or decision support systems for parents and health workers, especially in controlled image capture scenarios.

However, this study has several limitations, including controlled data collection conditions, which limit generalisation to real-world environments, regional food representation, and sample imbalance between classes. Future research could focus on developing a hybrid CNN–Transformer model, applying adaptive domain fine-tuning, and integrating nutrient content estimation by linking classification results with a nutrient database or recipe metadata. The ComFoodID21 dataset and its annotations are provided to encourage further research in the field of artificial intelligence-based food recognition.

Supplementary Materials: The following supporting information can be downloaded at: <https://tinyurl.com/datasetcomfoodid21>.

Author Contribution: Nani Purwati contributed to conceptualization, methodology, data acquisition, software development, validation, formal analysis, investigation, resources, initial draft writing, review and editing, visualization, and funding procurement. R Rizal Isnanto contributed to validation, review and editing, and supervision, while Martha Irene Kartasurya contributed to validation, review and editing, and supervision. Andino Maselena contributed to validation, review and editing, and supervision.

Funding: The authors would like to express their sincere appreciation to Universitas Diponegoro, particularly through the Institute for Research and Community Service (LPPM), for its support of this research. The institution's commitment to academic development and sustainable finance research has played a pivotal role in the successful completion of this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

- [1] C. K. Lutter, L. Grummer-Strawn, and L. Rogers, "Complementary feeding of infants and young children 6 to 23 months of age," *Nutrition Reviews*, vol. 79, no. 8, pp. 825-846, 2021, <https://doi.org/10.1093/nutrit/nuaa143>.
- [2] B. Vlieg-Boerstra *et al.*, "Guidance for healthy complementary feeding practices for allergy prevention in developed countries: An EAACI interest group report," *Pediatric Allergy and Immunology*, vol. 36, no. 7, p. e70150, 2025, <https://doi.org/10.1111/pai.70150>.
- [3] P. Bafila, K. Tiwari, G. Bisla, V. Yadav, and P. Negi, "Nutritional and compositional aspects of complementary feeding," *AIP Conference Proceedings*, vol. 2978, no. 1, 2024, <https://doi.org/10.1063/5.0184255>.
- [4] A. B. Olubunmi *et al.*, "Evaluation of Nutrient Composition of Complementary Foods Made From Locally Available Cereals and Legumes," *Nigerian Journal of Nutritional Sciences*, vol. 45, no. 2, pp. 205-213, 2025, <https://eurekamag.com/research/096/975/096975873>.
- [5] F. Indrio *et al.*, "Optimizing Weaning Strategies: A Consensus-Based Approach for Complementary Feeding in the Middle East and North Africa Region," *Pediatric Gastroenterology, Hepatology & Nutrition*, vol. 28, no. 4, pp. 201-214, 2025, <https://doi.org/10.5223/pghn.2025.28.4.201>.
- [6] H. Kalhoff and M. Kersting, "Programming long-term health: Nutrition and diet in infants aged 6 months to 1 year," *Early Nutrition and Long-Term Health (Second Edition)*, pp. 563-595, 2022, <https://doi.org/10.1016/B978-0-12-824389-3.00003-9>.

- [7] Y. Shi *et al.*, “Impact of complementary feeding on infant gut microbiome, metabolites and early development,” *Food & Function*, vol. 15, no. 21, pp. 10663-10678, 2024, <https://doi.org/10.1039/D4FO03948C>.
- [8] R. Vázquez-Frias *et al.*, “Consensus on complementary feeding from the Latin American Society for Pediatric Gastroenterology, Hepatology and Nutrition: COCO 2023,” *Revista de Gastroenterología de México*, vol. 88, no. 1, pp. 57-70, 2023, <https://doi.org/10.1016/j.rgm.2022.11.001>.
- [9] D. E. Akiso and S. Laelago, “Complementary feeding hygienic practices and its associated factors among mothers of children aged 6–23 months old in Central Ethiopia, 2023,” *Frontiers in Public Health*, vol. 13, 2025, <https://doi.org/10.3389/fpubh.2025.1475315>.
- [10] I. Dwijayanti, M. Sulistyowati, E. R. Isaura, A. L. Mutiarani, and G. M. S. Noor, “Exploring factors influencing complementary feeding practices of mothers with infants aged 6-23 months in Sidoarjo Regency, Indonesia: A qualitative study,” *African Journal of Reproductive Health*, vol. 28, no. 10, pp. 25-33, 2024, <https://doi.org/10.29063/ajrh2024/v28i10s.3>.
- [11] A. A. Crystal, M. Valero, V. Nino, and K. H. Ingram, “Empowering Diabetics: Advancements in Smartphone-Based Food Classification, Volume Measurement, and Nutritional Estimation †,” *Sensors*, vol. 24, no. 13, p. 4089, 2024, <https://doi.org/10.3390/s24134089>.
- [12] T. Ghosh *et al.*, “I2N: image to nutrients, a sensor guided semi-automated tool for annotation of images for nutrition analysis of eating episodes,” *Frontiers in Nutrition*, vol. 10, 2023, <https://doi.org/10.3389/fnut.2023.1191962>.
- [13] S. Malik, K. Krishnaswamy, and A. Mustapha, “Hazard analysis and risk-based preventive controls (Harpc): Current food safety and quality standards for complementary foods,” *Foods*, vol. 10, no. 9, p. 2199, 2021, <https://doi.org/10.3390/foods10092199>.
- [14] V. W. K. Tan, A. J. Lim, K. McCrickerd, and C. G. Forde, “Sensory profiles and mothers’ expectations and beliefs about age appropriate snacks for infants and toddlers in Singapore,” *Food Quality and Preference*, vol. 97, p. 104474, 2022, <https://doi.org/10.1016/j.foodqual.2021.104474>.
- [15] A. Tahosin *et al.*, “Assessment of the quality of flour from diverse ingredients and their fortified complementary foods,” *Applied Food Research*, vol. 5, no. 2, p. 101455, 2025, <https://doi.org/10.1016/j.afres.2025.101455>.
- [16] D. Zhiyong, “Status and prospects of the infants and young children complementary foods step-feeding system,” *Food and Machinery*, vol. 40, no. 8, 2024, <https://doi.org/10.13652/j.spjx.1003.5788.2024.60092>.
- [17] A. D. Oguizu and A. James, “Nutrient Composition and Sensory Attributes of Complementary Foods made from Maize (zea Mays), Soybean (glycine Max), Groundnut (arachis Hypogea) and Dates (Phoenix Dactylifera),” *Nigerian Journal of Nutritional Sciences*, vol. 44, no. 1, pp. 244-255, 2023, <https://dx.doi.org/10.4314/njns.v44i1.23>.
- [18] E. Mario, A. Belay, and E. Amare, “Evaluation of the quality and safety of commercial complementary foods: Implications for nutrient adequacy and conformance with national and international standards,” *PLOS One*, vol. 19, no. 2, p. e0294068, 2024, <https://doi.org/10.1371/journal.pone.0294068>.
- [19] T. Selvaraju, V. Dakshinamurthi, G. Badurudeen, N. Prabhakaran, P. Gopi, and M. A. N. Hussain, “Food Detection and Estimation of Calories and Other Macro-Nutrients and Features to Support Healthy Lifestyle,” *AIP Conference Proceedings*, vol. 3279, no. 1, 2025, <https://doi.org/10.1063/5.0263063>.
- [20] R. Tiwari, D. Koundal, and S. Upadhyay, “Image Based Computing for Food and Health Analytics: Requirements, Challenges, Solutions and Practices,” *Springer Cham*, 2023, <https://doi.org/10.1007/978-3-031-22959-6>.
- [21] O. I. Ogidi, C. M. Wenapere, and O. A. Chukwunonso, “Enhancing Food Safety and Quality Control With Computer Vision Systems,” *IGI Global*, 2025, <https://doi.org/10.4018/979-8-3693-8019-2.ch003>.
- [22] P. Indhumathi, J. T. R. K and S. B. S, “Food Recognition System and Nutrition Tracking Using Convolutional Neural Network Model,” *2024 Asia Pacific Conference on Innovation in Technology (APCIT)*, pp. 1-6, 2024, <https://doi.org/10.1109/APCIT62007.2024.10673543>.

- [23] S. Rokhva, B. Teimourpour, and A. H. Soltani, "Computer vision in the food industry: Accurate, real-time, and automatic food recognition with pretrained MobileNetV2," *Food Humanities*, vol. 3, p. 100378, 2024, <https://doi.org/10.1016/j.foohum.2024.100378>.
- [24] Z.-K. Zhang, H.-Z. Han, X.-Q. Zhao, and Z. Li, "Automatic recognition method for soft packaged food based on improved YOLOv3 model," *Food Machinery*, vol. 39, no. 5, pp. 95-100, 2023, <https://doi.org/10.13652/j.spjx.1003.5788.2022.60167>.
- [25] Y. Liu, "Automatic food recognition based on efficientnet and ResNet," *Journal of Physics: Conference Series*, vol. 2646, no. 1, 2023, <https://doi.org/10.1088/1742-6596/2646/1/012037>.
- [26] X. Dai, "Robust deep-learning based refrigerator food recognition," *Frontiers in Artificial Intelligence*, vol. 7, 2024, <https://doi.org/10.3389/frai.2024.1442948>.
- [27] C. A. Iheanacho and O. R. Vincent, "Classification and recommendation of food intake in West Africa for healthy diet using Deep Learning," *2022 5th Information Technology for Education and Development (ITED)*, pp. 1-6, 2022, <https://doi.org/10.1109/ITED56637.2022.10051387>.
- [28] J. Zhao *et al.*, "Convolutional neural networks in the realm of food quality and safety evaluation: Current achievements and future prospects," *Trends in Food Science & Technology*, vol. 163, p. 105162, 2025, <https://doi.org/10.1016/j.tifs.2025.105162>.
- [29] K. A. Nfor, T. P. Theodore Armand, K. P. Ismaylova, M.-I. Joo, and H.-C. Kim, "An Explainable CNN and Vision Transformer-Based Approach for Real-Time Food Recognition," *Nutrients*, vol. 17, no. 2, p. 362, 2025, <https://doi.org/10.3390/nu17020362>.
- [30] A. Chaitanya, J. Shetty, and P. Chiplunkar, "Food Image Classification and Data Extraction Using Convolutional Neural Network and Web Crawlers," *Procedia Computer Science*, vol. 218, pp. 143-152, 2022, <https://doi.org/10.1016/j.procs.2022.12.410>.
- [31] J. Zhang, H. Liu, and Q. Li, "Deep Migration Learning-Based Aerospace Food Image Classification," *Lecture Notes in Electrical Engineering*, pp. 281-288, 2023, https://doi.org/10.1007/978-981-99-4882-6_40.
- [32] K. Rathod, C. Thacker and R. Adhvaryu, "Food Recognition Using Deep Learning: An Automated Approach for Accurate Food Item Identification," *2024 Parul International Conference on Engineering and Technology (PICET)*, pp. 1-6, 2024, <https://doi.org/10.1109/PICET60765.2024.10716185>.
- [33] S. W. Tan, C. P. Lee, K. M. Lim and J. Y. Lim, "Food Detection and Recognition with Deep Learning: A Comparative Study," *2023 11th International Conference on Information and Communication Technology (ICoICT)*, pp. 283-288, 2023, <https://doi.org/10.1109/ICoICT58202.2023.10262523>.
- [34] P. Bohlol, S. Hosseinpour, and M. S. Firouz, "A deep learning model with machine vision system for recognizing type of the food during the food consumption," *Scientific Reports*, vol. 15, no. 1, 2025, <https://doi.org/10.1038/s41598-025-15755-6>.
- [35] 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, 2023, <https://doi.org/10.1109/ACCESS.2023.3234519>.
- [36] Y. C. Lee *et al.*, "Using Machine Learning to Predict Obesity Based on Genome-Wide and Epigenome-Wide Gene-Gene and Gene-Diet Interactions," *Frontiers in Genetics*, vol. 12, 2022, <https://doi.org/10.3389/fgene.2021.783845>.
- [37] W. Min *et al.*, "Large Scale Visual Food Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 9932-9949, 2023, <https://doi.org/10.1109/TPAMI.2023.3237871>.
- [38] N. F. Nishat, T. Biswas, K. R. Noor, M. S. Islam, and H. Ullah, "AI-Based Approaches for Bengali Food Image Recognition: A Review," *Lecture Notes in Networks and Systems*, pp. 591-605, 2025, https://doi.org/10.1007/978-981-96-6300-2_40.
- [39] X. Yuan, Y. Chen and F. Wang, "Hash-Based Food Image Retrieval with Enhanced Feature Correlation Vision Transformer," *2024 17th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1-8, 2024, <https://doi.org/10.1109/CISP-BMEI64163.2024.10906165>.
-

- [40] Y. Han, Q. Cheng, W. Wu, and Z. Huang, "DPF-Nutrition: Food Nutrition Estimation via Depth Prediction and Fusion," *Foods*, vol. 12, no. 23, p. 4293, 2023, <https://doi.org/10.3390/foods12234293>.
- [41] S. P. Mohanty *et al.*, "The Food Recognition Benchmark: Using Deep Learning to Recognize Food in Images," *Frontiers in Nutrition*, vol. 9, 2022, <https://doi.org/10.3389/fnut.2022.875143>.
- [42] F. S. Konstantakopoulos, E. I. Georga and D. I. Fotiadis, "Mediterranean Food Image Recognition Using Deep Convolutional Networks," *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1740-1743, 2021, <https://doi.org/10.1109/EMBC46164.2021.9630481>.
- [43] N. Theera-Ampornpant and P. Treepong, "Thai Food Recognition Using Deep Learning With Cyclical Learning Rates," *IEEE Access*, vol. 12, pp. 174204-174221, 2024, <https://doi.org/10.1109/ACCESS.2024.3503672>.
- [44] J.-J. Tsai, Y.-Y. Su, H. Chang, H.-C. Lin, J.-T. Ku, and M.-M. Chien, "Practices for introducing complementary foods in Taiwan: A nationwide survey from 2013 to 2016," *Journal of the Formosan Medical Association*, vol. 124, no. 11, pp. 1045-1049, 2024, <https://doi.org/10.1016/j.jfma.2024.09.026>.
- [45] C. Tournier, L. Demonteil, E. Ksiazek, A. Marduel, H. Weenen, and S. Nicklaus, "Factors Associated With Food Texture Acceptance in 4- to 36-Month-Old French Children: Findings From a Survey Study," *Frontiers in Nutrition*, vol. 7, 2021, <https://doi.org/10.3389/fnut.2020.616484>.
- [46] W. Zhixu, "New concepts of complementary feeding in infants and toddlers," *Chinese Journal of Child Health Care*, vol. 32, no. 4, pp. 362-367, 2024, <https://doi.org/10.11852/zgetbjzz2024-0250>.
- [47] M. Osório and A. Wichert, "Promoting the Shift From Pixel-Level Correlations to Object Semantics Learning by Rethinking Computer Vision Benchmark Data Sets," *Neural Computation*, vol. 36, no. 8, pp. 1626-1642, 2024, https://doi.org/10.1162/neco_a_01677.
- [48] B. Jamalpur, S. P. G, S. Venkat, A. R, S. Kavitha and G. C. Babu, "Optimizing Food Image Classification Using Black Widow Algorithm and Deep Learning Techniques," *2024 Second International Conference on Advances in Information Technology (ICAIT)*, pp. 1-6, 2024, <https://doi.org/10.1109/ICAIT61638.2024.10690324>.
- [49] S. Phiphitphatphisit and O. Surinta, "Multi-layer adaptive spatial-temporal feature fusion network for efficient food image recognition," *Expert Systems with Applications*, vol. 255, p. 124834, 2024, <https://doi.org/10.1016/j.eswa.2024.124834>.
- [50] T. Raman *et al.*, "CNN Based Study of Improvised Food Image Classification," *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 1051-1057, 2023, <https://doi.org/10.1109/CCWC57344.2023.10099066>.
- [51] A. Nazir, A. Hussain, and A. Assad, "CNN in Food Industry: Current Practices and Future Trends," *Artificial Intelligence in the Food Industry: Enhancing Quality and Safety*, 2025, <https://doi.org/10.1201/9781032633602-17>.
- [52] J. Patel and K. Modi, "Indian Food Image Classification and Recognition with Transfer Learning Technique Using MobileNetV3 and Data Augmentation †," *Engineering Proceedings*, vol. 56, no. 1, p. 197, 2024, <https://doi.org/10.3390/ASEC2023-15341>.
- [53] L. Nong, G. Peng, T. Xu, and J. Zhu, "From ensemble to knowledge distillation: Improving large-scale food recognition," *Engineering Applications of Artificial Intelligence*, vol. 151, p. 110727, 2025, <https://doi.org/10.1016/j.engappai.2025.110727>.
- [54] R. Mao, J. He, Z. Shao, S. K. Yarlagadda, and F. Zhu, "Visual Aware Hierarchy Based Food Recognition," *Lecture Notes in Computer Science*, pp. 571-598, 2021, https://doi.org/10.1007/978-3-030-68821-9_47.
- [55] M. Goel *et al.*, "Dish Detection in Indian Food Platters: A Computational Framework for Diet Management," *Communications in Computer and Information Science*, pp. 231-243, 2024, https://doi.org/10.1007/978-3-031-58181-6_20.
- [56] R. Nijhawan, G. Sinha, A. Batra, M. Kumar, and H. Sharma, "VTnet+Handcrafted based approach for food cuisines classification," *Multimedia Tools and Applications*, vol. 83, no. 4, pp. 10695-10715, 2024, <https://doi.org/10.1007/s11042-023-15800-4>.

- [57] P. Kathar, R. Khandare, M. Das, D. Gupta and S. Singh, "Vision Outlooker-Based Hierarchical Food Classification," *TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)*, pp. 317-322, 2023, <https://doi.org/10.1109/TENCON58879.2023.10322530>.
- [58] C. Mao, J. Li, T. Hu, and X. Zhao, "CMVT: ConVit Transformer Network Recombined with Convolutional Layer," *International Journal of Image and Graphics*, vol. 25, no. 1, p. 2450060, 2025, <https://doi.org/10.1142/S0219467824500608>.
- [59] J. Lee, D. Das, M. Hayat, S. Choi, K. Hwang and F. Porikli, "CustomKD: Customizing Large Vision Foundation for Edge Model Improvement via Knowledge Distillation," *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25176-25186, 2025, <https://doi.org/10.1109/CVPR52734.2025.02344>.
- [60] A. Todi, N. Narula, M. Sharma and U. Gupta, "ConvNext: A Contemporary Architecture for Convolutional Neural Networks for Image Classification," *2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT)*, pp. 1-6, 2023, <https://doi.org/10.1109/CISCT57197.2023.10351320>.
- [61] L. Zhang and K. Ma, "Structured Knowledge Distillation for Accurate and Efficient Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 12, pp. 15706-15724, 2023, <https://doi.org/10.1109/TPAMI.2023.3300470>.
- [62] A. I. Susilowardani and I. Budiono, "Pengaruh Edukasi Gizi Terhadap Pengetahuan dan Praktik Ibu Baduta Dalam Pemberian Makanan Pendamping ASI (MP ASI)," *Indonesian Journal of Public Health and Nutrition*, vol. 2, no. 2, pp. 131-136, 2022, <https://doi.org/10.15294/ijphn.v2i2.49868>.
- [63] Kemenkes RI, "Makanan Lokal," *Buku Resep Makanan Lokal Bayi, Balita dan Ibu Hamil*, pp. 1-52, 2023, <https://ayosehat.kemkes.go.id/pub/files/d8a32723535961f3f2a6e44f0f8ba915.pdf>.
- [64] L. Bu, C. H. Id, and X. Zhang, "Recognition of food images based on transfer learning and ensemble learning," *PLOS ONE*, vol. 19, no. 1, p. e0296789, 2024, <https://doi.org/10.1371/journal.pone.0296789>.
- [65] P. C. Patil and V. C. Burkapalli, "Food Cuisine Classification by Convolutional Neural Network based Transfer Learning Approach," *2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, pp. 1-5, 2021, <https://doi.org/10.1109/ICMNWC52512.2021.9688333>.
- [66] S. N, H. S, R. S. R, A. Sungheetha, M. M and K. Adare, "Fruit and Vegetable Detection with Calorie Estimation Built on Mobilenetv2," *2024 Second International Conference on Advances in Information Technology (ICAIT)*, pp. 1-6, 2024, <https://doi.org/10.1109/ICAIT61638.2024.10690758>.
- [67] A. R. Kandula, K. B. Koushik, K. Prashanth and K. C. V. S. Raviteja, "From Image to Recipe: Transfer Learning for Identifying Food Ingredients," *2025 International Conference on Advanced Computing Technologies (ICoACT)*, pp. 1-6, 2025, <https://doi.org/10.1109/ICoACT63339.2025.11005331>.