

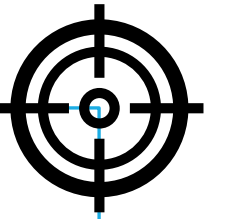
Comparative Analysis of State-of-the-Art Deep Learning Models for Food Image Recognition

Researcher: Atheer Mohammed Al-Shahrani
College of Informatics — Midocean University
Year: 2026

Abstract

- Food image recognition is a key research area in computer vision and artificial intelligence.
- It has important applications in healthcare, nutrition monitoring, and smart food systems.
- However, accurate classification is challenging due to:
 - High intra-class variation
 - Visual similarity between different food categories
- **This study presents a comparative analysis of three deep learning models:**
 - ResNet50
 - MobileNetV2
 - EfficientNetB0
- The models were evaluated on three datasets with different complexity levels:
 - Food-11, Food-34, and Food-101
- **The results show that:**
 - EfficientNetB0 achieved the best overall performance and generalization
 - ResNet50 provided strong and stable results
 - MobileNetV2 offered high computational efficiency for resource-constrained environments

Introduction



- Food image recognition has become an important area in computer vision and artificial intelligence.
- It has wide applications in healthcare, nutrition monitoring, dietary assessment, agriculture, and smart food systems.
- Deep learning has improved food classification significantly compared with traditional image-processing methods.
- However, the task remains difficult because food images show high intra-class variation and strong similarity between different categories.

Problem Statement



- Food recognition is challenging because the same food can look very different depending on ingredients, cooking style, presentation, lighting, and camera angle.
- Different food categories may also look very similar in terms of color, shape, and texture.
- Many previous studies focused on only one model or one dataset, which makes it difficult to judge which architecture works best.
- So, a systematic comparison of modern deep learning models is needed to understand accuracy, efficiency, and generalization.

Research Objectives

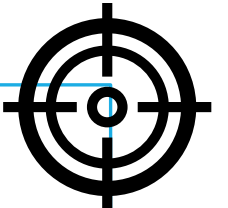
Compare the performance of state-of-the-art deep learning models in food image recognition.

Evaluate how well the models handle visual complexity, intra-class variation, and inter-class similarity.

Compare models based on both accuracy and computational efficiency.

Provide recommendations for practical applications such as dietary tracking and smart food systems.

Research Contribution



- A structured comparative evaluation framework was created for fair testing.
- The study analyzes how architectural design affects accuracy and robustness.
- It highlights trade-offs between model size, speed, and performance.
- The results can help researchers and practitioners choose suitable models for real-world food recognition systems.

Background



- Food image recognition is a rapidly growing field in computer vision and artificial intelligence due to its applications in healthcare, nutrition monitoring, and smart food systems.
- Early approaches relied on handcrafted features such as color, texture, and shape, but they struggled with real-world variations like lighting and background noise.
- The emergence of deep learning, especially Convolutional Neural Networks (CNNs), enabled automatic feature extraction and significantly improved classification performance.
- Advanced architectures such as deep CNNs and attention-based models further enhanced accuracy by capturing complex visual patterns and focusing on relevant regions in images.

(Liu et al., 2025; Rasyidi et al., 2024; Suddul & Seguin, 2023)

Deep Learning Models in Food Recognition



- CNNs are the most widely used models due to their ability to learn hierarchical image features automatically.
- Deep CNN architectures such as ResNet, DenseNet, and EfficientNet improve learning efficiency and overcome issues like vanishing gradients.
- Attention-based CNN models enhance classification by focusing on important image regions and reducing background noise.
- Recent approaches also explore state-space models for capturing global context in images.

Related Work

Deep learning models consistently outperform traditional machine learning approaches in food image recognition tasks.

(Liu et al., 2025)

Large-scale comparative studies highlight the importance of dataset diversity and benchmarking multiple architectures.

(Rasyidi et al., 2024)

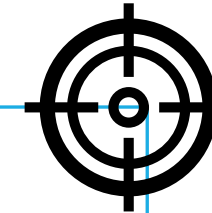
Transfer learning and data augmentation significantly improve model generalization and reduce overfitting.

(Suddul & Seguin, 2023)

Studies from related domains such as agriculture confirm the scalability and robustness of deep learning models.

(Farahnakian et al., 2023)

Research Gap



- Most studies focus on limited models or single datasets.
- There is a lack of fair comparison under unified experimental conditions.
- Limited research evaluates model performance across datasets with different complexity levels.

Methodology Overview

The study used a controlled experimental design. The same preprocessing steps, augmentation methods, and training settings were applied to all models.

Three datasets were selected with increasing difficulty:

- Food-11
- Food-34
- Food-101

Three models were compared:

- ResNet50
- MobileNetV2
- EfficientNetB0

The goal was to see how each model behaves as the classification task becomes harder.

Experimental Environment

- Experiments were run on a local laptop for preprocessing and initial exploration.
- Training was done on cloud platforms such as Google Colab and Kaggle because deep learning training needs GPU support.
- The work used Python, TensorFlow, and Keras.
- GPU acceleration was important for reducing training time and improving efficiency.
- Using the same environment helped support reproducibility.

Datasets

Food-11

- 16,643 images
- 11 food classes
- RGB JPEG images
- Low-complexity benchmark dataset
- Classes include bread, dairy products, desserts, eggs, fried food, meat, noodles/pasta, rice, seafood, soup, and vegetables/fruits

Food-101

- 101,000 images
- 101 classes
- High-complexity benchmark
- Very fine-grained classification task
- Real-world noise and strong similarity between classes

Food-34

- About 24,000 images
- 34 classes
- Medium complexity
- Includes Indian and Western foods
- Has class imbalance and moderate visual similarity

Exploratory Data Analysis



The datasets were examined before training to understand image counts, class distribution, and complexity.

Random image samples were inspected manually.

The analysis focused on:

- lighting variation

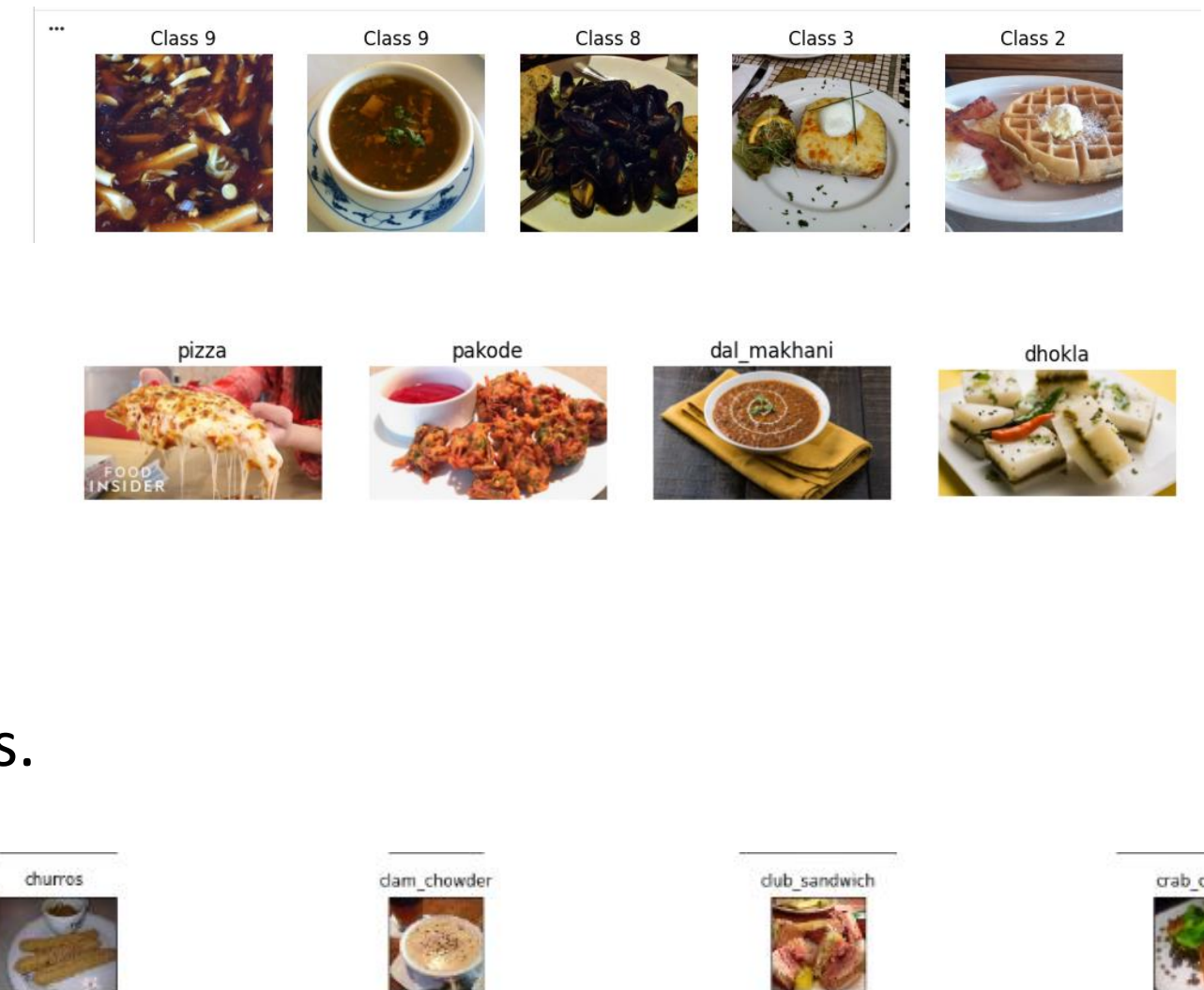
- camera angle variation

- background complexity

- food presentation style



This step helped interpret the later experimental results.



Data Preprocessing



- All images were resized to 224×224 pixels.
- Pixel values were normalized to the range 0 to 1.
- A validation split was used to separate training and validation data.
- These steps ensured consistency across all datasets and fair comparison between models.

Resized 224x224



Resized 224x224



Resized 224x224



Resized 224x224



Resized 224x224



Normalized



Normalized



Normalized



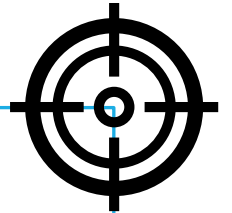
Normalized



Normalized



Data Augmentation



To reduce overfitting and improve generalization, augmentation was applied during training.

The augmentation methods used were:

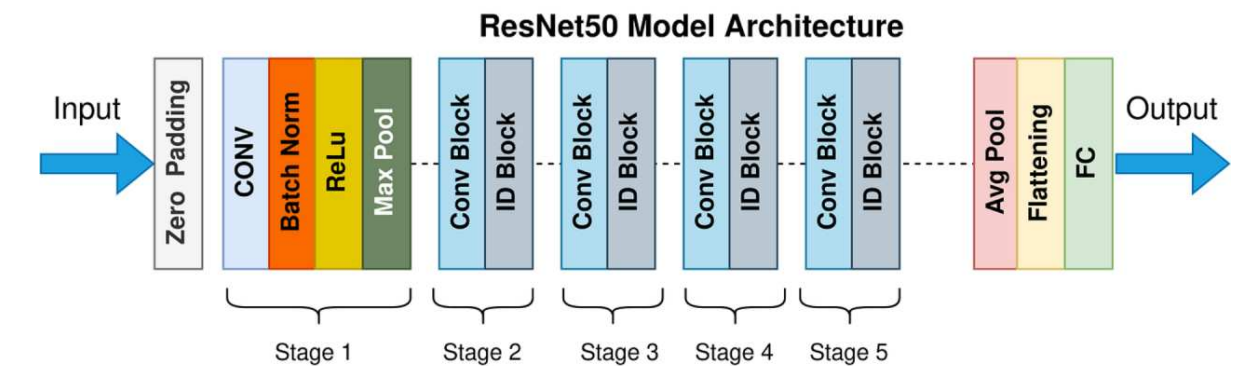
- image rotation
- horizontal flipping
- slight zoom in and zoom out
- These transformations simulate natural variations in real images.
- Strong random cropping was avoided so the food content would not be distorted too much.



Model Selection

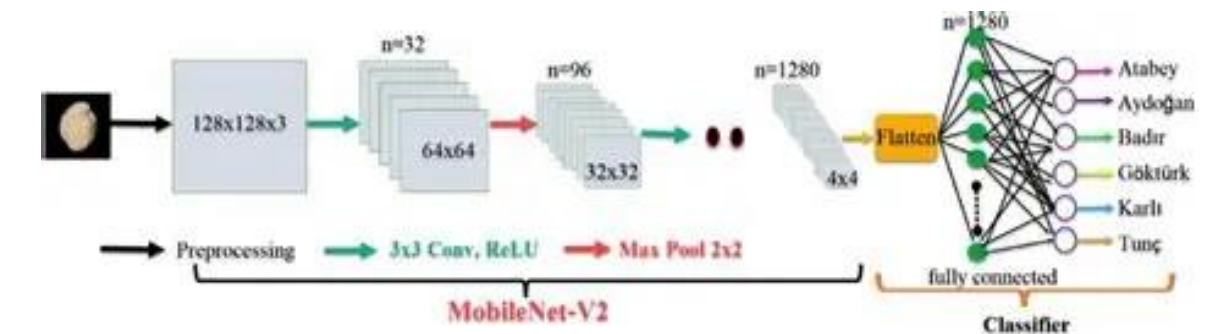
ResNet50

- Deep residual network
- Uses skip connections
- Helps avoid vanishing gradients and supports deeper learning



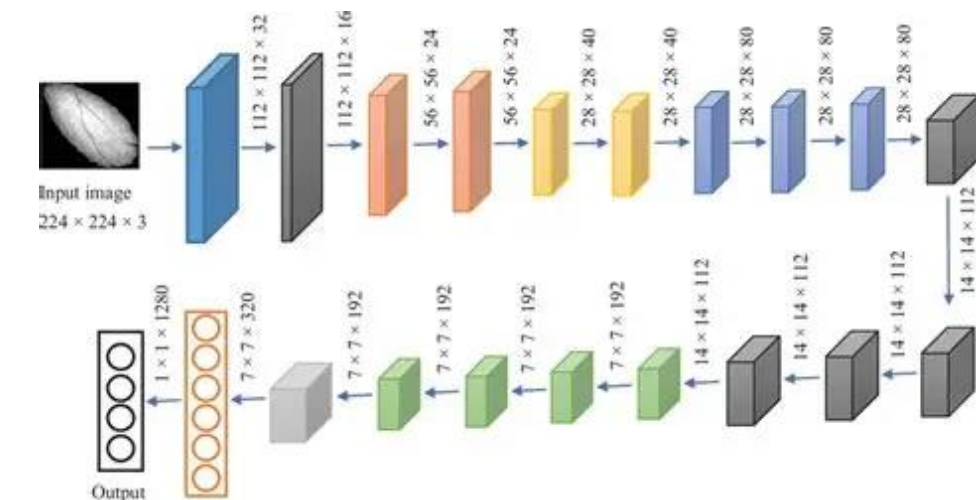
MobileNetV2

- Lightweight architecture
- Designed for resource-constrained systems
- Suitable for mobile and embedded deployment

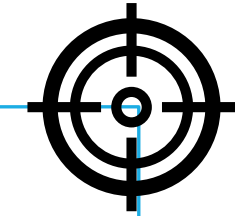


EfficientNetB0

- Uses compound scaling
- Balances depth, width, and resolution
- Offers strong accuracy with lower computational cost

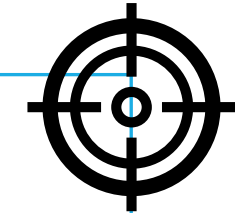


Training Configuration and Evaluation Metrics



- Same training setup was used for all models to keep the comparison fair.
- **Optimizer:** Adam
- **Loss function:** Categorical Cross-Entropy
- **Evaluation metric:** Accuracy
- **Batch size:** 32
- Besides accuracy, precision, recall, and F1-score were also discussed as useful evaluation measures.

Results on Food-11



Training Accuracy

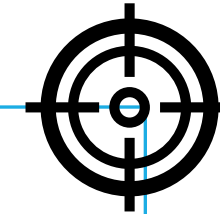
- ResNet50: **0.9937**
- MobileNetV2: **0.9272**
- EfficientNetB0: **0.9912**

Validation Accuracy

- ResNet50: **0.9239**
- MobileNetV2: **0.9093**
- EfficientNetB0: **0.9207**

- ❖ ResNet50 achieved the highest training accuracy.
- ❖ EfficientNetB0 was very close and showed strong balance.
- ❖ MobileNetV2 performed well but was slightly lower than the other two.
- ❖ The gap between training and validation accuracy shows some generalization difficulty.

Results on Food-34



Training Accuracy

- EfficientNetB0: **0.9864**
- ResNet50: **0.9861**
- MobileNetV2: **0.9850**

Best Validation Accuracy

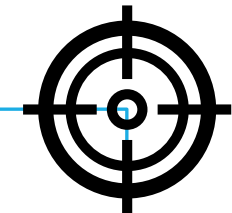
- EfficientNetB0: **0.9557**
- ResNet50: **0.9525**
- MobileNetV2: **0.9525**

EfficientNetB0 gave the best validation result.

ResNet50 and MobileNetV2 were very close.

EfficientNetB0 had the best balance of accuracy and efficiency.

Results on Food-101



Training Accuracy

- EfficientNetB0: **0.9864**
- ResNet50: **0.9899**
- MobileNetV2: **0.9144**

Best Validation Accuracy

- EfficientNetB0: **0.9494**
- ResNet50: **0.9367**
- MobileNetV2: **0.8892**

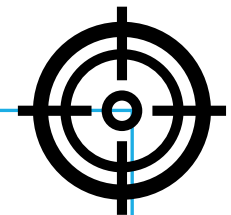
Food-101 was the hardest dataset because of its 101 categories and strong visual similarity.

EfficientNetB0 achieved the highest validation accuracy.

ResNet50 also performed well.

MobileNetV2 was the weakest on this dataset, but it remained the most lightweight option.

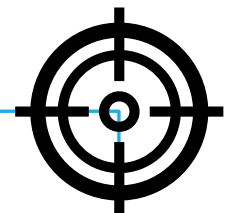
Comparative Analysis



Validation Accuracy Comparison

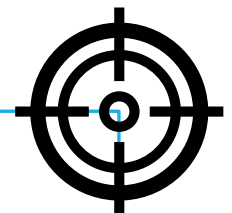
- **ResNet50:** 0.9239 on Food-11, 0.9525 on Food-34, 0.9367 on Food-101
- **MobileNetV2:** 0.9093 on Food-11, 0.9525 on Food-34, 0.8892 on Food-101
- **EfficientNetB0:** 0.9207 on Food-11, 0.9557 on Food-34, 0.9494 on Food-101
- EfficientNetB0 gave the most consistent overall performance.
- ResNet50 was also strong, especially on simpler and moderate datasets.
- MobileNetV2 was the most efficient model, but its accuracy dropped more on the hardest dataset.
- The results show that model choice should depend on the trade-off between accuracy and efficiency.

Discussion of Findings



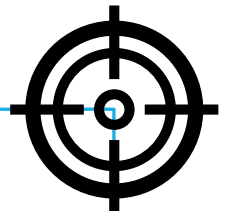
- As dataset complexity increased, performance became more difficult to maintain.
- All models did well on Food-11 because it is relatively easier.
- Food-34 showed that the models can still generalize strongly when the dataset becomes more diverse.
- Food-101 highlighted the challenge of fine-grained classification.
- EfficientNetB0 appears to be the best overall trade-off between accuracy and efficiency.
- MobileNetV2 is best when computational resources are limited.

Conclusion



- Deep learning is highly effective for food image recognition.
- Among the tested models, EfficientNetB0 delivered the best overall and most stable performance.
- ResNet50 also showed strong results due to its residual architecture.
- MobileNetV2 is attractive for deployment in resource-limited settings because of its efficiency.
- The study confirms that dataset complexity strongly affects classification performance.

Future Work



- Test more modern architectures such as attention-based models and state-space models.
- Use larger and more diverse food datasets.
- Improve performance on highly fine-grained categories.
- Explore explainable AI methods to better understand model decisions.
- Study real-world deployment in mobile nutrition and smart food systems.

References

- Farahnakian, F., et al. (2023). *Deep learning architectures for agricultural image classification*.
- Liu, Y., et al. (2025). *A comprehensive review of deep learning for food image recognition*.
- Rasyidi, M. A., et al. (2024). *Comparative study of deep learning architectures for food image classification*.
- Srivastava, S., et al. (2021). *Comparative analysis of object detection algorithms using deep learning*.
- Suddul, G., & Seguin, B. (2023). *Food image classification using deep learning and data augmentation*.
- Terzioğlu, Y., et al. (2025). *Deep learning-based feature extraction for agricultural image classification*.
- Farahnakian, F., Sheikh, J., Farahnakian, F., & Heikkonen, J. (2023). A comparative study of state-of-the-art deep learning architectures for rice grain classification. *Journal of Agriculture and Food Research*, 13, 100890.
- Suddul, G., & Seguin, J. F. L. (2023). A comparative study of deep learning methods for food classification with images. *Food and Humanity*, 1, Article 100018.

THANK YOU