Contents

Deep Learning-Based Classification of Healthy Food Images Using Custom and Pretrained Models

By Shrutika Gupta (2448387)

Abstract

Recent efforts in the automation of dietary assessment underline the urgent need for accurate nutrition monitoring, motivated in part by rising rates of diet-related diseases. Central to these systems, image-based food recognition faces persistent hurdles when differentiating healthy foods, primarily driven by high intra-class variability, uneven representation across categories, and frequently constrained dataset sizes. We investigate the classification of healthy food images across twelve subclasses, deploying and comparing several deep learning architectures: a novel, custom Convolutional Neural Network (CNN), and established transfer-learning backbones—ResNet50, EfficientNetB0, and MobileNetV2. Calculated class weights are used during model training to address the notable class imbalance in the dataset. Metrics such as accuracy, precision, recall, and F1-score, were used for the evaluation of the model performance. The results show that transfer learning works well for small-scale, domain-specific food classification tasks, with pretrained models—in particular, EfficientNetB0—performing noticeably better than the custom CNN.

***Keywords**: Healthy Food Classification, Deep Learning, CNN, Transfer Learning, EfficientNetB0, Class Imbalance, Image Recognition, Dietary Assessment.*

1. Introduction

Encouragement of healthy eating habits is more crucial than ever in a period when lifestyle diseases are on the rise. Automated food recognition systems have become a critical tool for dietary monitoring where they enable the quick and efficient assessment of food choices by individuals or health professionals. At the heart of these systems is the tough problem of food classification based only on visual appearance-a problem that needs robust image classification models that are able to generalize to various categories.

This study focuses on deep learning-based methods to classify images of healthy foods into categories.

Items from the same nutritional group can look extremely similar depending on the presentation style, background, and lighting, which is why food images are especially challenging to categorise. Unlike general object recognition, food classification must take into account finer visual cues such as texture, color gradients, and shape that could be only slightly different across categories-which differ from more general object recognition scenarios, thus making food image classification a challenging concern in the field of deep learning.

Since deep learning is beginning to extend the horizons of computer vision, this field of modern technology has useful promise to its application in healthy food classification. In perfectly constructed datasets and thoughtfully orchestrated models, an automated system can be trained to recognize and classify many varieties of healthy food, helping an end user inculcate better dietary choices.

2. Problem Statement

Because there are many image-based assessment systems that are becoming popular, obtaining high classification accuracies for the healthy food items has proven difficult due to:

- Intra-class similarity among food categories (e.g., fruits vs. vegetables),
- Class imbalance in available datasets,
- Limited interpretability of deep models in the food domain.

There arises a need to compare and contrast deep networks-adapted and pretrained-based classifying various healthy food items, especially if they were trained over smaller and/or imbalanced datasets.

3. Objectives

1. To explore and tackle the issue of class imbalance in a multi-class dataset of healthy food images.
2. To create a baseline CNN model and assess how well it performs in classifying images of healthy foods.
3. To implement and compare various transfer learning models (like ResNet50, EfficientNetB0, and MobileNetV2) to enhance classification accuracy.
4. To visualize training metrics such as accuracy and loss, along with predictions, to gain insights into model behavior and interpret the results.
5. To study performance metrics like accuracy, precision, recall, and F1-score to identify the most effective deep learning models for classifying healthy food images.

4. Literature Review

While no significant prior research has been conducted specifically using the dataset employed in this study, two notable works in the broader domain of food image classification offer valuable insights and benchmarks, as discussed below.

One notable study, titled **"Healthy vs. Unhealthy Food Images: Image Classification of Twitter Data"** by Tejaswini et al. (2022)[2], applied ResNet-152 to differentiate between healthy and unhealthy images. The model was trained on web-sourced food images and evaluated on real-world Twitter data where it achieved an impressive overall accuracy of 77.25% and an F1 score of 78.78% , reflecting the model's ability to generalize to diverse, noisy inputs.

However, this study faced some limitations, such as overly broad binary class definitions, a lack of structured datasets, and no comparative analysis across various models, since only one pretrained model (ResNet-152) was evaluated, with no comparison to other architectures such as EfficientNet, MobileNet, or custom CNNs. On the other hand, our current study aims to fill these gaps by conducting a detailed multi-class classification of healthy foods. We are using an organized and balanced dataset, evaluating several architectures (including CNN, ResNet50, EfficientNetB0, and MobileNetV2), and adding visualizations to enhance interpretability.

Then, in 2016, Liu et al. introduced **"DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment"**[3], a unique deep learning method for food image identification, tested on two extensive datasets—Food-101 and UEC-256. By utilizing a specially designed CNN architecture that features Inception-style modules and transfer learning, this model achieved an impressive top-1 classification accuracy of around 77.4% on Food-101, surpassing previous records of about 56% on UEC-256.

However, this research was concentrating solely on single-item food images, not comparing different architectures, and lacking visual interpretability of the results. The current study aims to fill these gaps by focusing on multi-class classification of healthy food types, creating a well-structured and balanced dataset, comparing various models (CNN, ResNet50, EfficientNetB0, MobileNetV2), and providing a thorough visualization of the training dynamics and prediction behaviors.

5. Methodology

Using both custom-built and pretrained convolutional neural network (CNN) models, this study employs a systematic approach to deep learning with the goal of classifying images of healthy foods into various groups. Preparing the data, managing class imbalances, creating the model, training it, assessing its effectiveness, and visualising the outcomes are some of the crucial steps in the process.

5.1. Dataset

The dataset from Kaggle by Rafly Ramadan presents images of healthy food categories [1], classifiable into subfolders per category, ideal for training image classifiers. The dataset contains 1568 images in the following 12 food categories, along with their respective class counts:

(i) Avocado Toast - 158
(ii) Boiled Eggs - 179
(iii) Fruit Salad - 69
(iv) Grilled Chicken Breast - 140
(v) Oatmeal - 173
(vi) Quinoa Salad - 77
(vii) Roasted Vegetables - 106
(viii) Smoothie Bowl - 56
(ix) Steamed Tempeh - 110
(x) Steamed Tofu - 173
(xi) Tuna Salad - 169
(xi) Vegetable Salad - 158



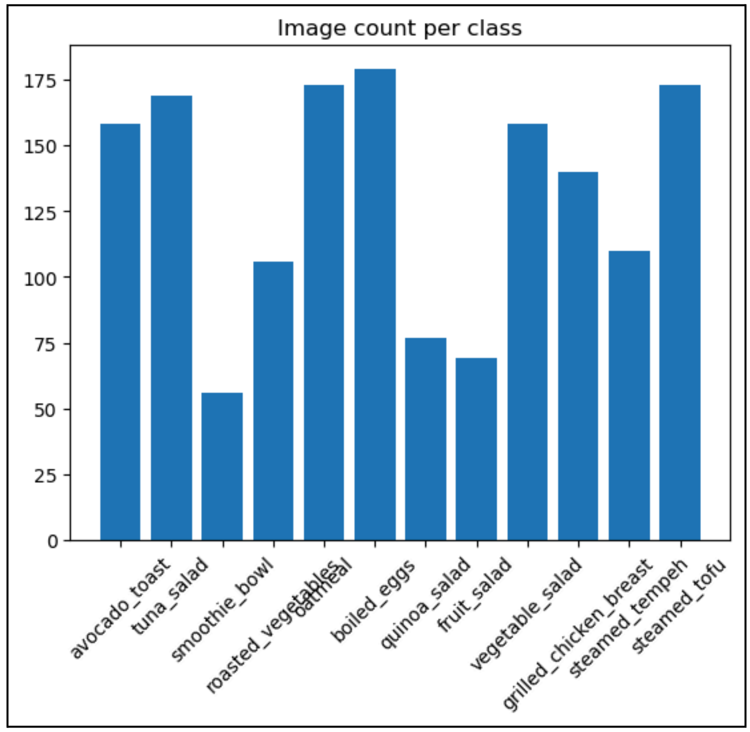

5.2. Dataset Preparation

(i) The labelled photos of healthy food items in the dataset are arranged into subdirectories that correspond to each class.
(ii) TensorFlow's "image_dataset_from_directory" tool is used to load the dataset, automatically dividing the images into training and validation sets.
(iii) To ensure input compatibility with pretrained models, all images are resized to a standard shape of 224 × 224 × 3.

5.3. Class Distribution Analysis & Balancing

The number of images in each category is counted in order to examine the class imbalance. To ensure that minority classes are not under-represented in model learning, class weights are applied during training if there is a significant imbalance. This ensures balanced model performance during training by adjusting each class's contribution to the overall loss.

Class Label	Class Count	Class Weight
avocado_toast	158	0.82
tuna_salad	169	0.68
smoothie_bowl	56	1.94
roasted_vegetables	106	1.04
oatmeal	173	0.76
boiled_eggs	179	1.71
quinoa_salad	77	1.25
fruit_salad	69	2.18
vegetable_salad	158	1.15
grilled_chicken_breast	140	0.77
steamed_tempeh	110	0.78
steamed_tofu	173	0.81



5.4. Model Architectures

The study compares four model architectures:
(i) A custom CNN built from scratch using standard convolutional, pooling, and dense layers.
(ii) Three transfer learning models: ResNet50, EfficientNetB0, and MobileNetV2, each initialized with pretrained ImageNet weights and modified by adding global average pooling, dropout, and dense output layers suited to the number of classes.
(iii) All pretrained models use their respective preprocessing functions to ensure input compatibility with the backbone architecture.

5.5. Training Procedure

(i) Each model is trained using the '*Adam*' optimizer and the '*sparse_categorical_crossentropy*' loss function.
(ii) The training is conducted over 10 epochs, with validation accuracy and loss monitored at each step.
(iii) Class weights are used to balance imbalanced data during training.
(iv) The models are trained using the same train/validation split for fair comparison.

5.6. Model Evaluation

After training, each model is evaluated on the validation dataset using:

- Accuracy and loss metrics
- Classification reports including precision, recall, and F1-score for each class
- Prediction visualizations, where a subset of images is displayed alongside true and predicted labels
- Training curves showing accuracy and loss over epochs, aiding in analysis of convergence and overfitting

5.7. Visualization Tools

To improve model interpretability and analysis:

- Matplotlib is used to plot training vs. validation accuracy and loss for each model.
- Sample predictions are visualized to assess qualitative performance.
- Confusion matrices and classification reports further aid in understanding model strengths and weaknesses.

6. Results

6.1. Basic CNN Evaluation

Train Accuracy:	0.3833
Train Loss:	1.8223
Validation Accuracy:	0.3674
Validation Loss:	1.8690

Classification Report:

	precision	recall	f1-score	support
avocado_toast	0.3333	0.0323	0.0588	31
boiled_eggs	0.7000	0.8077	0.7500	26
fruit_salad	0.2143	0.2000	0.2069	15
grilled_chicken_breast	0.3529	0.1538	0.2143	39
oatmeal	0.4130	0.5429	0.4691	35
quinoa_salad	0.3409	0.9375	0.5000	16
roasted_vegetables	0.3793	0.5000	0.4314	22
smoothie_bowl	0.0811	0.3750	0.1333	8
steamed_tempeh	0.2500	0.1053	0.1481	19
steamed_tofu	0.4118	0.3684	0.3889	38
tuna_salad	0.3889	0.2000	0.2642	35
vegetable_salad	0.3939	0.4483	0.4194	29
accuracy			0.3674	313
macro avg	0.3550	0.3893	0.3320	313
weighted avg	0.3829	0.3674	0.3411	313

6.2. ResNet50 Evaluation

Train Accuracy:	0.9737
Train Loss:	0.1449
Validation Accuracy:	0.8147
Validation Loss:	0.7092

Classification Report:

	precision	recall	f1-score	support
avocado_toast	0.8056	0.9355	0.8657	31
boiled_eggs	1.0000	0.9231	0.9600	26
fruit_salad	0.7333	0.7333	0.7333	15
grilled_chicken_breast	0.9231	0.9231	0.9231	39
oatmeal	0.8333	0.8571	0.8451	35
quinoa_salad	0.9333	0.8750	0.9032	16
roasted_vegetables	0.7600	0.8636	0.8085	22
smoothie_bowl	0.4444	0.5000	0.4706	8
steamed_tempeh	0.6250	0.5263	0.5714	19
steamed_tofu	0.8750	0.7368	0.8000	38
tuna_salad	0.8485	0.8000	0.8235	35
vegetable_salad	0.6667	0.7586	0.7097	29
accuracy			0.8147	313
macro avg	0.7874	0.7860	0.7845	313
weighted avg	0.8195	0.8147	0.8148	313

6.3. EfficientNetB0 Evaluation

Train Accuracy:	0.9267
Train Loss:	0.3121
Validation Accuracy:	0.8019
Validation Loss:	0.6378

Classification Report:

	precision	recall	f1-score	support
avocado_toast	0.8235	0.9032	0.8615	31
boiled_eggs	1.0000	0.9231	0.9600	26
fruit_salad	0.7222	0.8667	0.7879	15
grilled_chicken_breast	0.8974	0.8974	0.8974	39
oatmeal	0.9355	0.8286	0.8788	35
quinoa_salad	0.6667	0.8750	0.7568	16
roasted_vegetables	0.8696	0.9091	0.8889	22
smoothie_bowl	0.7500	0.7500	0.7500	8
steamed_tempeh	0.5217	0.6316	0.5714	19
steamed_tofu	0.8667	0.6842	0.7647	38
tuna_salad	0.7188	0.6571	0.6866	35
vegetable_salad	0.7000	0.7241	0.7119	29
accuracy			0.8019	313
macro avg	0.7893	0.8042	0.7930	313
weighted avg	0.8122	0.8019	0.8035	313

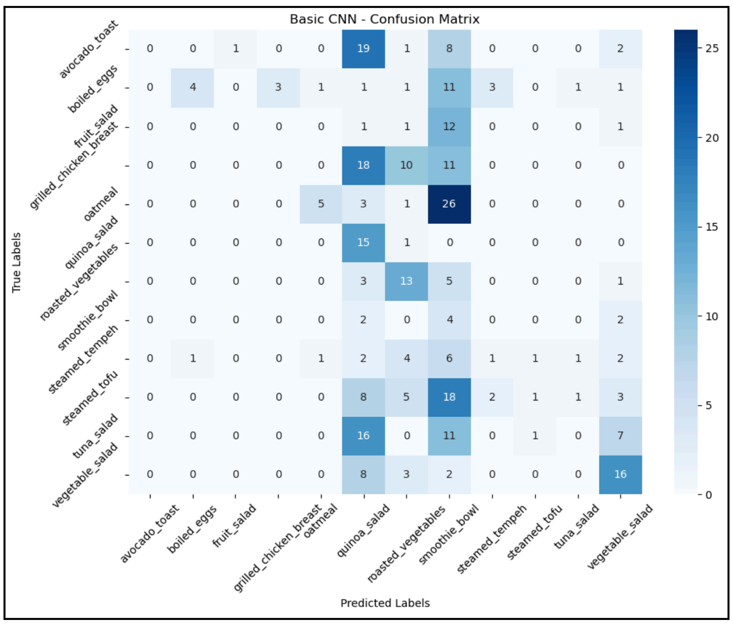
6.4. MobileNetV2 Evaluation

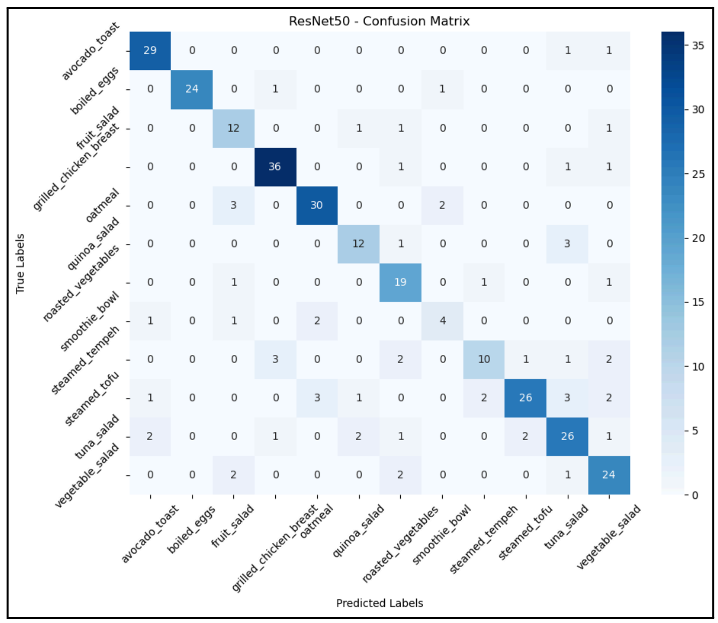
Train Accuracy:	0.9434
Train Loss:	0.2823
Validation Accuracy:	0.7125
Validation Loss:	0.7997

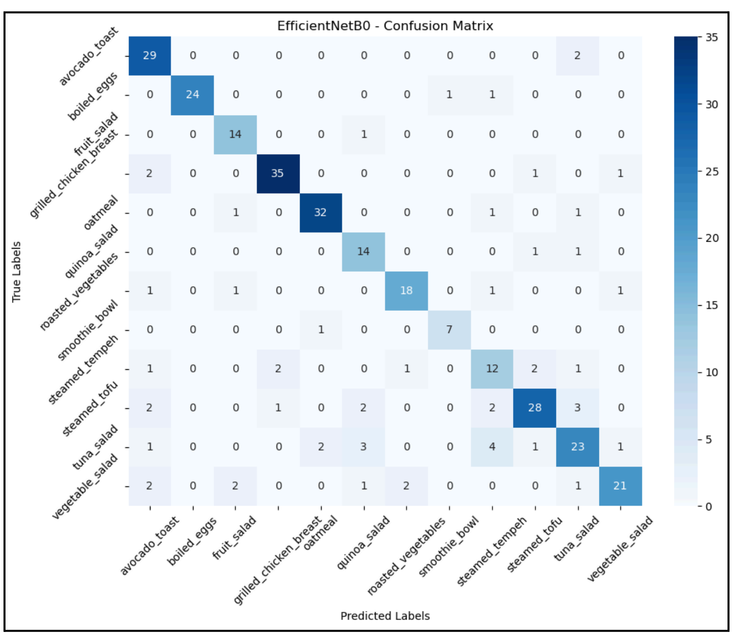
Classification Report:

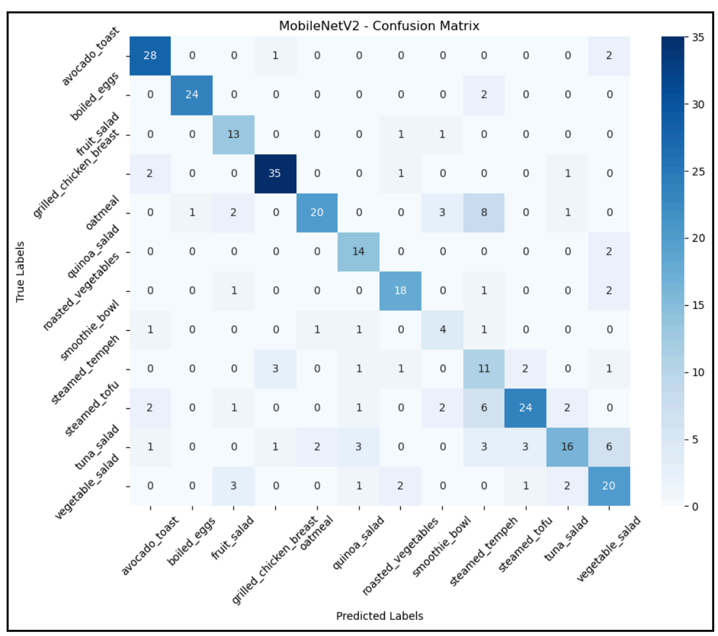
	precision	recall	f1-score	support
avocado_toast	0.8710	0.8710	0.8710	31
boiled_eggs	0.9600	0.9231	0.9412	26
fruit_salad	0.7857	0.7333	0.7586	15
grilled_chicken_breast	0.8684	0.8462	0.8571	39
oatmeal	0.8750	0.6000	0.7119	35
quinoa_salad	0.7778	0.8750	0.8235	16
roasted_vegetables	0.7500	0.8182	0.7826	22
smoothie_bowl	0.2500	0.5000	0.3333	8
steamed_tempeh	0.3548	0.5789	0.4400	19
steamed_tofu	0.7667	0.6053	0.6765	38
tuna_salad	0.7273	0.4571	0.5614	35
vegetable_salad	0.5250	0.7241	0.6087	29
accuracy			0.7125	313
macro avg	0.7093	0.7110	0.6972	313
weighted avg	0.7532	0.7125	0.7208	313

7. Confusion Matrices

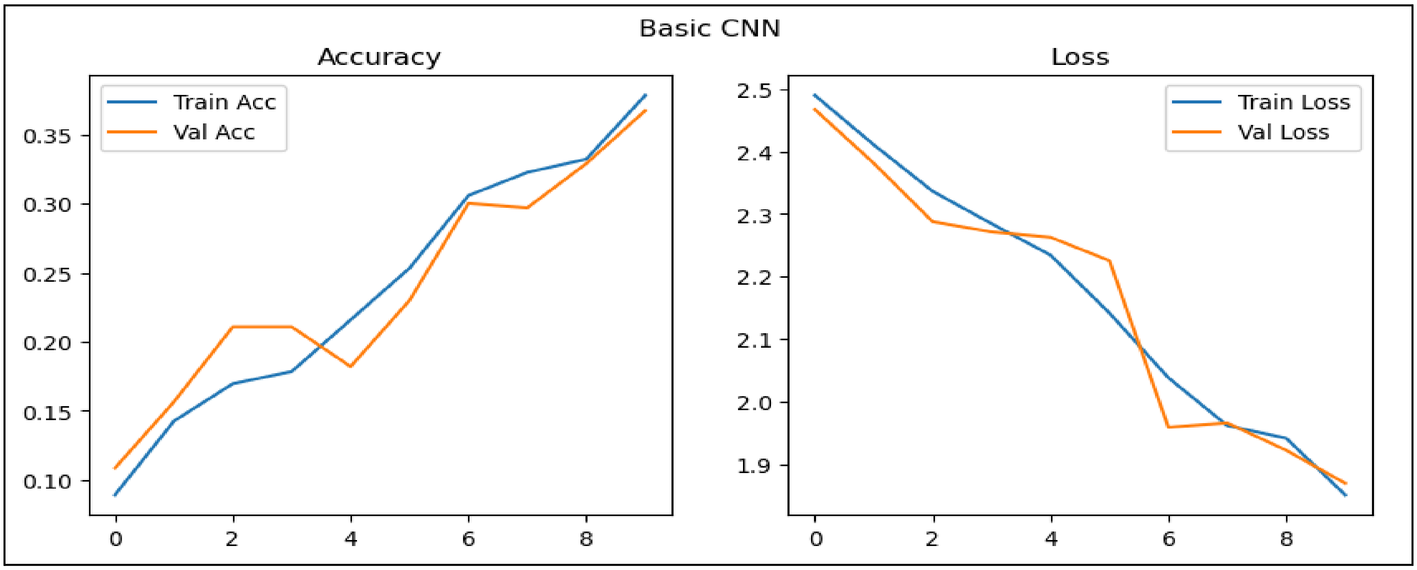


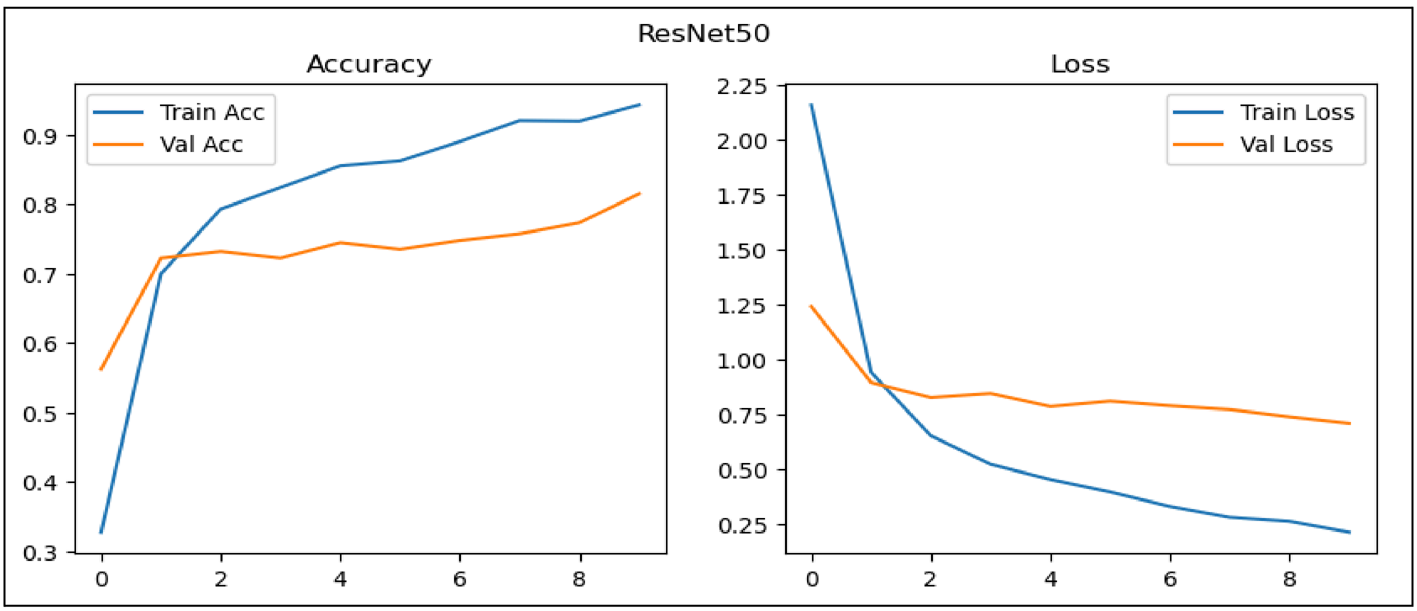


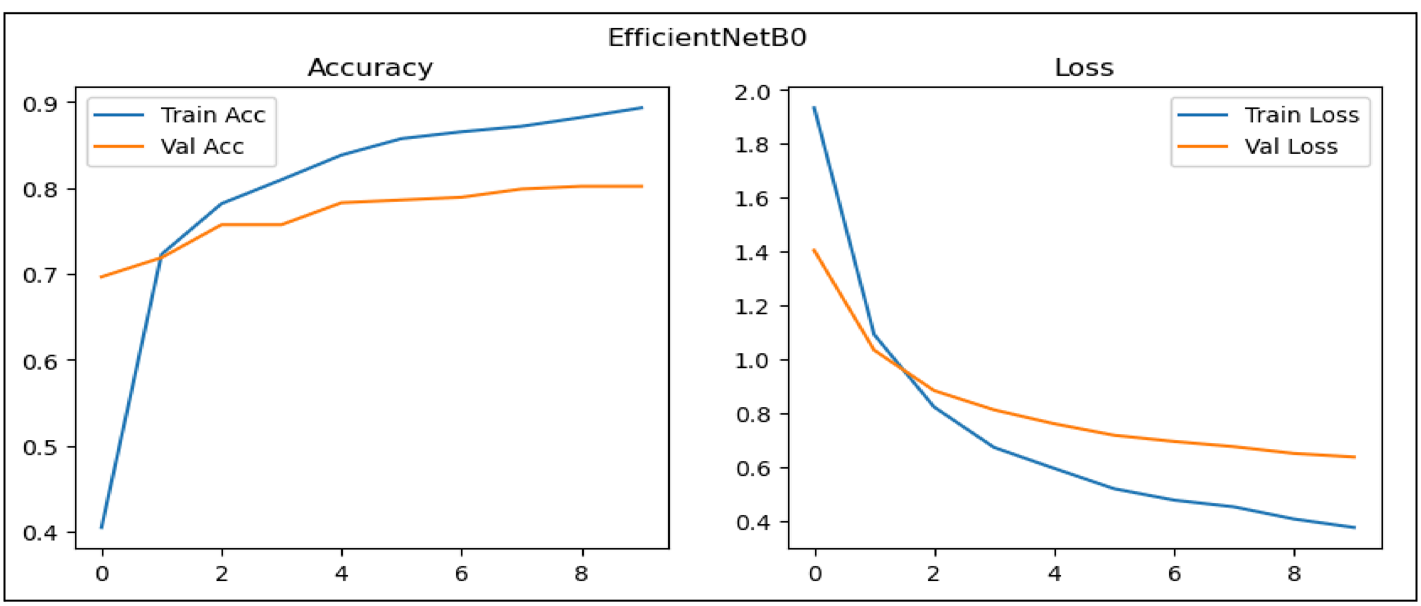


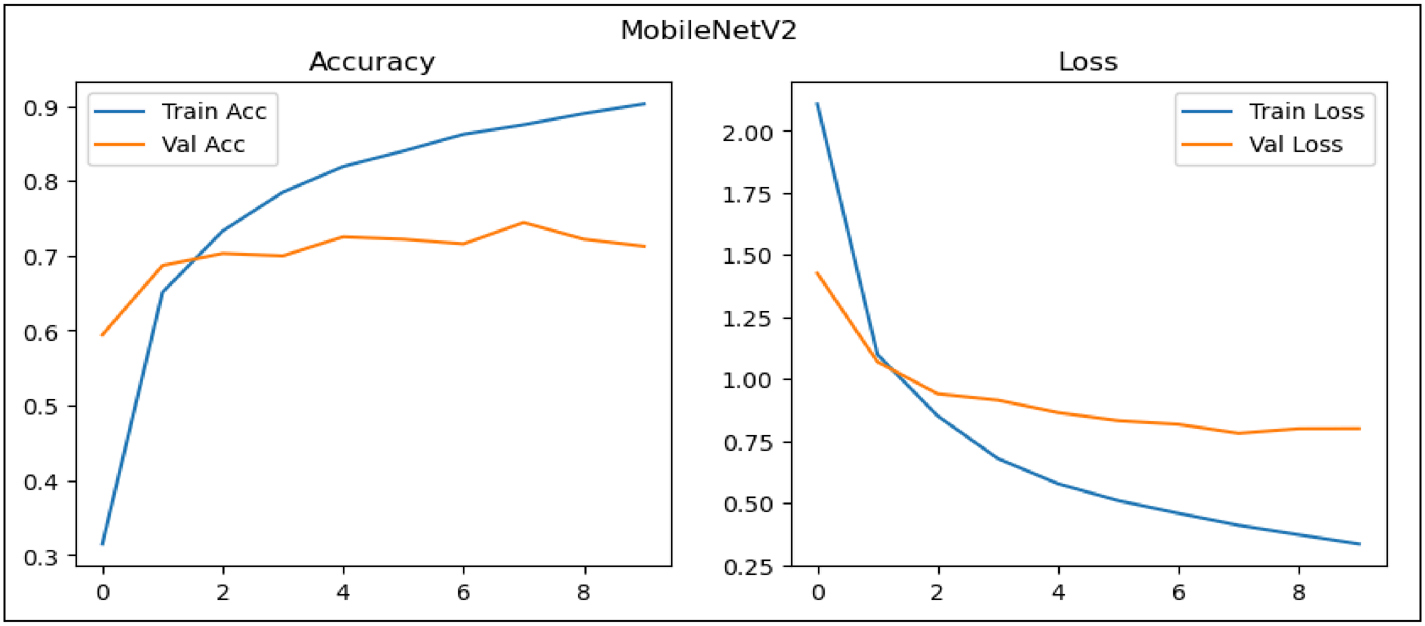


8. Visualizations









9. Analysis

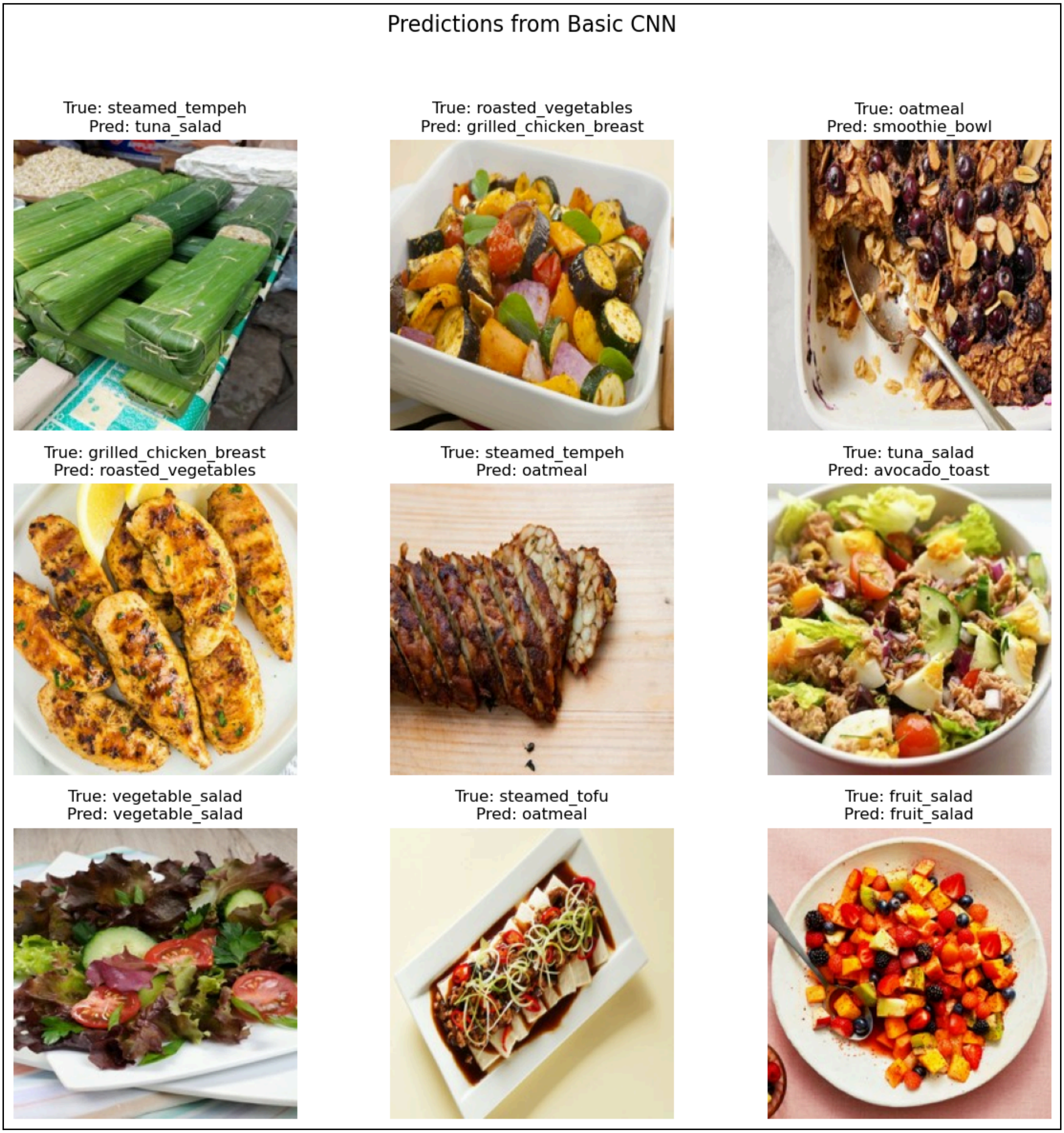
Four models—a basic CNN, ResNet50, EfficientNetB0, and MobileNetV2— were tested to determine how well different deep learning architectures can classify healthy food items. The same dataset was used for all models, and the accuracy, loss, precision, recall, and F1-score were assessed. The baseline CNN did not perform well, achieving only 36.74% validation accuracy and a high loss of 1.8690. It struggled with the dataset's class imbalance and the complexity of visual patterns, failing to generalize for less common classes like avocado_toast (F1 = 0.05) and smoothie_bowl (0.13).

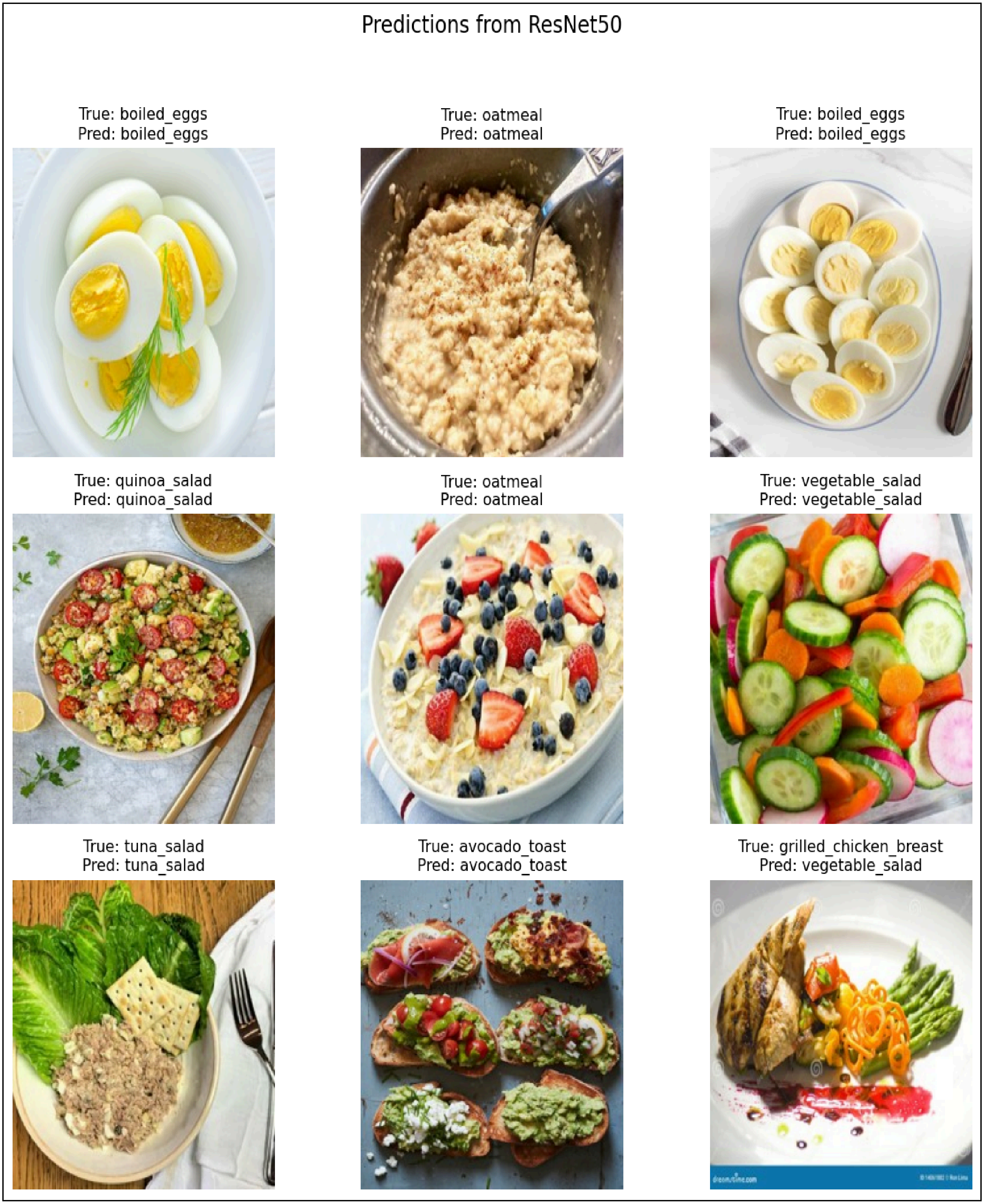
On the other hand, ResNet50, which is a deeper pretrained architecture, excelled with the best performance, reaching 81.47% accuracy and showing consistently high scores across most classes, including boiled_eggs (F1 = 0.96) and grilled_chicken_breast (0.92). EfficientNetB0 was a close second with 80.19% accuracy and a slightly lower loss of 0.6378, maintaining balanced results across both common and smaller classes like fruit_salad (0.79) and roasted_vegetables (0.89). On the other hand, MobileNetV2, being lighter and faster, managed to achieve 71.25% accuracy but had difficulties with rare categories such as smoothie_bowl (0.33) and steamed_tempeh (0.44).

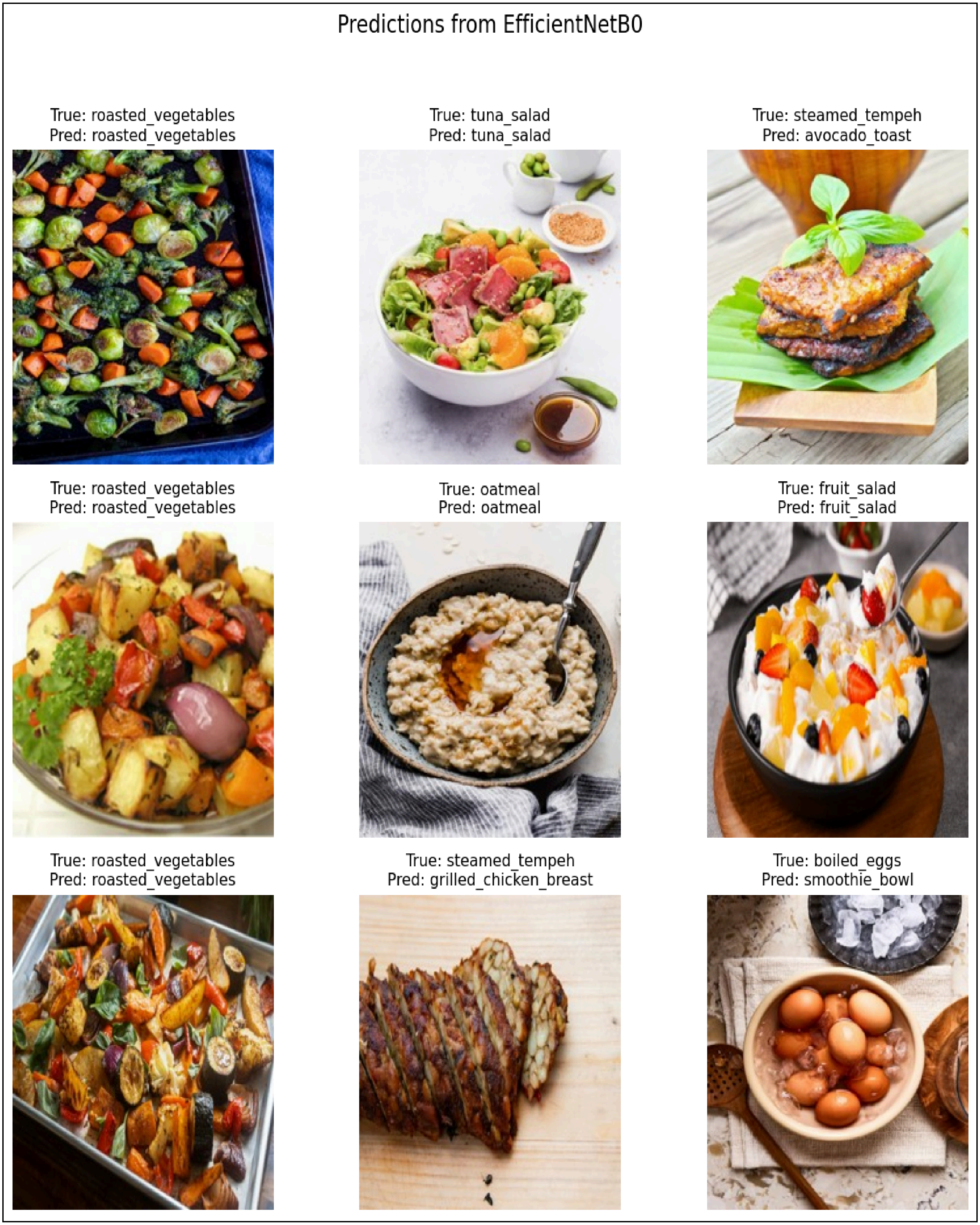
The confusion matrices shed more light on these findings. The CNN had a lot of misclassifications, often mixing up visually similar categories like smoothie_bowl with fruit_salad or avocado_toast with bruschetta. On the other hand, the pretrained models like ResNet50 and EfficientNetB0 made lesser errors, with most misclassifications occurring between classes that share similar features. MobileNetV2 fell in the middle, making fewer mistakes than the CNN but still showing significant confusion for the less represented categories. These trends highlight that better feature extraction in pretrained models leads to more accurate differentiation between visually similar food items.

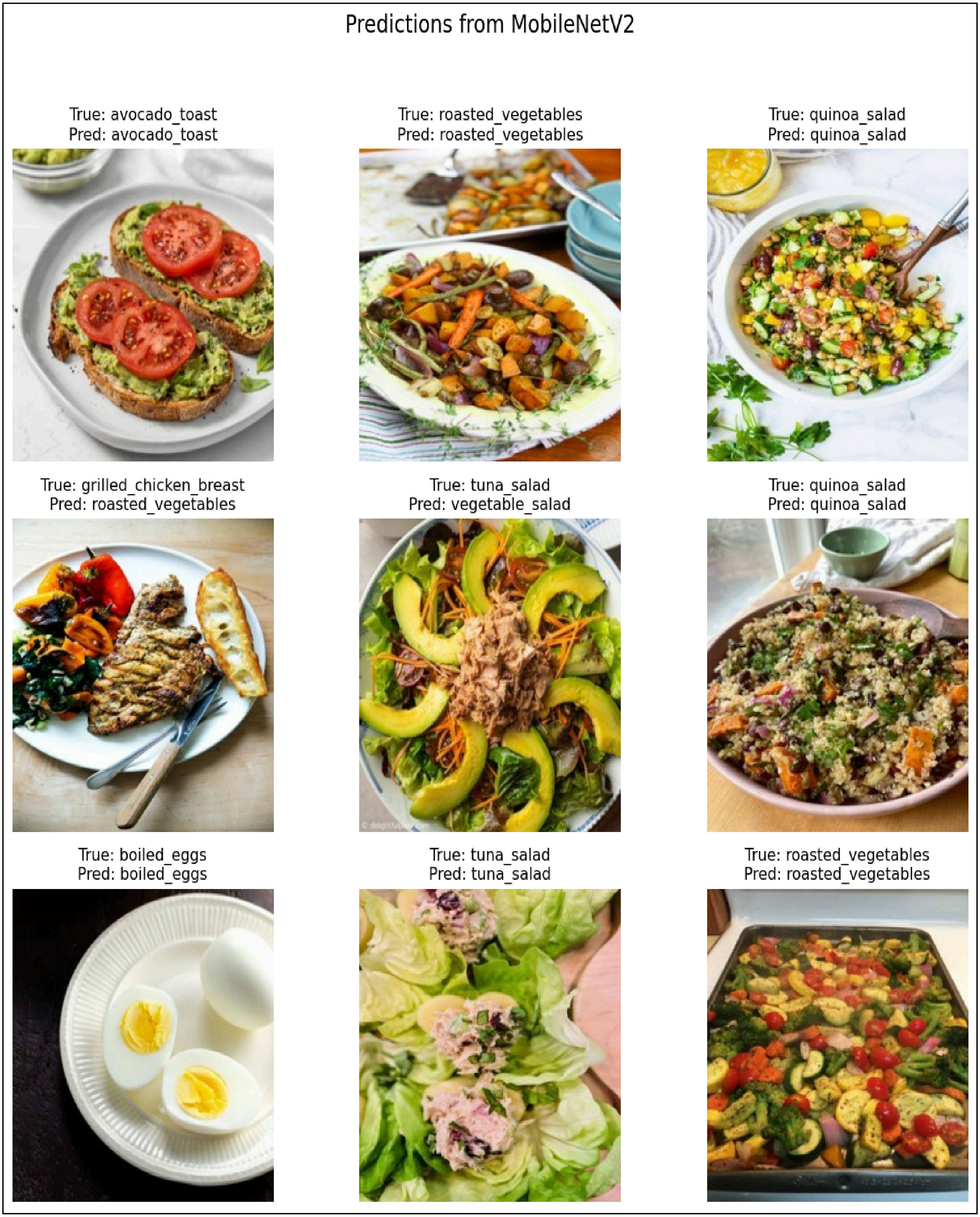
The results show that pretrained models have a significant improvement over basic CNN, with ResNet50 and EfficientNetB0 getting the best results. Although using weights helped to fix the class imbalance, some rare classes like smoothie_bowl and fruit_salad still posed a challenge. With additional fine-tuning, focused augmentation for the less represented categories, and perhaps the inclusion of attention-based mechanisms, these pretrained architectures could greatly enhance the classification performance when it comes to identifying healthy food images.

10. Predictions




Predictions from ResNet50
True: boiled_eggs
Pred: boiled_eggs
True: oatmeal
Pred: oatmeal
True: boiled_eggs
Pred: boiled_eggs
True: quinoa_salad
Pred: quinoa_salad
True: oatmeal
Pred: oatmeal
True: vegetable_salad
Pred: vegetable_salad
True: tuna_salad
Pred: tuna_salad
True: avocado_toast
Pred: avocado_toast
True: grilled_chicken_breast
Pred: vegetable_salad


Predictions from EfficientNetB0
True: roasted_vegetables
Pred: roasted_vegetables
True: tuna_salad
Pred: tuna_salad
True: steamed_tempeh
Pred: avocado_toast
True: roasted_vegetables
Pred: roasted_vegetables
True: oatmeal
Pred: oatmeal
True: fruit_salad
Pred: fruit_salad
True: roasted_vegetables
Pred: roasted_vegetables
True: steamed_tempeh
Pred: grilled_chicken_breast
True: boiled_eggs
Pred: smoothie_bowl


Predictions from MobileNetV2
True: avocado_toast
Pred: avocado_toast
True: roasted_vegetables
Pred: roasted_vegetables
True: quinoa_salad
Pred: quinoa_salad
True: grilled_chicken_breast
Pred: roasted_vegetables
True: tuna_salad
Pred: vegetable_salad
True: quinoa_salad
Pred: quinoa_salad
True: boiled_eggs
Pred: boiled_eggs
True: tuna_salad
Pred: tuna_salad
True: roasted_vegetables
Pred: roasted_vegetables

11. Limitations and Future Scope

This study highlights the application of deep learning models like ResNet50 and EfficientNetB0, in classifying images of healthy foods. However, there are a few limitations in this study:

1. ***Limited Dataset Size***: The dataset has a relatively small number of images per class, with certain categories being particularly underrepresented. This imbalance affected how well the model performed on these less common classes, even after we applied class weights.
2. ***Visual Similarity Among Classes***: A lot of healthy food items have similar colors and textures, making it hard for the models to correctly classify them.
3. ***Lack of Real-World Diversity***: The training images were likely captured under ideal lighting and composition conditions, because of which the models struggle to perform well on real-world photos taken in less-than-perfect situations, like poor lighting or cluttered backgrounds.
4. ***Interpretability***: We did not focus on using XAI techniques like Grad-CAM or SHAP to clarify which visual features influenced predictions, although this aspect is essential for building trust in AI systems related to health.

Future research can build upon this work in the following ways:

1. ***Larger and More Diverse Datasets***: By diversifying the dataset to include more samples for each class and incorporating a variety of backgrounds, lighting conditions, and food presentations, generalization can be enhanced and bias can be minimised.
2. ***Data Augmentation and Synthesis***: Utilizing advanced augmentation techniques or creating synthetic samples with GANs can help balance the dataset and boost its robustness.
3. ***Model Explainability***: Incorporating XAI techniques can improve the model's transparency and understanding.
4. ***Multi-label and Nutritional Analysis***: Future studies could delve into multi-label classification (like identifying ingredients) and connect image-based predictions to nutritional information, allowing for more thorough dietary evaluations.
5. ***Real-Time Deployment***: Fine-tuning models for mobile or embedded use (through techniques like quantization or pruning) can make healthy food recognition systems practical for real-time applications in dietary tracking apps.

12. Conclusions

In this study, we investigated the use of deep learning techniques to categorise images of nutritious foods into 12 distinct groups. The development of precise food recognition systems has emerged as a major area of research interest due to the increasing significance of monitoring our diets in order to prevent lifestyle-related illnesses. This study evaluated the ability of robust pretrained models such as ResNet50, EfficientNetB0, and MobileNetV2 as well as custom convolutional neural networks (CNNs) to generalise across visually similar food items despite class imbalance and a small dataset.

Based on the findings, the pretrained models, especially EfficientNetB0, performed better than the custom CNN in terms of generalisation and accuracy. This demonstrates how effective transfer learning can be when working with smaller datasets of food images. The models produced remarkable classification results despite obstacles like class similarities and a lack of diversity in the real world, especially when class weight strategies were added to correct the imbalance.

By explaining the advantages and disadvantages of the existing deep learning techniques for categorising healthful foods, this study contributes significantly to automated dietary assessment tools. These systems have the potential to become crucial components of individualised nutrition and health management with continued improvements in data diversity, explainability, and real-time application.

13. References

1. Oduru, Tejaswini et al. "Healthy vs. Unhealthy Food Images: Image Classification of Twitter Images." International journal of environmental research and public health vol. 19,2 923. 14 Jan. 2022, doi:10.3390/ijerph19020923. https://pmc.ncbi.nlm.nih.gov/articles/PMC8775411/
2. Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., & Ma, Y. (2016, June 17). DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment. arXiv.org. https://arxiv.org/abs/1606.05675

*****