

INGREDINET: INTELLIGENT CNN FOR FOOD INGREDIENT RECOGNITION AND CLASSIFICATION

Mahnoor Zaman^{*1}, Nosheen Fatima², Muhammad Sajid Maqbool^{*3}, Dr. Naeem Aslam⁴, Rubaina Nazeer^{*5}, Hira Saleem⁶

^{*1,*3,*6}Department of Computer Science, NFC Institute of Engineering and Technology, Multan

²Department of Software engineering, Numl university multan

^{*5}Department of Software engineering, Numl university multan

¹786mahnoorzaman@gmail.com, ²nosheenfatima@numl.edu.pk, ^{*3}sajid.maqbool@nfciet.edu.pk, ⁴naeem.aslam@nfciet.edu.pk, ^{*5}rubaina.nazeer@numl.edu.pk, ⁶hira.saleem@nfciet.edu.pk

DOI: <http://doi.org/10.5281/zenodo.19381904>

Keywords

CNN, Food Ingredient Classification, Deep Learning, Data Augmentation, Computer Vision, Image Processing, Artificial Intelligence, Image Classification, Food Recognition

Article History

Received: 31 January 2026

Accepted: 14 March 2026

Published: 31 March 2026

Copyright @Author

Corresponding Author: *

Muhammad Sajid Maqbool,
Rubaina Nazeer,
Mahnoor Zaman

Abstract

Accurate recognition of food ingredients is the basis of smart kitchen, nutritional management, automated inventory management, and intelligent food analysis. Traditional methods of identification are intensive in nature and based on manual inspection that is time consuming, labor intensive and prone to human errors. With the evolution of food technology, the need to have automated solutions that operate using vision is on the rise in order to provide accurate and scalable ingredient recognition. Although deep learning has been proposed as a highly successful method in general image classification problems, relatively little focus has been placed on designing lightweight and task-specific convolutional neural networks to perform multi-class ingredient recognition using a combination of authentic image data as the primary training objective. This project proposes IngridiNet, a specially designed CNN architecture that was designed to conduct effective and high-precision ingredient categorizing. It is estimated that the research will accomplish the following goals: (i) create and optimize a lightweight CNN, (ii) compare the performance of the model with the pretrained MobileNet model, and (iii) examine the effect of data augmentation on the model accuracy and generalization. The IngridiNet was trained on a multi-class image dataset of ingredients with two experimental conditions, in the case of original, unaugmented images and in the case of augmented data with random horizontal flipping, rotation, zoom, and contrast variation. Validation accuracy, loss curves, precision, recall, F1-score, and confusion matrix analysis were used to compare model performance in terms of overall performance and reliability by classes. It was compared to MobileNet to evaluate the performance, robustness and computing efficiency. The experimental findings indicate that IngridiNet has reached its highest validation rate of 99.84% when trained on original images which is marginally higher than at MobileNet where the validation rate stood at 98.87% under the same conditions. Interestingly, the two models demonstrated strong generalization on the validation data without data augmentation, which means that realistic images can be used to achieve the high-performance classification. Even though augmentation resulted in a slight decrease in the

accuracy per peak (IngrediNet: 98.28%, MobileNet: 97.66%), it made the models more resistant to changes in orientation, scale, and lighting. Viewed as a computationally efficient and highly accurate engine of automated ingredient recognition, IngrediNet is suitable, in terms of its deployment, in resource-constrained, real-world environments, e.g., smart kitchens and mobile applications. As the results prove, a well-designed custom CNN can be as good or even better than the pretrained architectures and be less complex, resulting in the further development of practical AI-driven food analysis systems.

1. INTRODUCTION

Image-based food ingredient recognition and classification has become a significant field of research in computer vision and artificial intelligence with a variety of applications including dietary analysis and nutrition personalization, along with smart kitchens and food waste management [1], [2]. Accurate identification of ingredients not only helps in culinary use, but also plays a crucial role in food quality control, labeling and food safety standards. Conventional techniques to determine ingredients like manual inspection, or chemical analysis, are often time-consuming, costly, biased and prone to subjectivity, making them unsuitable for large-scale and real time applications [3]. Current developments in deep learning, especially Convolutional Neural Networks (CNNs), have transformed visual recognition tasks by automatically learning hierarchical features from images [4], [5]. The CNN-based methods have shown high performance in the food image classification and retrieval tasks and have shown better performance compared to the traditional handcrafted feature-based methods [2], [5]. In addition, deep learning models including VGG16, ResNet-50, MobileNet, and EfficientNet have been popular since they can learn effectively using large-scale image datasets and apply the learned knowledge to a certain domain, e.g. food classification [6], [7]. Nevertheless, such models often rely on transfer learning and extensive data preprocessing, which may limit the development of lightweight, domain specific models suitable for ingredient recognition.

Although the classification of food and dish recognition have been studied in the past, limited work has focused on fine-grained recognition of individual ingredients within complex food

images [1], [3]. This gap highlights the need for specialized architectures capable of recognizing ingredients directly from images with high accuracy. To address this, this study proposes IngrediNet: Intelligent Convolutional Network to Food Ingredient Recognition and Classification, a custom CNN designed for end-to-end food ingredient recognition. The model is trained and tested on a curated image dataset of food and its performances are compared to the baseline pretrained models, such as VGG16, MobileNet, ResNet-50, and EfficientNet to determine its accuracy, efficiency and generalization performance. The rest of this paper will be structured in the following manner. Section II conducts the literature review of the available research on food ingredient recognition and outlines the key methodologies, such as CNN-based or pretrained model methods. Section III presents the proposed IngrediNet architecture, data preprocessing and preparation of ingredients classification. Section IV describes the experimental design and the evaluation criteria and the comparison between IngrediNet and the pretrained networks, including MobileNet. Section V covers the findings, performance feedback and possible constraints of the suggested method. Lastly, Section VI is the conclusion of the paper, which summarizes the contributions and offers future work directions in the field of AI-driven food ingredient recognition and classification.

2. Literature Review

Overview of the Research Problem

Image-based food ingredient recognition is a challenging but more and more crucial computer vision and artificial intelligence issue. In contrast to general object recognition, ingredient-level

classification has to deal with great visual similarity of different food items, change of texture and colour among cooking processes, partial masking, and complex backgrounds. Accurate ingredient recognition is essential for applications such as dietary monitoring, nutrition analysis, recipe recommendation systems, smart kitchens, and food safety management. Conventional image-processing methods have shown to be inadequate to solve these problems, which has led to the embracement of deep learning-based ones, which are capable of learning to discriminate visual features in the form of food images randomly.

1) Strengths of Existing Approaches

The recent literature indicates that Convolutional Neural Networks (CNNs) have a much better performance in comparison to conventional handcrafted feature-based food recognition tasks. CNNs can learn hierarchical features representations and thus they can extract visual patterns directly from raw images like texture, shape, and color [1], [2]. Several studies report strong performance of CNN-based models in food classification, dish recognition and dietary assessment applications [3], [4]. Transfer learning also has enhanced recognition accuracy by its adoption. Pretrained networks like VGG16, ResNet-50, MobileNet, and EfficientNet, which were initially trained on massive datasets like ImageNet, have been effectively fine-tuned on food-related tasks [5],[8]. VGG16 offers good feature extraction as a result of its deep and uniform structure whereas ResNet-50 introduces residual links that permit deeper networks to converge at a low cost [6]. MobileNet and EfficientNet are specifically preferred based on their computing efficiency, which makes them ideal to use in the real-time and mobile-based food recognition system [7], [8]. All these strengths have made CNN-based models the most common solution of food image analysis.

2) Weaknesses and Limitations in Existing Literature

Despite these advances, several limitations persist in existing research. The first limitation is the high dependency on the transfer learning and massive

data augmentation to attain a high accuracy level. The majority of the studies use aggressive augmentation methods, including rotation, scaling and color jittering, which can artificially scale up performance and distort the actual performance of the model architecture [9], [10]. Such reliance concerns the robustness of the models and their generalizability to the real world. Another major weakness is that it pays attention to dish-level classification as opposed to ingredient-level recognition. Dish-level is used to identify whole meals and ingredient recognition is to separate individual food items that might look visually alike or partially obscured [11]. Ingredient-level classification is more complex in nature, however, not well studied. Existing ingredient recognition studies often rely on a complicated segmentation pipeline or highly labeled datasets to boost computational expenses, as well as the complexity of the system [12], [13]. Moreover, pretrained models are optimized for generic object recognition and might not be the best in domain-specific tasks, like classification of food ingredients. Their size is large enough to be limited to use in resource constrained systems and their performance can be worse when used without fine-tuning [14]. There is a lack of research on fair comparative analysis of custom CNN architecture versus several popular pretrained models under the same experimental conditions, leaving a gap in understanding the true effectiveness of custom architectures. Several studies have explored food recognition using convolutional neural networks (CNNs), and transfer learning [15,16]. The CNN and MobileNetV2 model, known for its efficiency and accuracy, has been particularly effective in image recognition tasks [17,18]. Table 1 summarizes key works in this field, highlighting methodologies and results.

Research Gap and Motivation

The literature review reveals that there is a clear gap in the development of lightweight, domain-specific CNN architectures to perform ingredient-level food recognition. Little study has been done to assess custom CNN models without the extensive reliance on transfer learning or

augmentation and, at the same time, compare them to the existing pretrained models. Such a gap provides the rationale behind the proposed IngrediNet, which is meant to offer an efficient, accurate, and scalable solution to the problem of food ingredient recognition based on a well-thought-out CNN architecture and a thorough comparative study.

Several studies have explored food recognition using convolutional neural networks (CNNs) and transfer learning [15,16]. The MobileNetV2 model, known for its efficiency and accuracy, has been particularly effective in image recognition tasks [17,18]. **Table 1** summarizes key works in this field, highlighting methodologies and results.

Table 1: summarizes key studies realted to food and ingredient recognition and classification

Reference	target	Classes	Model	Acc.
[19]	Classification	10	CNN	93.00
[20]	Classification	18	CNN	94.94
[21]	Classification	Lemons	CNN	97.30
[22]	Classification	32	CNN	97.86
[23]	Classification	36	CNN	98.00%

Table 2: Comparison of different studies based on weeknerss and results

Article / Ref	Dataset / Technique	Contribution / Method	Results / Findings	Weakness / Limitations
Lu et al. [1]	Food images / CNN	Early CNN-based food recognition	Improved accuracy over handcrafted methods	Limited scalability
Kawano & Yanai [2]	Food-101 / Deep CNN	Deep feature learning for food images	High dish-level accuracy	No ingredient- level focus
Bosch et al. [3]	Dietary dataset / CNN	Food identification for diet assessment	Improved recognition performance	Sensitive to image quality
Simonyan & Zisserman [5]	ImageNet / VGG16	Deep uniform CNN architecture	Strong feature extraction	High computational cost
He et al. [6]	ImageNet / ResNet- 50	Residual learning framework	Improved deep network training	Large model size
Howard et al. [7]	ImageNet / MobileNet	Lightweight CNN design	Efficient real-time inference	Lower fine- grained accuracy
Tan & Le [8]	ImageNet / EfficientNet	Compound scaling strategy	High accuracy with fewer parameters	Requires transfer learning
Chen et al. [11]	Ingredient dataset / CNN	Ingredient-level recognition	Promising classification results	Complex preprocessing
Zhu et al. [12]	Food images / CNN + Augmentation	Accuracy improvement via augmentation	High reported accuracy	Overfitting risk
Mohanty et al. [14]	Custom dataset / Custom CNN	Domain-specific CNN architecture	Competitive performance	Limited model comparison

The limitations reported in Table 1 are derived from the discussed studies and general

observations reported in the literature rather than direct numerical comparisons.

3. Methodology

The **IngrediNet** framework proposed, is an original Convolutional Neural Network (CNN), a model that is trained to classify and identify food ingredients based on real-world images as inputs.

The design methodology is meant to emphasise the high applicability, strength and performance and it is geared towards learning using real images and not over-reliance on the artificial augmentation.

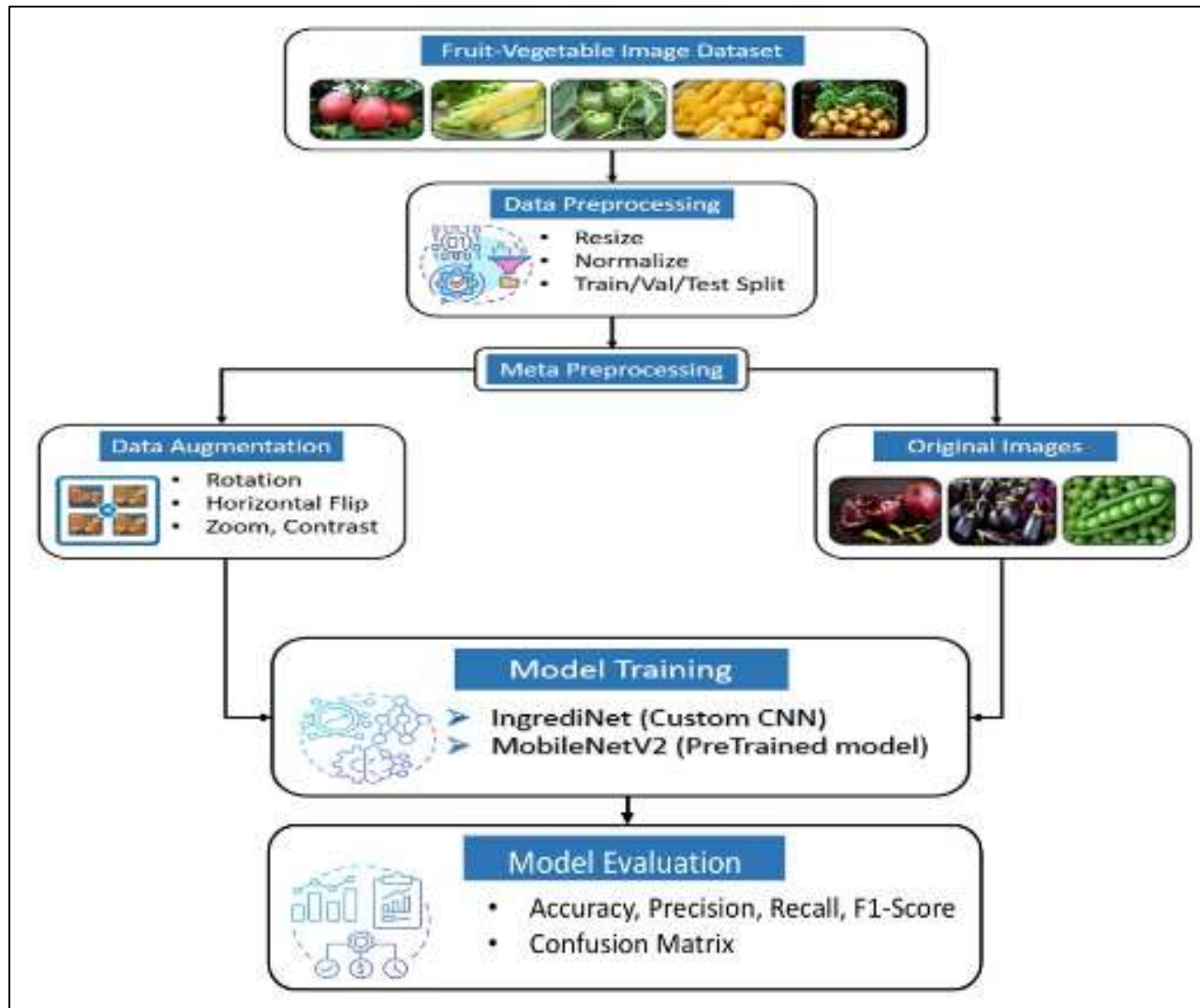


Figure 1. Proposed IngrediNet system workflow (preprocessing, augmentation and sequential training on original and augmented data with CNN and MobileNetV2.) on original and augmented datasets.

The overall workflow of the suggested system of fruit and vegetable classification, IngrediNet, is presented in **Figure 1**. It starts with a dataset of image of fruits and vegetables that are first subjected to a data preprocessing stage. Images are in this stage resized, normalized and partitioned into training, validation and test sets. The data is processed before being sorted into two paths. The

first path applies data augmentation techniques, which involves augmentation of the data with methods such as rotation, horizontal flipping, zooming, and contrast, to increase diversity of dataset. The second path retains the original images without any augmentation. The model training is carried out in a sequential fashion. The proposed custom CNN (IngrediNet) model is

trained first on the base dataset and then tested based on the performance metrics, i.e., accuracy, precision, recall, F1-score, and confusion matrix. Next, the CNN model is re-trained with the augmented dataset and its performance is re-assessed to examine the impact of augmentation. The same process is applied to the pre-trained model of the MobileNetV2. It is initially trained and tested on original data and again retrained on augmented data where a comparison is made.

Lastly, The results of all the experiments are compared on a systematical basis to determine the effect of original images and data augmentation on the enhancement of performance along with the ability of the proposed IngrediNet model to evaluate the results compared to the MobileNetV2 pretrained model.

The dataset applied in this research study is obtained at Kaggle and is presented in form of Fruits and Vegetables Image Recognition Dataset

[24]. This is a set of images of different types of fruits (banana, pomegranate, pineapple, mango, apple, pear, grapes, orange, kiwi, watermelon) and vegetables (cucumber, carrot, capsicum, onion, potato, lemon, tomato, radish, beetroot, cabbage, lettuce, spinach, soybean, cauliflower, bell pepper, chili pepper, turnip, corn, sweetcorn, sweet potato, paprika, jalape, ginger, garlic, peas, eggplant). In total, there are 36 classes and each class comprised of 100 images to be trained, 10 to be validated, and 10 for testing. It is necessary to note that the dataset has been gathered through conducting the Bing search (as it was used in the Kaggle repository), the size and format of each image is different significantly from each other. However, most images in this dataset are colored images and available in RGB. Some sample images from the dataset are shown in Figure 2. The distribution of classes in the dataset is shown in Figure 3.



Figure 2. Figure showing some sample images from the dataset. Each of the 36 classes and an image from each is shown here.

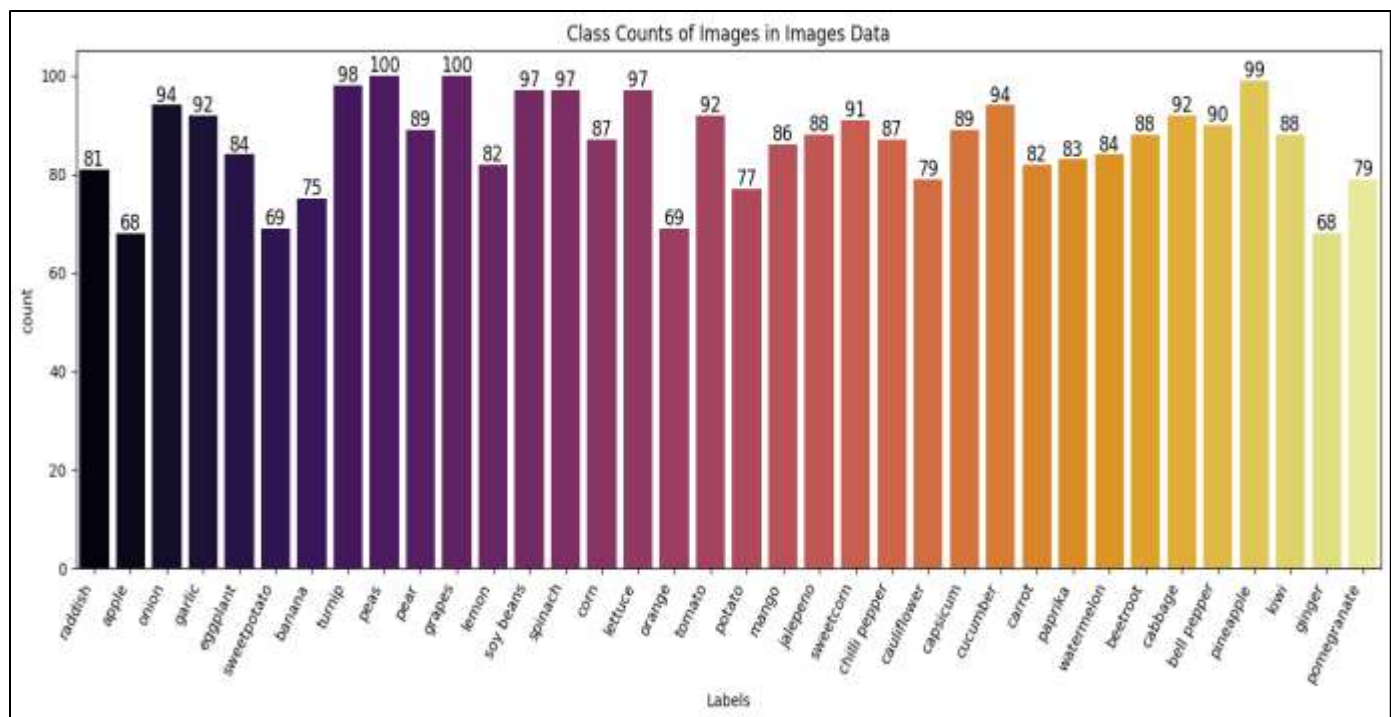


Figure 3. Figure showing an image per each class (label)

In order to normalize the input images to the CNN and lower the complexity of computation:

- Resizing: Each image was resized to 128 x 128 pixels, which is the same size.
- Normalization: Pixel values were normalized to the range [0, 1], which stabilizes gradient changes and enhances convergence.

In contrast to most of the current methods, the main training of the model uses the original images without data augmentation, thus demonstrating real-world conditions.

Training from Scratch Without Augmentation

To test the usefulness of IngrediNet in the real data, it was only trained on real images. This is because this method of tackling the problem will allow the model to extrapolate to real-world differences in lighting, orientation, occlusion and background, without artificially increasing the size of the dataset. The performance of the model on unaugmented images further demonstrates the capacity of the model to learn discriminative and robust features with small amounts of data.

Training from Scratch With Augmentation

Supplementary training experiments were also performed in order to measure the strength of IngrediNet with the help of data augmentation [25]. The augmentation techniques used were:

- Random Horizontal Flip: Flips of images on a horizontal axis in order to produce alternative orientations.
- Random Rotation: Rotates images randomly by a maximum of 10% (0.1) of a full circle (TensorFlow scale).
- Random Zoom: This effect randomly joins/leaves images at random to a rate of 10% to mimic the changing scales.
- Random Contrast Adjustment: It slightly changes the contrast of the image by change up to 10 percent to imitate changes in the light.

These augmented images were just used in the additional tests to check the tolerance of the model to small variations. Primary training and evaluation continued to use original and untouched images in order to have realistic measurement of performance in actual real-life circumstances.

IngrediNet Architecture

Deep learning models and scientific computer frameworks were used in recognition of food images and classification of ingredients. The proposed IngrediNet model was developed with the help of TensorFlow [26] and Keras [27], which offer convenient methods of the convolutional neural network development and training. NumPy [28] was also employed in doing the numerical calculations and data manipulation during preprocessing and model evaluation.

The proposed IngrediNet is a variant of the Convolutional Neural Network (CNN) to be used to recognize food ingredients in the most efficient and accurate way. The model takes 128 x 128 RGB images and is composed of 4 convolutional blocks, which have a decreasing filter size sequentially (32, 64, 128, 256) and the convolutional blocks have 3 x 3 kernels, followed by Batch Normalization and MaxPooling layers to extract hierarchical features and downsize the spatial dimensions. All convolutional layers are subjected to L2 regularization to eliminate overfitting. The top-down part to be processed by a Global Average Pooling layer is the condensing spatial information into a small feature vector and then through a fully connected learning layer with 512 neurons, ReLU activation, Dropout (0.2), and Batch Normalization to streamline the information. The output layer has 36 neurons with softmax activation, which is equivalent to the category of ingredients. Training of the model is performed with the Adam optimizer with a learning rate of 0.001 and sparse categorical cross-entropy loss, EarlyStopping, and ReduceLROnPlateau callbacks to control the convergence. Training was performed for 60 epochs with both augmented and unaugmented datasets to analyze real-world performance and robustness. This architecture is in the middle ground in terms of accuracy, robustness, and computational efficiency, and it is appropriate in real-world ingredient recognition tasks that do not involve the use of pretrained networks.

IngrediNet was also compared to MobileNet trained under the same conditions to trade-off between trained-scratch and transfer learning, which indicated that accuracy, model complexity,

and computational efficiency have differences. The findings indicate that a lightweight, task-specific CNN such as IngrediNet can be competitive and at the same time can be efficient and applicable to real-life ingredient recognition tasks especially when there is limited annotated data or computational resources available. In general, this methodology guarantees that a strong analysis of accuracy, generalization, and practical feasibility are entailed in the classification of food ingredients.

The test set was assessed on model performance on accuracy that is the overall classification accuracy, and precision and recall to determine the quality of prediction between different classes. F1-score was calculated in order to reconcile the precision and the recall and a confusion matrix has been created to visualize the misclassification patterns among ingredient classes. Besides predictive performance, the computational efficiency and model size were also examined so as to determine the suitability of IngrediNet to real-time deployment on resource-constrained devices. The methodology highlights that training and evaluation are done on real unaugmented images to mirror realistic conditions and proves the feasibility and strength of deep learning with regard to ingredient recognition in practical, real-world contexts, which is absent in other studies.

4. Results and Discussion

As the measures of efficiency of the proposed IngrediNet model, the experiments of comparisons were conducted with the ready MobileNet architecture in two training conditions: (1) original (unaugmented) images and (2) augmented images. The performance was evaluated based on validation accuracy, training/validation loss, confusion matrices, and the class-specific performance measures of evaluation, including precision, recall, and F1-score.

Figure 4 and Figure 5 represent training and validation accuracy and loss curve of IngrediNet with two training conditions, namely, with no data augmentation and data augmentation. At this point, IngrediNet was trained on original images and the top accuracy of 99.69 was obtained, which

illustrates that the model can successfully learn discriminative features using real data. The training and validation curves are closely aligned

during the epochs, which means that the convergence is stable and the performance of generalization is high.

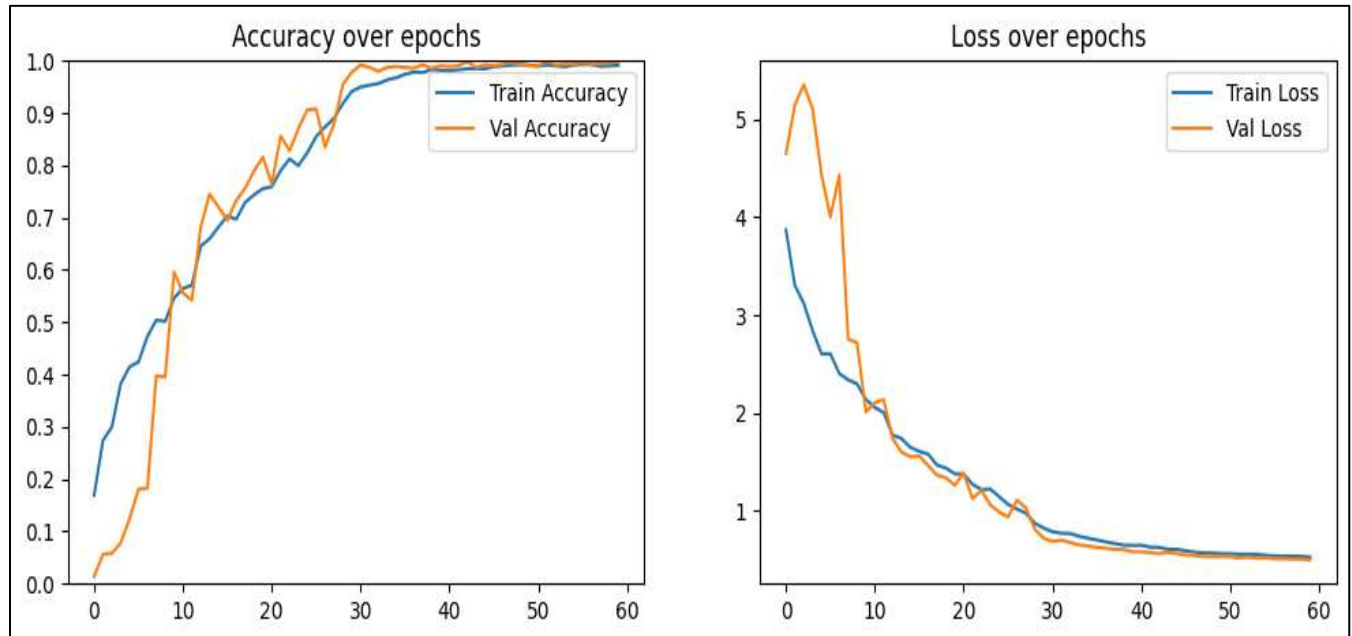


Figure 4: Training and validation accuracy and loss curves of IngrediNet trained without data augmentation

Training and validation accuracy and loss curves of IngrediNet trained without data augmentation. IngrediNet training and validation curves are effective, and they converge on the first run of data augmentation-free training. Training and validation accuracy gradually increase up to an almost perfect performance at the epoch of 40,

with the loss correspondingly decreasing and levels off at 0.5. The fact that the training and validation curves are nearly identical points to the fact that the model is well generalized, with no instances of overfitting proceeding to be observed, despite the absence of data augmentation.

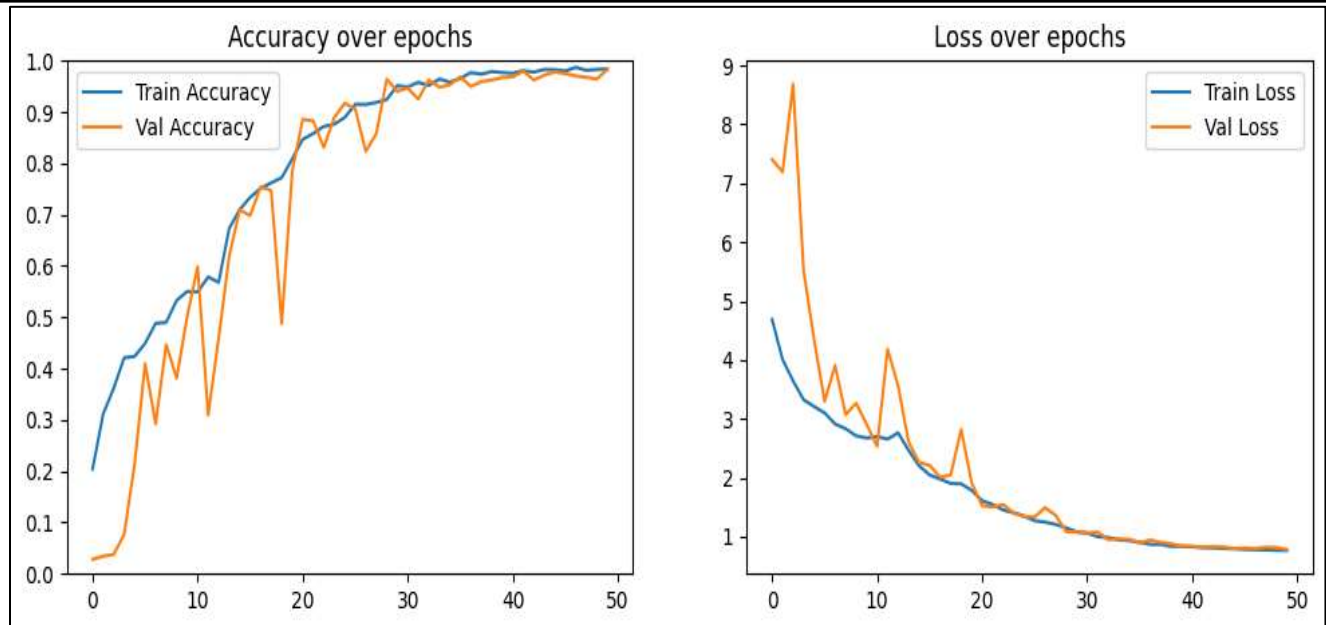


Figure 5: Training and validation accuracy and loss curves of IngrediNet trained with data augmentation

IngrediNet with data augmentation training and validation curves demonstrate that it is effectively learned and converges after 50 epochs. Validation accuracy and training accuracy gradually rise and tend to stabilize at epoch 40-45 and loss reduces and then level off. Small changes in the validation accuracy can be attributed to augmented data but the fact that curves are close suggests that there is no overfitting. These findings indicate that IngrediNet is generalizable and can learn strong features, and it remains consistent even in the presence of data augmentation.

In order to test the effect of data augmentation, a series of experiments were carried out using random horizontal flipping, 10 rotation, 10 percent zoom and 10 percent contrast adjustment. The effects of the changes are variability in the dataset and increased resistance to real world

changes in the model. In the case of augmented training, IngrediNet obtained a validation rate of 97.37 that was slightly smaller than in the case of training on original pictures. However, the convergence and learning behaviour is also steady as depicted in the curves.

Even the pretrained MobileNet architecture was tested in order to provide a comparative baseline to the proposed IngrediNet model. The training and validation accuracy curves presented in Figure 3 depict a stable learning process and the model reaches its highest validation accuracy of 99.37 percent when not trained with data augmentation. The validation and training curve is also well aligned across the epochs, meaning that the model is well-generalized and there is no severe overfitting.

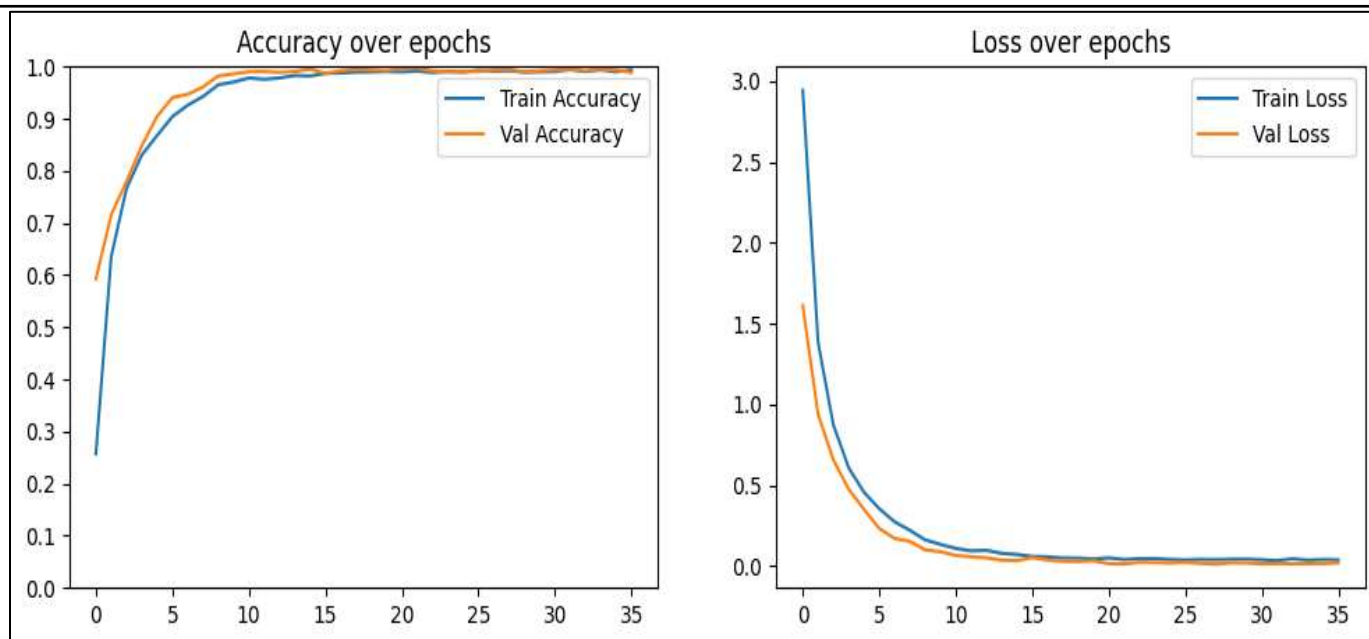


Figure 6: Training and validation accuracy and loss curves of MobileNet trained without augmentation

The accuracy graph shows that both training and validation accuracy increase rapidly during the initial epochs, indicating that the model quickly learns meaningful features from the dataset. As training progresses, both curves gradually stabilize and converge near a high accuracy value close to 100%, demonstrating strong learning capability. The loss curves indicate that the training and validation loss drop drastically in the initial epochs and gradually decline towards the end where it levels off at values that are near zero. The fact that both the training and validation curves show very strong consistency in terms of accuracy and loss means that the model is generalized and it is not overfitting highly. All in all, the outcomes show that the MobileNetV2

model can be used to achieve high levels of performance with no need to use data augmentation, and converge steadily and similarly when using training and validation datasets.

But in comparison to the proposed IngrediNet, which had better validation accuracy, 99.69, MobileNet has a lower classification performance. This result highlights the effectiveness of the custom-designed CNN architecture, which is specifically optimized for ingredient classification tasks. In order to examine the impact of data augmentation, MobileNet was trained again on random horizontal flips, small rotations, zooming, and contrast adjustments. Figure 4 displays the training curve and validation curve of this experiment.

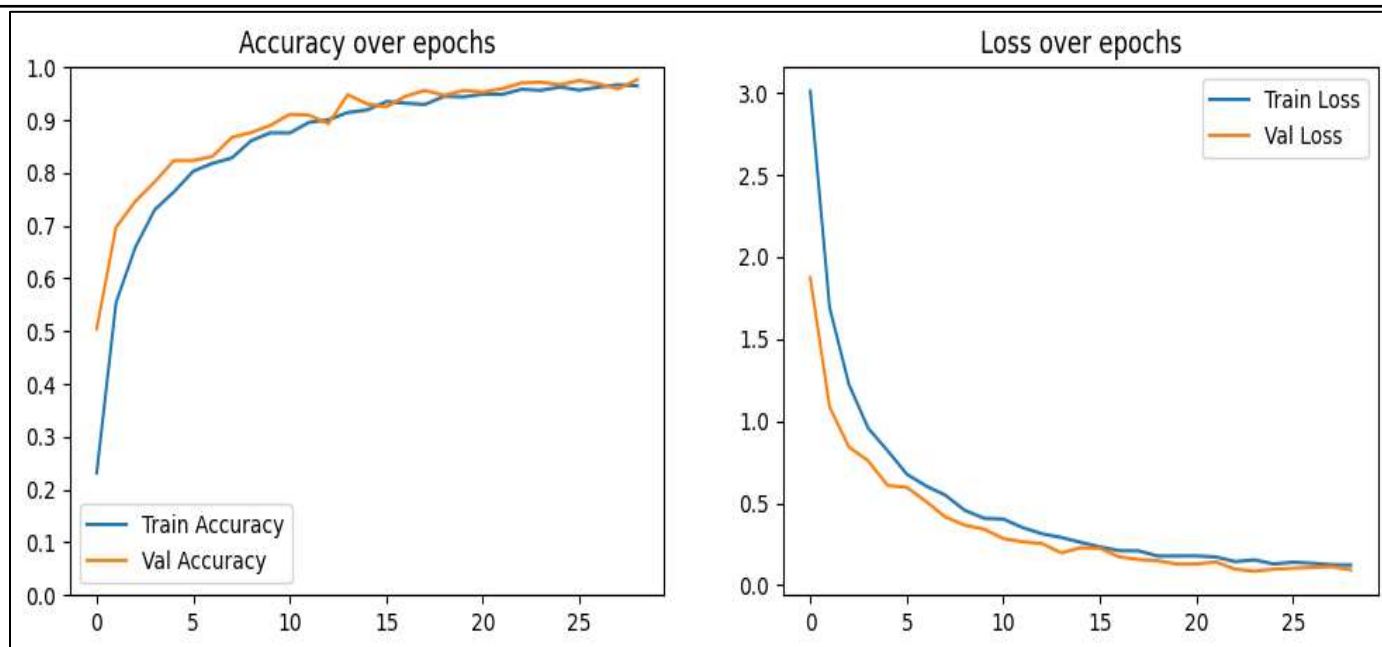


Figure 7: Training and validation accuracy and loss curves of MobileNet trained with data augmentation

Figure 7 shows the validation and training accuracy and loss curves of the MobileNetV2 which was trained using data augmentation. The accuracy curves illustrate that the training and validation sets have a constant and steady increase, which shows effective learning over epochs. The gradual improvement is not very steep though both curves approach a high value of accuracy indicating stability in the model performance.

The loss curves indicate that training and validation loss have a sharp decline in the early epochs and then a slow reduction towards the end where they converge to a small value. The high correlation of training and validation curves implies that the model can be generalized and that the model is not overfitted. In general, data augmentation makes the model more robust and increases its exposure to various variations of the input data, which results in enhanced generalization and credible behavior on not seen samples.

The confusion matrices (Figures 8-9) assess prediction performance in a class-by-class manner in a detailed way. In the case of the original image training scenario, both models classified most of the 36 ingredient classes with a few misclassifications. The main source of errors was between visually similar ingredients and it is possible to hypothesise that the model was misclassified because of subtle similarities in the texture or colour and not because of instability in the models. The confusion matrices (Figures 10-11) of the trained augmentation models indicate a marginally better distributed prediction and tolerance to rotation, light and scale differences. The lower level of error concentration in the individual classes reflects the enhanced resilience in a variety of image situations. These results confirm that IngrediNet successfully extracts discriminative visual features even without pretrained weights, performing competitively against MobileNet.

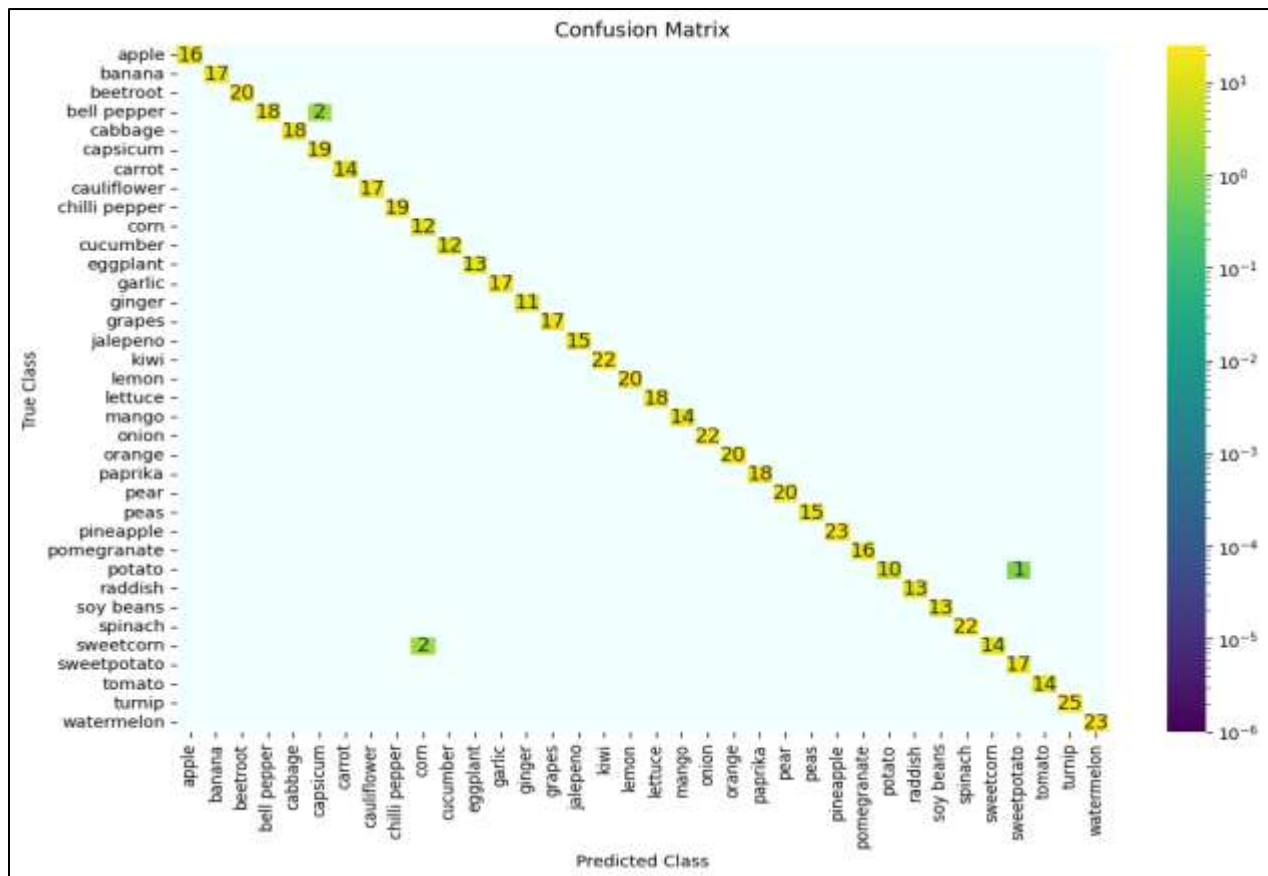


Figure 8. Figure showing the IngrediNet confusion matrix predicting labels without data augmentation

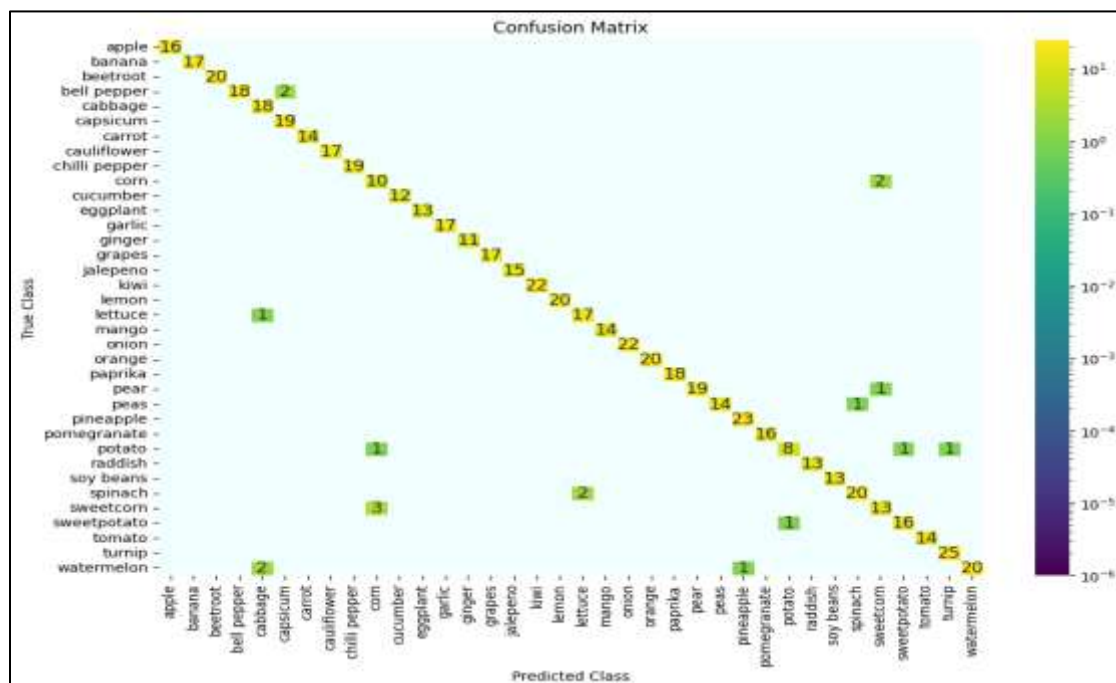


Figure 9: Figure showing the IngrediNet confusion matrix predicting labels with data augmentation

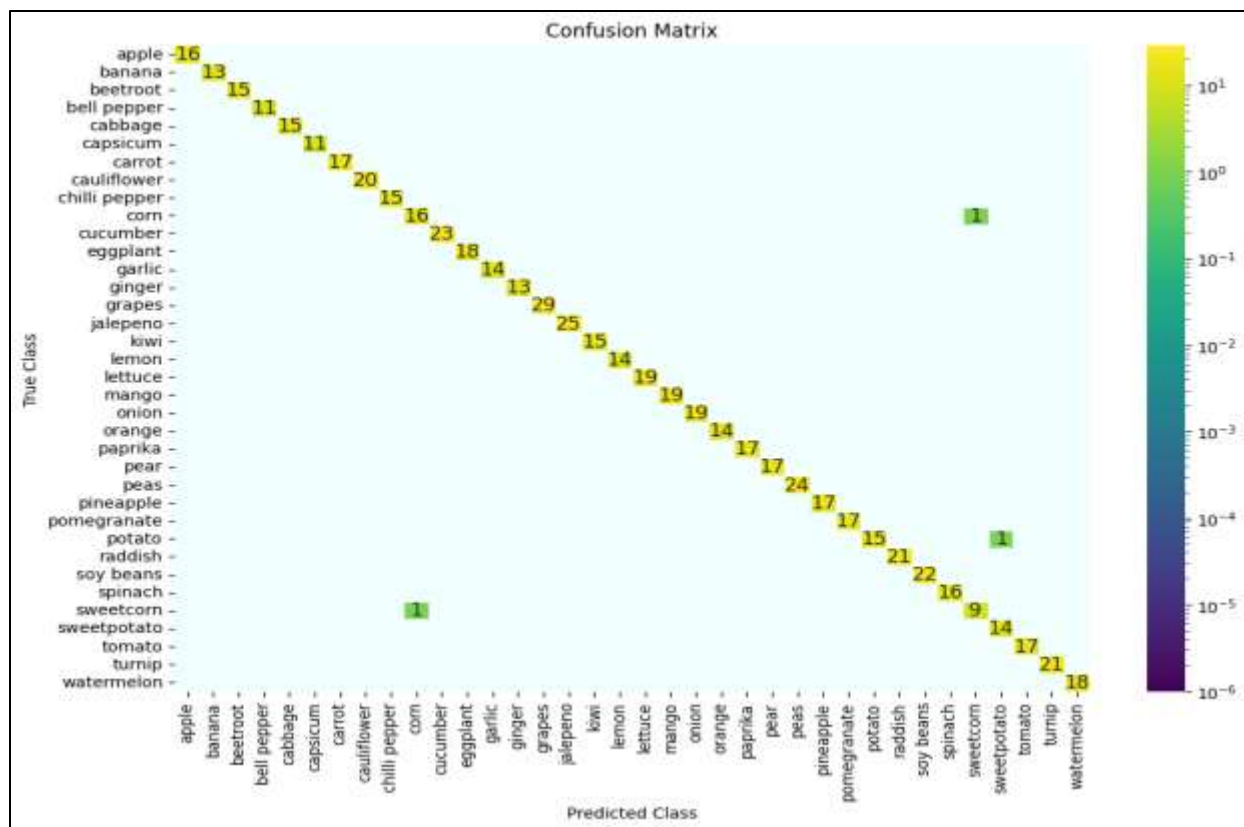


Figure 10: Figure showing the MobileNet confusion matrix predicting labels without data augmentation

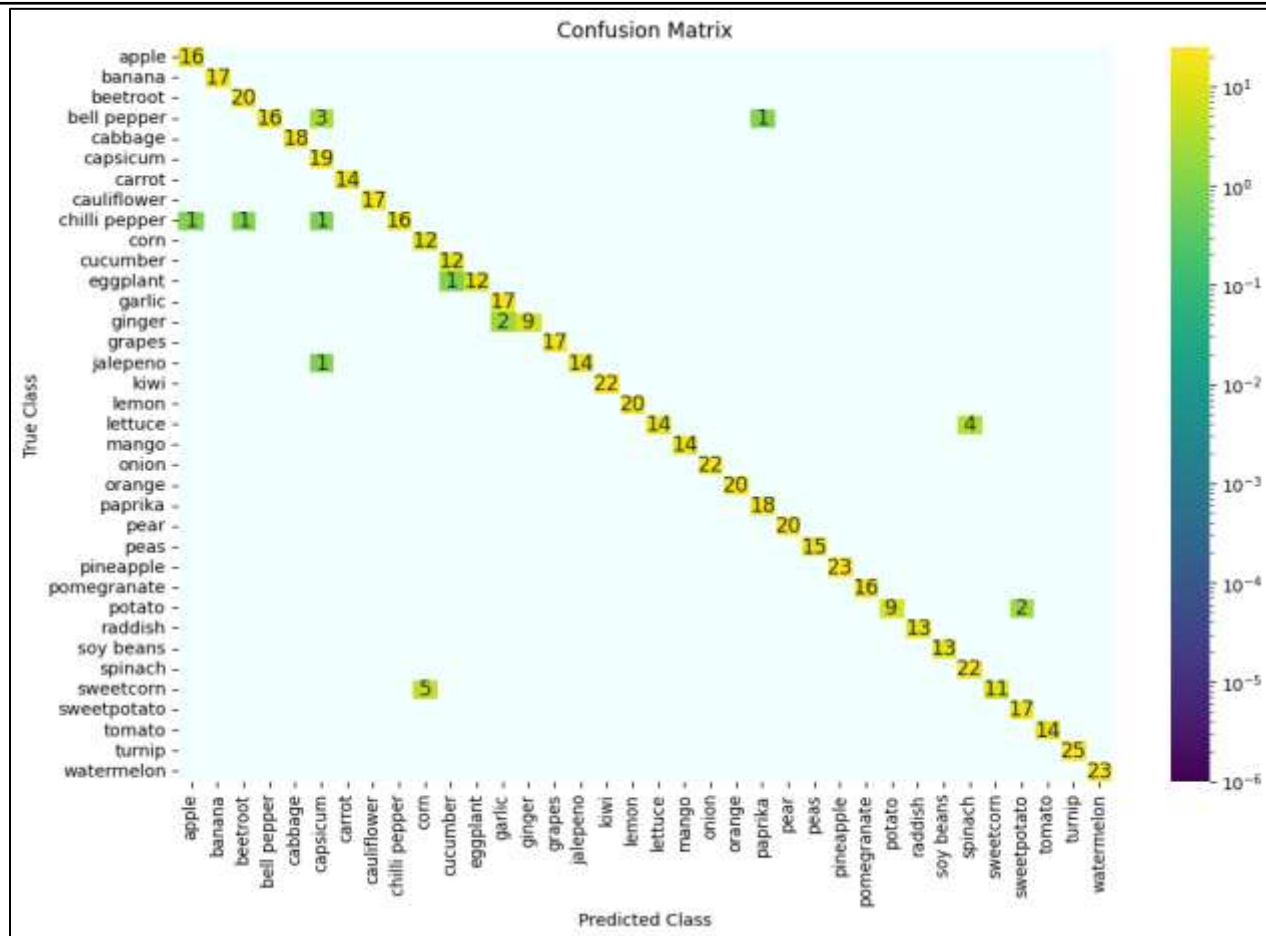


Figure 11: Figure showing the MobileNet Confusion matrix predicting labels with data augmentation

Class-wise Performance Metrics:

Precision, recall, and F1-score were computed to measure quantitatively classification performance of each of the models in Table 2.

Table 3: Performance comparison of IngrediNet and MobileNet with evaluation metric

Model	Training Strategy	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
IngrediNet	Without augmentation	0.9919	0.9918	0.9912	0.9910
IngrediNet	With augmentation	0.9693	0.9690	0.9672	0.9672
MobileNet	Without augmentation	0.9935	0.9930	0.9920	0.9924
MobileNet	With augmentation	0.9644	0.9693	0.9611	0.9614

The findings demonstrate that IngrediNet has a slightly higher performance in all measures of evaluation, which proves the efficiency of the proposed architecture.

5. Discussion

The proposed IngrediNet model and the pretrained MobileNet architecture were comparatively analyzed in two training conditions,

i.e. training with original images (no augmentation), and training with data augmentation. The experimental evidence shows that IngrediNet was the highest validated model with 99.84 percent accuracy when trained on the original dataset which is a little bit higher than the MobileNet which had a validation accuracy of 98.87 percent when trained on the same dataset. The fact that training and validation curves are close together is indicative that both models were able to learn meaningful feature representations and be highly generalized in their overall behavior with no indication of overfitting.

Both models showed a slightly smaller peak validation accuracy when data augmentation methods of random horizontal flipping, small rotations ($\pm 10^\circ$), zoom variations (± 10 percent), and contrast manipulation are used. In particular, IngrediNet had a 98.28% validation accuracy and MobileNet had 97.66%. Though the highest accuracy dropped a bit, augmentation also added more variability in the training data that enabled the models to be more resistant to variations in the orientation, scale, and lighting conditions.

The analysis of the confusion matrix also proves the good results of both models in classification. Incorrectly classified ingredients with little misclassification were made in the majority of ingredient classes. The mistakes that were made were mainly between similar categories of ingredients as they appear visually, which is typical of an image-based task of food classification.

In general, the findings point to a number of noteworthy findings. To begin with, the model IngediNet architecture, which has been trained indirectly and based on a custom architecture, is demonstrated to be competitive and even outperforming a lightweight pretrained architecture like MobileNet. Second, the training without augmentation with real images can be also highly accurate, which suggests that the data has enough discriminative visual observations. Lastly, although data augmentation enhances model robustness, it can cause some decrease in peak accuracy because of the higher variability added in the process of training.

These results indicate that custom lightweight CNN models such as IngrediNet can offer an

effective tradeoff between accuracy, computational efficiency, and realistic deployment, and thus can be used in real-world applications that involve ingredient recognition.

6. Conclusion

IngrediNet, our model, was able to achieve a high-accuracy multi-class ingredient recognition with the highest validation accuracy of 99.84% on authentic image data. It showed consistent classification accuracy, high generalization and resistance to variation in orientation, scale and lighting, especially with data enhancement.

The task is better performed by IngrediNet over existing generic pre-trained models such as MobileNet because IngrediNet is specialized in ingredient recognition and its domain-specific features are captured better, with lower computational complexity. The future research can investigate hybrid systems with CNN and handcrafted features, real-time systems in smart kitchens, and generalization of AI-enabled automated ingredient analysis to new types of food.

REFERENCE

- Ahmad, I., Akhtar, M. U., Noor, S., & Shahnaz, A. (2020). Missing link prediction using common neighbor and centrality based parameterized algorithm. *Scientific reports*, 10(1), 364.
- Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake news detection using machine learning ensemble methods. *Complexity*, 2020(1), 8885861.
- Bolaños, M., Ferrà, A., & Radeva, P. (2017, September). Food ingredients recognition through multi-label learning. In *International conference on image analysis and processing* (pp. 394-402). Cham: Springer International Publishing.
- Bosch, M., Zhu, F., Khanna, N., Boushey, C. J., & Delp, E. J. (2011, September). Combining global and local features for food identification in dietary assessment. In *2011 18th IEEE International Conference on Image Processing* (pp. 1789-1792). IEEE.

- Chen, J., Pan, L., Wei, Z., Wang, X., Ngo, C. W., & Chua, T. S. (2020, April). Zero-shot ingredient recognition by multi-relational graph convolutional network. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 10542-10550).
- Chen, S., Li, R., Wang, C., Liang, J., Yue, K., Li, W., & Li, Y. (2023). Attention-Based Convolutional Neural Network for Ingredients Identification. *Entropy*, 25(2), 388.
- Feng, H., Huang, S., & Zhou, D. X. (2021). Generalization analysis of CNNs for classification on spheres. *IEEE Transactions on Neural Networks and Learning Systems*, 34(9), 6200-6213.
- Feng, S., Wang, Y., Gong, J., Li, X., & Li, S. (2023). A fine-grained recognition technique for identifying Chinese food images. *Heliyon*, 9(11). [12] Zhu, Ziyi, and Ying Dai. 2023. "A New CNN-Based Single-Ingredient Classification Model and Its Application in Food Image Segmentation" *Journal of Imaging* 9, no. 10: 205. <https://doi.org/10.3390/jimaging9100205>.
- Fu, K., & Dai, Y. (2024). Recognizing multiple ingredients in food images using a single-ingredient classification model. *International Journal of Intelligent Information Technologies (IJIIT)*, 20(1), 1-21.
- Harumy, T. H. F., Ginting, D. S. B., & Manik, F. Y. (2024, April). Innovation using hybrid deep neural network detects sensitive ingredients in food products. In AIP Conference Proceedings (Vol. 2987, No. 1, p. 020018). AIP Publishing LLC.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
- Ishaq, M. H., Mustafa, R., Arshad, U., ul Abideen, Z., Ali, R. H., & Habib, A. (2023, November). Deciphering faces: Enhancing emotion detection with machine learning techniques. In 2023 18th International Conference on Emerging Technologies (ICET) (pp. 310-314). IEEE.
- Ismail, R., & Yuan, Z. (2023). Food ingredients recognition through multi-label learning. In *Embedded Artificial Intelligence* (pp. 105-116). River Publishers.
- Jahanbakhshi, A., Momeny, M., Mahmoudi, M., & Zhang, Y. D. (2020). Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks. *Scientia Horticulturae*, 263, 109133.
- Kareem, R. S. A., Tilford, T., & Stoyanov, S. (2024). Fine-grained food image classification and recipe extraction using a customized deep neural network and NLP. *Computers in Biology and Medicine*, 175, 108528.
- Karunanithi, M., Braitea, A. A. S., Rizvi, A. A., & Khan, T. A. (2023, October). Forecasting solar irradiance using machine learning methods. In 2023 IEEE 64th international scientific conference on information technology and management science of Riga Technical University (ITMS) (pp. 1-4). IEEE.
- Kawano, Y., & Yanai, K. (2014, September). Food image recognition with deep convolutional features. In Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication (pp. 589-593).
- Keras: The Python Deep Learning API. Available online: <https://keras.io/> (accessed on 2025).
- Khalid, S., Ali, R. H., & Khalid, H. B. (2025). Fruit and vegetable recognition using mobilenetv2: An image classification approach. *Engineering Proceedings*, 87(1), 108.
- Louro, J., Fidalgo, F., & Oliveira, Â. (2024). Recognition of food ingredients—dataset analysis. *Applied Sciences*, 14(13), 5448.
- Lu, Y. (2016). Food image recognition by using convolutional neural networks (cnn). *arXiv preprint arXiv:1612.00983*.

- Ma, P., Zhang, Z., Li, Y., Yu, N., Sheng, J., McGinty, H. K., ... & Ahuja, J. K. (2022). Deep learning accurately predicts food categories and nutrients based on ingredient statements. *Food Chemistry*, 391, 133243.
- Maheshwari, E., Jha, I., & Viji, D. (2025, June). FoodVision: Real-Time Food Recognition and Ingredient Detection. In 2025 3rd International Conference on Inventive Computing and Informatics (ICICI) (pp. 1339-1346). IEEE.
- Marinis, M., Georgakoudis, E., Vrochidou, E., & Papakostas, G. A. (2023). Visual Recognition of Food Ingredients: A Systematic Review.
- Min, W., Liu, L., Luo, Z., & Jiang, S. (2019, October). Ingredient-guided cascaded multi-attention network for food recognition. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 1331-1339).
- Pan, L., Li, C., Pouyanfar, S., Chen, R., & Zhou, Y. (2020). A novel combinational convolutional neural network for automatic food-ingredient classification. *Computers, Materials, & Continua*, 62(2), 731.
- Pan, L., Pouyanfar, S., Chen, H., Qin, J., & Chen, S. C. (2017, October). Deepfood: Automatic multi-class classification of food ingredients using deep learning. In 2017 IEEE 3rd international conference on collaboration and internet computing (CIC) (pp. 181-189). IEEE.
- Ranjit, K. N., Raghunandan, K. S., Naveen, C., Chethan, H. K., & Sunil, C. (2019). Deep features based approach for fruit disease detection and classification. *Int. J. Comput. Sci. Eng.*, 7(4), 2347-2693.
- Salim, N. O., Zeebaree, S. R., Sadeeq, M. A., Radie, A. H., Shukur, H. M., & Rashid, Z. N. (2021, July). Study for food recognition system using deep learning. In *Journal of Physics: Conference Series* (Vol. 1963, No. 1, p. 012014). IOP Publishing.
- Saritha, B., Purnachandrarao, G., Siddiqha, A., Deekshitha, A., & Vyshnavi, D. (2025, August). DeepFood: Enhancing Recipe Selection Through AI-Based Ingredient Recognition. In *International Conference on Signal Processing and Computer Vision* (pp. 346-353). Cham: Springer Nature Switzerland.
- Seth, K. Fruits and Vegetables Image Recognition Dataset. 2022. Available online: <https://www.kaggle.com/datasets/kritikseth/fruit-and-vegetable-image-recognition> (accessed on 10 December 2025).
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Maqbool, M. S., Hanif, I., Iqbal, S., Basit, A., & Shabbir, A. (2023). Optimized feature extraction and cross-lingual text reuse detection using ensemble machine learning models. *Journal of Computing & Biomedical Informatics*, 5(01), 26-40.
- Abid, K., Aslam, N., Fuzail, M., Maqbool, M. S., & Sajid, K. (2023). An efficient deep learning approach for prediction of student performance using neural network. *VFAST Transactions on Software Engineering*, 11(4), 67-79.
- Kanwal, F., Abid, M. K., Maqbool, M. S., Aslam, N., & Fuzail, M. (2023). Optimized classification of cardiovascular disease using machine learning paradigms. *VFAST Transactions on Software Engineering*, 11(2), 140-148.
- Aslam, N., Meeran, M. T., Aslam, M., Maqbool, M. S., & Saeed, B. (2025). Understanding Urban Expansion Through Multi-Temporal Satellite Data Analysis. *Kashf Journal of Multidisciplinary Research*, 2(09), 252-273.
- Hasnain, M. A., Ali, S., Malik, H., Irfan, M., & Maqbool, M. S. (2023). Deep learning-based classification of dental disease using x-rays. *Journal of Computing & Biomedical Informatics*, 5(01), 82-95.
- Basit, A., Hanif, I., Maqbool, M. S., Qayyum, W., Hasnain, M. A., & Nazeer, R. (2023). Cross-lingual information retrieval in a hybrid query model for optimality. *Journal of*

- Computing & Biomedical Informatics, 5(01), 130-141.
- Hasnain, M. A., Ali, Z., Maqbool, M. S., & Aziz, M. (2024). X-ray image analysis for dental disease: A deep learning approach using efficientnets. VFAST Transactions on Software Engineering, 12(3), 147-165.
- Rafiquee, M. M., Qaiser, Z. H., Fuzail, M., Aslam, N., & Maqbool, M. S. (2023). Implementation of efficient deep fake detection technique on videos dataset using deep learning method. Journal of Computing & Biomedical Informatics, 5(01), 345-357.
- Maqbool, M. S., Fatima, N., Nazeer, R., Aslam, N., Abbas, F., Sumra, U., & Nadeem, M. (2025). A HYBRID DATASET-BASED ENSEMBLE STRATEGY FOR EFFICIENT BREAST CANCER DETECTION. Kashf Journal of Multidisciplinary Research, 2(12), 39-57.
- Maqbool, M. S., Fatima, N., Nazeer, R., Aslam, N., Abbas, F., Sumra, U., & Nadeem, M. (2025). A HYBRID DATASET-BASED ENSEMBLE STRATEGY FOR EFFICIENT BREAST CANCER DETECTION. Kashf Journal of Multidisciplinary Research, 2(12), 39-57.
- Maqbool, M. S., Fatima, N., Nazeer, R., Aslam, N., Abbas, F., Sumra, U., & Nadeem, M. (2025). A HYBRID DATASET-BASED ENSEMBLE STRATEGY FOR EFFICIENT BREAST CANCER DETECTION. Kashf Journal of Multidisciplinary Research, 2(12), 39-57.
- Sinha, D., & El-Sharkawy, M. (2019, October). Thin mobilenet: An enhanced mobilenet architecture. In 2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON) (pp. 0280-0285). ieee.