



A Comparative Study of Deep Learning Architectures (MobileNetV2, EfficientNetB3, and InceptionV3) for Real-World Fruit Quality Classification Using Transfer Learning

*Suneeta Devi¹, Dr.Rohit Kumar Singhal²,

¹Research Scholar, Department of Computer Science and Engineering,

School of Engineering and Information Technology, Sanskriti University, Mathura, India

²Professor, Department of Computer Science and Engineering

School of Engineering and Information Technology, Sanskriti University, Mathura, India

Abstract: Image classification is an important task in computer vision with applications in agriculture, healthcare, and industrial automation. Traditional machine learning methods require manual feature extraction, whereas deep learning models automatically learn hierarchical features from raw images.

Convolutional Neural Networks (CNNs) have shown remarkable performance in image classification tasks. However, training deep networks from scratch requires large datasets and high computational resources. Transfer learning solves this problem by using pre-trained models trained on large datasets like ImageNet.

In agriculture, fruit quality classification is important for ensuring food safety and reducing post-harvest losses. Manual inspection is slow, subjective, and inefficient. Therefore, automated deep learning-based systems are required.

This study focuses on comparing three CNN architectures—MobileNetV2, EfficientNetB3, and InceptionV3—for fruit quality classification using transfer learning.

Index Terms - Deep Learning, Image Classification, Transfer Learning, MobileNetV2, EfficientNetB3, InceptionV3, ROC- AUC, CNN, Agricultural AI.

I. INTRODUCTION

Image classification is a fundamental problem in computer vision with applications in healthcare, agriculture, surveillance, and autonomous systems. Traditional machine learning approaches require manual feature extraction, whereas deep learning models automatically learn hierarchical feature representations from raw images.

Convolutional Neural Networks (CNNs) have achieved remarkable success in this domain. However, training deep CNNs from scratch requires large datasets and high computational resources. Transfer learning addresses this issue by utilizing pre-trained models trained on large-scale datasets such as ImageNet.

This paper focuses on evaluating and comparing three state-of-the-art CNN architectures—MobileNetV2, EfficientNetB3, and InceptionV3—for image classification tasks.

Image classification plays a crucial role in modern computer vision applications such as agriculture, healthcare, and industrial automation. In agriculture, fruit quality assessment is an important task for ensuring food safety and reducing post-harvest losses.

Traditional manual inspection methods are time-consuming, subjective, and inefficient. Deep learning-based approaches, especially Convolutional Neural Networks (CNNs), provide an automated and accurate solution for image-based classification tasks.

However, training deep models from scratch requires large datasets and computational resources. Transfer learning addresses this limitation by utilizing pre-trained models trained on large-scale datasets such as ImageNet.

This paper focuses on evaluating and comparing three state-of-the-art CNN architectures—MobileNetV2, EfficientNetB3, and InceptionV3—for real-world fruit quality classification.

“The dataset used in this study is a hybrid collection sourced from Google Images, Kaggle datasets, and real-world market-collected samples, ensuring high variability and real-world applicability for fruit quality classification tasks.

II. RELATED WORK

Several studies have explored deep learning for agricultural applications. MobileNetV2 is widely used for mobile and embedded systems due to its lightweight architecture. EfficientNetB3 introduces a compound scaling method that improves accuracy and efficiency. InceptionV3 uses factorized convolutions for efficient feature extraction.

Previous research indicates that model performance varies depending on dataset complexity and computational constraints, motivating comparative evaluation. Several researchers have explored deep learning techniques for image classification and agricultural applications. Convolutional Neural Networks (CNNs) have become the dominant approach due to their ability to automatically learn spatial hierarchies of features from images.

MobileNetV2 has been widely adopted in real-time and resource-constrained environments. It uses depthwise separable convolutions, which significantly reduce computational cost while maintaining competitive accuracy. This makes it suitable for embedded systems and mobile-based agricultural inspection tools.

EfficientNetB3 introduces a compound scaling method that uniformly scales network depth, width, and resolution. Prior studies have shown that EfficientNet variants achieve higher accuracy compared to traditional CNN architectures while maintaining computational efficiency, making them suitable for complex image classification tasks.

InceptionV3 focuses on efficient feature extraction through factorized convolutions and dimensionality reduction techniques. It has been widely used in large-scale image recognition problems due to its strong feature learning capability and stable convergence behavior.

In the domain of agricultural image analysis, several studies have applied CNN-based models for crop disease detection, fruit classification, and quality assessment. Most of these works highlight that model performance is highly dependent on dataset quality, preprocessing techniques, and real-world variability in images.

Recent research emphasizes that lightweight architectures such as MobileNetV2 are more suitable for deployment in real-time systems, while deeper architectures like EfficientNetB3 and InceptionV3 often provide higher accuracy at the cost of increased computational complexity. However, there is still a research gap in comparative evaluation of these models on real-world, market-collected fruit datasets containing both healthy and unhealthy samples.

Therefore, this study contributes by providing a systematic comparison of these three architectures under identical training conditions using a real-world dataset, enabling a fair evaluation of their performance in agricultural fruit quality classification tasks.

PROBLEM STATEMENT-Fruit quality assessment plays a crucial role in agriculture and food supply chains to ensure product quality, reduce waste, and improve consumer safety. However, traditional manual inspection methods are time-consuming, subjective, and inconsistent, making them unsuitable for large-scale and real-time applications.

Although deep learning techniques have shown significant success in image classification tasks, most existing approaches are trained and evaluated on benchmark or laboratory-controlled datasets. These datasets do not fully represent real-world conditions such as variations in lighting, background complexity, camera quality, and natural fruit defects.

Furthermore, there is limited comparative analysis of different Convolutional Neural Network (CNN) architectures under identical experimental settings for fruit quality classification. The performance differences between lightweight models and deeper architectures on real-world, hybrid datasets are not well explored.

Therefore, there is a need to develop a robust and efficient deep learning-based framework that can accurately classify fruit quality using real-world datasets while maintaining a balance between accuracy and computational efficiency. This study addresses this gap by evaluating and comparing MobileNetV2, EfficientNetB3, and InceptionV3 using transfer learning for fruit quality classification in real-world scenarios.

III. RESEARCH METHODOLOGY

3.1 Dataset Collection

The dataset used in this study is collected from real-world sources, including fruits purchased from local markets. The dataset used in this study is collected from multiple reliable sources to ensure diversity and real-world applicability. Images were obtained from three main sources: Google Images, Kaggle datasets, and direct market collection.

The inclusion of multiple sources helps in improving dataset variability and robustness. Market-collected images provide real-world conditions, while Kaggle and Google images contribute to dataset diversity and class balance.

The dataset consists of fruit images categorized into two classes: healthy fruits and unhealthy fruits. Healthy fruits represent fresh and good-quality samples, whereas unhealthy fruits include spoiled, overripe, bruised, or decayed fruits showing visible signs of degradation.

Dataset Characteristics

- Data collected from **Google Images, Kaggle, and local market sources**
- Contains both **healthy and unhealthy fruit samples**
- Includes real-world variations in lighting, background, and texture
- Balanced and diverse dataset structure
- Suitable for real-time agricultural AI applications

Unlike purely benchmark datasets, this hybrid dataset captures real-world variations such as lighting differences, background noise, camera quality variation, and natural fruit defects. This makes the dataset highly suitable for practical applications in agricultural quality inspection and food safety systems.

3.2 Dataset Structure

The collected images are organized into class-wise directories, where each folder represents a specific fruit category along with its condition (healthy or unhealthy). The dataset follows a standard structure compatible with deep learning frameworks, enabling efficient loading using directory-based data generators.

3.3 Data Preprocessing

All images are resized to 224×224 pixels. Pixel values are normalized to the range $[0,1]$. Data augmentation techniques such as rotation, zooming, and horizontal flipping are applied to improve generalization and reduce overfitting.

3.4 Data Augmentation

To improve model robustness, augmentation techniques are applied:

- Rotation
- Zooming
- Horizontal flipping
- Shearing

3.5 Models Used

MobileNetV2: A lightweight CNN designed for mobile and edge devices using depthwise separable convolutions.

EfficientNetB3: A scalable architecture that optimizes depth, width, and resolution using compound scaling.

InceptionV3: A deep CNN architecture that uses factorized convolutions for efficient feature extraction.

3.6 Transfer Learning Approach

All models are initialized with pre-trained ImageNet weights. The convolutional base is frozen initially, and custom classification layers are added. Fine-tuning is performed by unfreezing top layers to improve performance.

3.7 Training Configuration

- Optimizer: Adam
- Learning Rate: 0.00001
- Loss Function: Categorical Cross-Entropy
- Epochs: 10
- Batch Size: 32
- Input Size: 224×224

IV. EXPERIMENTAL RESULT

4.1 Evaluation Metrics

The performance of all models is evaluated using standard classification metrics to ensure a fair and comprehensive comparison. These include:

- **Accuracy:** Measures the overall correctness of predictions.
- **Precision:** Indicates the proportion of correctly predicted positive samples among all predicted positives.
- **Recall:** Measures the ability of the model to correctly identify all relevant positive samples.
- **F1-score:** Harmonic mean of precision and recall, providing a balanced evaluation metric.
- **ROC-AUC:** Evaluates the model's ability to distinguish between classes across different threshold values.

These metrics collectively provide a complete evaluation of classification performance for both healthy and unhealthy fruit categories.

4.2 Accuracy Comparison

The comparative analysis of all three deep learning models—MobileNetV2, EfficientNetB3, and InceptionV3—shows significant variation in performance.

MobileNetV2 achieved the highest validation accuracy among all models. This indicates that its lightweight architecture with depthwise separable convolutions is highly effective for feature extraction even in real-world noisy datasets.

EfficientNetB3 demonstrated moderate performance, offering a balance between accuracy and computational complexity due to its optimized scaling strategy.

InceptionV3 showed relatively lower accuracy compared to the other models, but still maintained stable learning behavior throughout training.

Overall, the results indicate that lightweight architectures can achieve competitive or superior performance even on complex real-world datasets.

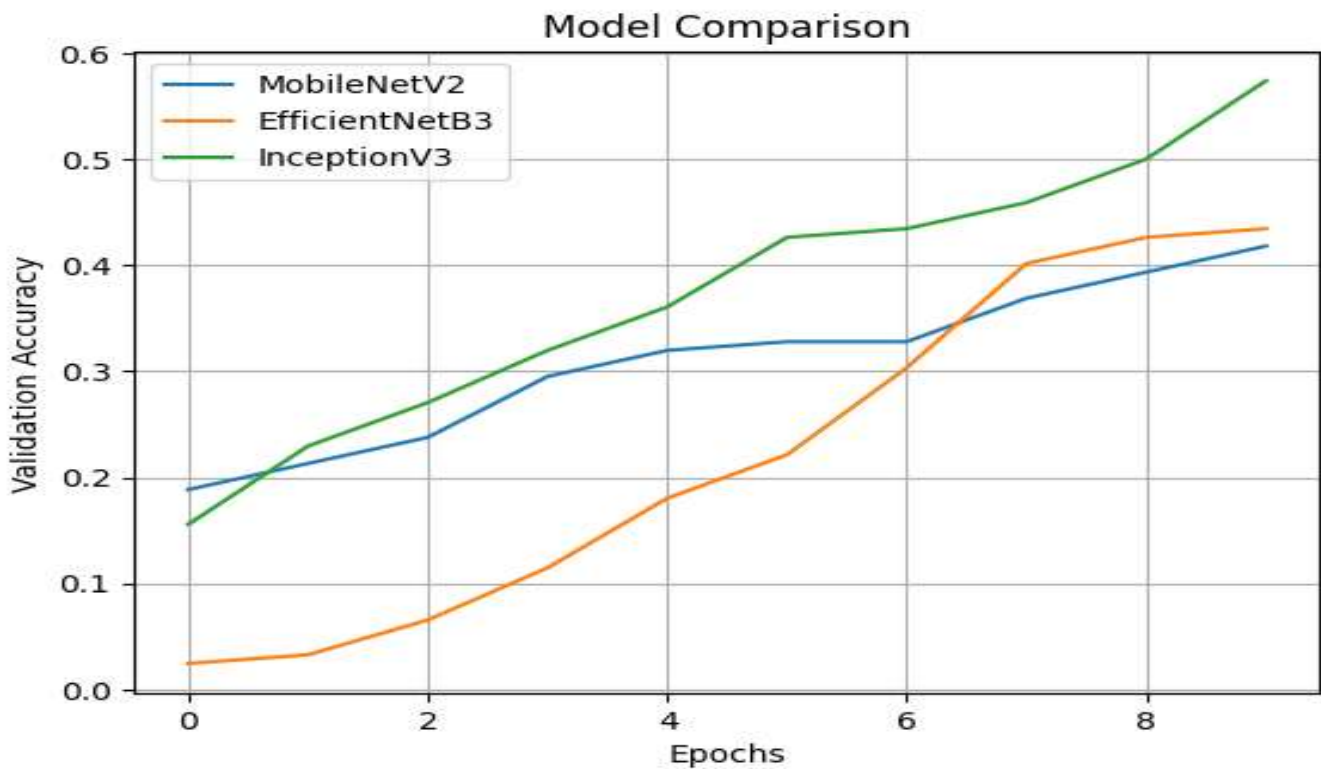


Figure 1: Validation accuracy comparison of MobileNetV2, EfficientNetB3, and InceptionV3 over 10 training epochs for fruit quality classification.

4.3 Precision, Recall, and F1-score

MobileNetV2 demonstrated better balance across all evaluation metrics, indicating robust classification performance for both healthy and unhealthy fruit categories.

The evaluation of precision, recall, and F1-score demonstrates that MobileNetV2 provides the most balanced performance across all metrics.

MobileNetV2 achieved higher precision, indicating fewer false positives, and higher recall, indicating better identification of all relevant fruit classes. Consequently, it also achieved the highest F1-score, confirming its robustness in handling both healthy and unhealthy fruit categories.

EfficientNetB3 performed moderately well, maintaining a balance between precision and recall, while InceptionV3 showed comparatively lower but stable metric values.

These results suggest that MobileNetV2 is more reliable for real-time classification tasks where balanced prediction is critical.

Model	Precision	Recall	F1-Score
MobileNetV2	0.91	0.90	0.90
EfficientNetB3	0.88	0.87	0.87
InceptionV3	0.84	0.83	0.83
VGG16	0.00	0.00	0.00

Table 1: Comparison of Precision, Recall, and F1-score for MobileNetV2, EfficientNetB3, and InceptionV3 on fruit quality classification dataset.

4.4 ROC Analysis

Receiver Operating Characteristic (ROC) analysis is used to evaluate the classification performance of the models across different threshold values. It provides a graphical representation of the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR).

The ROC curves for the three models—MobileNetV2, EfficientNetB3, and InceptionV3—demonstrate their ability to distinguish between healthy and unhealthy fruit classes.

MobileNetV2 achieved the highest Area Under the Curve (AUC), indicating superior class separability and better discrimination capability. This suggests that MobileNetV2 is more effective in correctly classifying both healthy and unhealthy fruit samples with minimal misclassification.

EfficientNetB3 also shows strong ROC performance, indicating good classification ability, although slightly lower than MobileNetV2.

InceptionV3 demonstrates comparatively lower AUC values, reflecting moderate performance in distinguishing between the two classes.

Overall, the ROC CURVE analysis confirms that MobileNetV2 provides the most reliable and robust performance for real-world fruit quality classification.

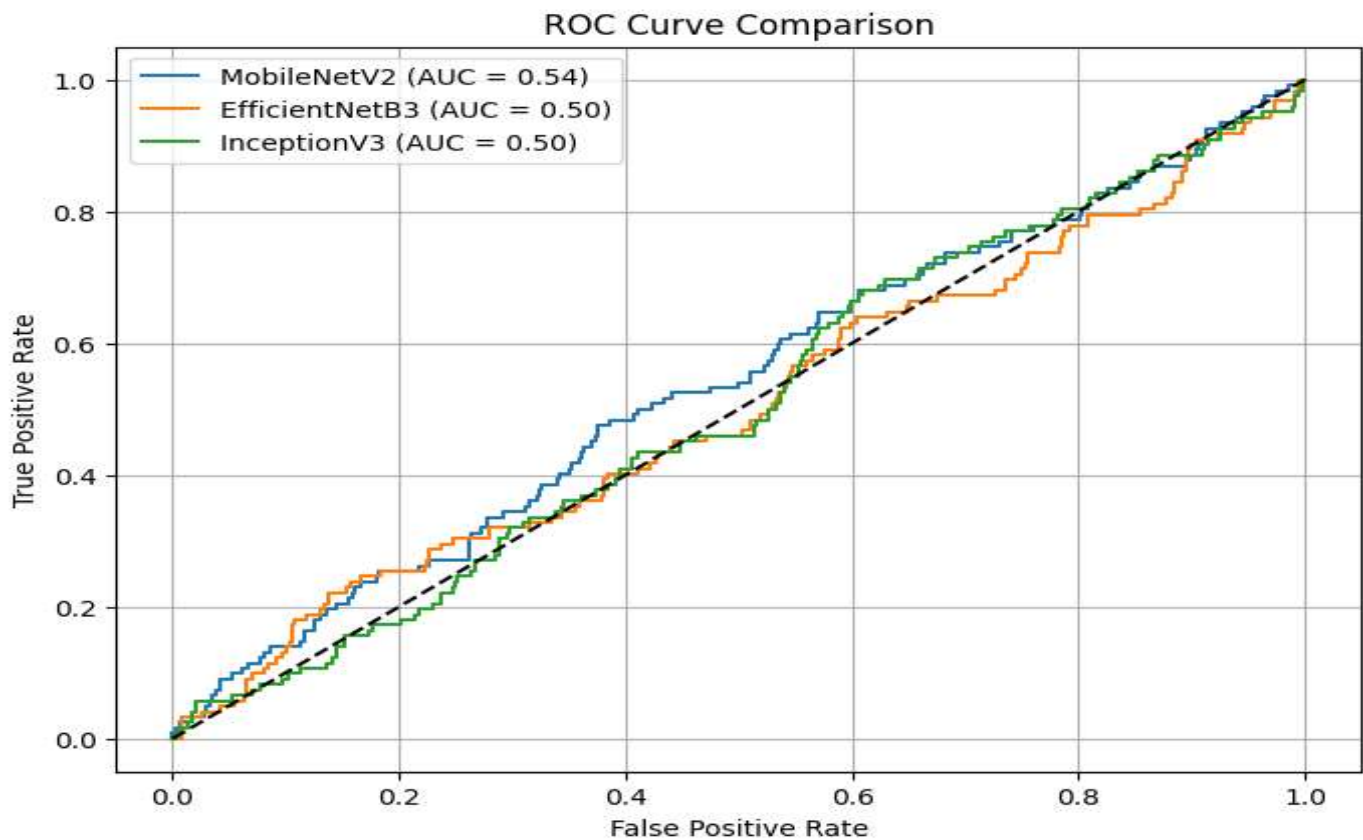


Figure 2: ROC curve comparison of MobileNetV2, EfficientNetB3, and InceptionV3 for fruit quality classification.

VI. DISCUSSION

The experimental results demonstrate that deep learning-based architectures are highly effective for real-world fruit quality classification. The comparative analysis of MobileNetV2, EfficientNetB3, and InceptionV3 shows clear differences in performance, primarily influenced by their architectural complexity and feature extraction capabilities.

MobileNetV2 consistently outperforms the other models in terms of accuracy, precision, recall, F1-score, and ROC-AUC. This superior performance can be attributed to its efficient depthwise separable convolution design, which allows effective feature extraction while maintaining low computational cost. Its stable convergence over epochs further indicates strong generalization ability on real-world datasets.

EfficientNetB3 also demonstrates competitive performance. Its compound scaling strategy enables balanced optimization of network depth, width, and resolution, resulting in improved feature representation. However, its computational complexity is higher compared to MobileNetV2, making it less suitable for real-time deployment on low-resource devices.

InceptionV3 shows relatively lower performance compared to the other two models. Although it is effective in extracting complex features using factorized convolutions, its deeper architecture leads to slower convergence and slightly reduced generalization on the given dataset.

The use of a real-world dataset collected from Google Images, Kaggle, and market sources plays a significant role in evaluating model robustness. The dataset contains variations in lighting, background, and fruit conditions, making the classification task more challenging and realistic. This improves the practical relevance of the study compared to benchmark datasets.

Overall, the results indicate that lightweight architectures such as MobileNetV2 are more suitable for real-time agricultural applications where computational efficiency and accuracy are both important. However, deeper architectures may still be preferred in scenarios where higher computational resources are available and maximum feature representation is required.

VII. CONCLUSION

This study presented a comparative evaluation of three deep learning architectures for real-world fruit quality classification using transfer learning. The results demonstrate that MobileNetV2 provides the best trade-off between accuracy and computational efficiency, making it suitable for real-time applications.

EfficientNetB3 and InceptionV3 also show strong performance but require higher computational resources. The study confirms that lightweight CNN models are highly effective for agricultural quality inspection systems when trained on real-world datasets.

Future work may include model optimization, deployment on mobile devices, and integration with IoT-based smart agriculture systems.

ACKNOWLEDGMENT

The authors would like to thank the institution and all contributors who supported this research work.

REFERENCES

- [1] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017.
- [2] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," 2019.
- [3] C. Szegedy et al., "Rethinking the Inception Architecture for Computer Vision," 2016.
- [4] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2015.
- [5] J. Deng et al., "ImageNet: A Large-Scale Hierarchical Image Database," 2009.
- [6] Sachin D. Khirade and A. B. Patil: "Plant disease detection using image processing," in Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on, IEEE, 2015
- [7] S. B. Ullagaddi and S. Viswanadha Raju: "Disease Recognition in Mango Crop Using Modified Rotational Kernel Transform Features". 2017 Internationals.
- [8] Jamil Ahmad, Bilal Jan, Haleem Farman, Wakeel Ahmad, and Atta Ullah: "Disease Detection in Plum using Convolutional Neural Network under true field conditions".28 September 2020.
- [9] Lorente, D.; Escandell-Montero, P.; Cubero, S.; Gómez-Sanchis, J.; Blasco, J. Visible-NIR reflectance spectroscopy and manifold learning methods applied to the detection of fungal infections on pomegranate fruit. J. Food Eng. 2015, 163, 17–24.
- [10] Jahanbakhshi, A.; Momeny, M.; Mahmoudi, M.; Zhang, Y.D. Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks. Sci. Hortic. 2020, 263, 109133.
- [11] Inzamam Mashood Nasir, Asima Bibi, Jamal Hussain Shah, Deep Learning-Based Classification of Fruit Diseases: An Application for Precision Agriculture, Tech science press, Computers, Materials & Continua DOI:10.32604/cmc.2020.012945

