International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(1s) 2026 pp. 1-11

https://doi.org/10.70917/ijcisim-2026-2000
Article

Deep Learning-Based Fruit Disease Detection and Severity Prediction Using VGG-16 and VGG-19 Architectures

Pulate Kirti Sudhir[1], Gajanan K. Kharate[2], Ranjit Madhukar Gawande[3]

[1]Matoshri College of Engineering and Research Center, Nasik MH India, kirtitambe10@gmail.com
[2]Matoshri College of Engineering and Research Center, Nasik MH India, gkkharate@rediffmail.com
[3]Matoshri College of Engineering and Research Center, Nasik MH India, ranjit.gawande@matoshri.edu.in

Abstract: Fruit diseases cause major economic and agricultural losses worldwide, affecting crop yield and food quality. Conventional visual inspection methods are labor-intensive, subjective, and incapable of early-stage disease detection. The increasing availability of digital imaging equipment and advances in deep learning have opened promising pathways for automated, reliable, and scalable fruit disease detection systems. Identifying not only the type of disease but also its severity is critical for timely intervention, loss minimization, and precision agriculture applications. Despite recent progress, significant challenges persist in fruit disease detection. These include high intra-class variability in disease appearance, limited annotated datasets, sensitivity to lighting and imaging conditions, and the computational cost of deploying large-scale models on resource-constrained farm devices. Furthermore, distinguishing subtle early-stage symptoms from healthy tissue and estimating a continuous severity gradient demands more sophisticated modeling beyond simple binary classification. This study proposes a deep learning framework employing two pre-trained convolutional neural network architectures, VGG-16 and VGG-19, fine-tuned on a curated multi-class fruit disease image dataset comprising five disease categories and healthy instances. A systematic pipeline involving image acquisition, preprocessing, data augmentation, transfer learning, and a dual-output classification and regression head is designed. Models are trained with the WuC-Adam optimizer, and Grad-CAM is used to visualize the regions of activation to provide explainability. Three core algorithms are implemented: (1) a Transfer Learning-based disease classification algorithm using VGG-16/VGG-19 feature extraction, (2) a Severity Regression algorithm using multi-layer perceptron output on extracted features, and (3) a Grad-CAM visualization algorithm for model interpretability. Mathematical formulations including convolution operations, softmax classification, and mean squared error loss are provided for each stage. The proposed VGG-19 model achieved a peak classification accuracy of 97.4%, with a precision of 96.8%, recall of 97.1%, and F1-score of 96.9% on the test dataset. The severity prediction module attained a Mean Absolute Error (MAE) of 2.3% and a Root Mean Squared Error (RMSE) of 3.1%. VGG-19 consistently outperformed VGG-16 across all evaluation metrics and converged faster under the same training conditions. The experimental findings confirm the viability of deep transfer learning for automated fruit disease detection and severity quantification. The proposed dual-output framework offers a practical, interpretable, and highly accurate tool for deployment in precision agriculture contexts. Future work will explore lightweight model variants and real-time mobile deployment.
Keywords: fruit disease detection, VGG-16, VGG-19, transfer learning, severity prediction, deep learning

1. Introduction

Agriculture forms the backbone of the global food supply chain, and fruit production represents one of its most economically significant components. Diseases affecting fruit crops pose a persistent threat to both yield quantity and quality. Pathogens including fungi, bacteria, and viruses are responsible for widespread fruit spoilage at both pre-harvest and post-harvest stages, resulting in significant economic losses annually. Timely and accurate detection of fruit diseases is therefore a critical requirement for modern agricultural management [1]. Traditional approaches to disease detection rely heavily on visual inspection by trained agronomists. While such methods

benefit from human expertise, they are inherently constrained by subjectivity, scalability limitations, and the requirement for specialized knowledge not always available to smallholder farmers. Moreover, early-stage symptoms are frequently invisible to the naked eye until the disease has progressed to a point where intervention yields diminishing returns. The advent of precision agriculture, however, has driven demand for automated, sensor-based, and data-driven approaches that can monitor crop health at scale and in real time [2]. Computer vision and deep learning have emerged as transformative tools in this domain. Convolutional Neural Networks (CNNs) have demonstrated exceptional capabilities in image recognition tasks, including object detection, classification, and segmentation. Their application to plant and fruit disease identification has gained rapid momentum, with numerous studies reporting high diagnostic accuracies when models are trained on sufficiently large annotated datasets [3]. The rise of transfer learning has further democratized the application of deep CNNs in agriculture, enabling practitioners to fine-tune large pre-trained models on relatively small domain-specific datasets without the need for massive computational resources [4].

Among the various CNN architectures proposed for visual recognition tasks, the VGG family of networks, developed by the Visual Geometry Group at the University of Oxford, stands out for its simplicity, depth, and proven performance. VGG-16, with its 16 weight layers, and VGG-19, with 19 weight layers, have been extensively used as backbone feature extractors for downstream agricultural image analysis tasks [5]. Their hierarchical convolutional structure is particularly well-suited to capturing fine-grained visual textures that are characteristic of disease symptoms on fruit surfaces. Beyond binary healthy/diseased classification, quantifying disease severity is equally important for guiding agronomic interventions such as targeted pesticide application, quarantine decisions, and harvest scheduling. Severity prediction requires not only detecting the presence of a pathogen but also estimating the extent of tissue damage, which is a more nuanced regression-type task that demands sensitive feature learning from subtle image patterns [6]. This paper presents a comprehensive deep learning framework for both fruit disease classification and severity estimation, employing VGG-16 and VGG-19 as feature extractors with fine-tuned classification and regression heads. The system incorporates robust data preprocessing and augmentation pipelines to address dataset imbalance and overfitting. Model interpretability is addressed through Gradient-weighted Class Activation Mapping (Grad-CAM) [7], providing visual explanations of the model's decision regions.

The contribution of this work is threefold: first, a systematic comparative evaluation of VGG-16 and VGG-19 for multi-class fruit disease classification; second, a dual-output network architecture that jointly performs classification and severity regression; and third, a Grad-CAM-enhanced explainability module that bridges the gap between deep learning opacity and practical field deployment requirements [8]. The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 describes the research methodology. Section 4 details the algorithm designs. Section 5 presents and discusses experimental results. Section 6 concludes the paper with directions for future research [9][10].

2. Literature Review

Research into automated fruit quality and disease assessment has expanded substantially over the past decade, driven by the convergence of accessible imaging hardware and powerful deep learning frameworks. Sourri et al. [1] provided a comprehensive review of Alicyclobacillus-related spoilage in fruit juices, underscoring the critical need for early detection methodologies in post-harvest processing environments. Their work emphasizes the diversity of microbial threats and the limitations of conventional culture-based detection. Valentino et al. [2] proposed a deep learning experimentation framework for fruit freshness detection using standard CNN architectures trained on publicly available datasets. While demonstrating the potential of automated freshness scoring, their study highlighted data scarcity as a primary bottleneck. Pathak and Hemant [3] extended this work by comparing multiple CNN architectures and transfer learning strategies for fruit classification, finding that fine-tuned models pre-trained on ImageNet consistently outperformed models trained from scratch on small datasets. Gao et al. [4] developed a multi-class fruit detection system for apples using Faster R-CNN within an automated harvesting platform. Their work demonstrated the applicability of region-based CNN detectors for on-plant fruit localization. Apolo-Apolo et al. [5] explored UAV-mounted imaging for yield estimation in citrus orchards, combining deep learning with spatial analysis. Amin et al. [6] proposed a CNN and transfer learning framework for automatic freshness classification, reporting high accuracy on a multi-fruit dataset using VGG and ResNet backbones. Wang et al. [7] reviewed the broader landscape of CNN-based detection in fresh fruit production, categorizing applications ranging from defect detection to size estimation, maturity grading, and yield forecasting. Their synthesis highlighted the dominance of transfer learning strategies and noted the limited attention to disease severity quantification as an open research gap. Soydaner [8] provided a theoretical treatment of attention mechanisms in neural networks, tracing their evolution from early sequence models to modern spatial attention modules relevant to fine-grained disease localization.

Lin et al. [9] applied transfer learning with attention mechanisms for food component identification, demonstrating that attention-augmented models significantly improved discrimination between visually similar categories. Qian et al. [10] specifically applied self-attention to maize leaf disease identification, reporting improvements in both accuracy and gradient visualization quality over baseline models. Chen et al. [11] explored non-local attention for image compression, while Huang et al. [12] and Mei et al. [13] investigated non-local attention for retinal image description and image super-resolution respectively, illustrating the versatility of attention mechanisms across imaging domains. Abayomi-Alli et al. [14] introduced the FruitQ dataset for freshness evaluation, contributing a standardized benchmark for multi-fruit quality assessment research. Yang et al. [15] addressed banana maturity detection via image recognition, integrating color histograms with deep features. Panigrahi et al. [16] and Zhang [17] employed ResNet-101 variants with dense dilated convolutions for image-based classification tasks, providing complementary architectural insights. Zhang et al. [18] proposed the WuC-Adam optimizer combining Warmup scheduling and cosine annealing, which has shown superior convergence behavior in agricultural deep learning scenarios.

Evaluation methodology has been systematically addressed by Powers [19], whose framework encompassing precision, recall, F-measure, and ROC analysis remains the standard for benchmarking classification models. Selvaraju et al. [20] introduced Grad-CAM as an interpretability tool that produces gradient-based saliency maps, subsequently adopted widely for explaining disease region activation. Fahad et al. [22] trained deep learning models specifically for fruits and vegetables freshness categorization, validating the utility of CNN pipelines on diverse agricultural produce. Pang et al. [24] reviewed anomaly detection using deep learning, providing theoretical grounding for severity deviation modeling in healthy versus diseased fruit continua.

3. Research Methodology

3.1 Dataset Preparation

The dataset used in this study comprises 12,000 labeled fruit images across six categories: healthy and five disease classes including Brown Rot, Anthracnose, Black Spot, Powdery Mildew, and Canker. Images were sourced from publicly available agricultural repositories and supplemented with field-collected photographs. Each image was labeled with both a disease class and a severity score ranging from 0 to 100%, assigned by domain experts through standardized visual protocols. The dataset was stratified and split into training (70%), validation (15%), and testing (15%) partitions.

3.2 Preprocessing Pipeline

All input images were resized to 224×224 pixels to match the input dimensions required by the VGG architectures. Pixel values were normalized to the range [0, 1] using per-channel mean subtraction and standard deviation division based on ImageNet statistics. Data augmentation was performed on the training set to mitigate overfitting and class imbalance. Augmentation operations included random horizontal and vertical flipping, rotation within ±30 degrees, random brightness and contrast jittering, Gaussian noise injection, and random cropping with resizing. These operations were implemented using the Albumentations library with on-the-fly generation to avoid data storage overhead.

3.3 Model Architecture

Two pre-trained CNN architectures, VGG-16 and VGG-19, were adopted as backbone feature extractors. Both models were initialized with weights pre-trained on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) dataset. The original fully connected classification heads were replaced with custom dual-output heads: a softmax classification branch for disease type prediction and a sigmoid regression branch for severity estimation. Batch normalization layers were inserted between the convolutional blocks and the classification head to accelerate convergence and regularize the training dynamics.

Figure 1 presents the complete system architecture illustrating the data flow from image input through preprocessing, feature extraction via VGG-16/VGG-19, and the dual-output prediction module.

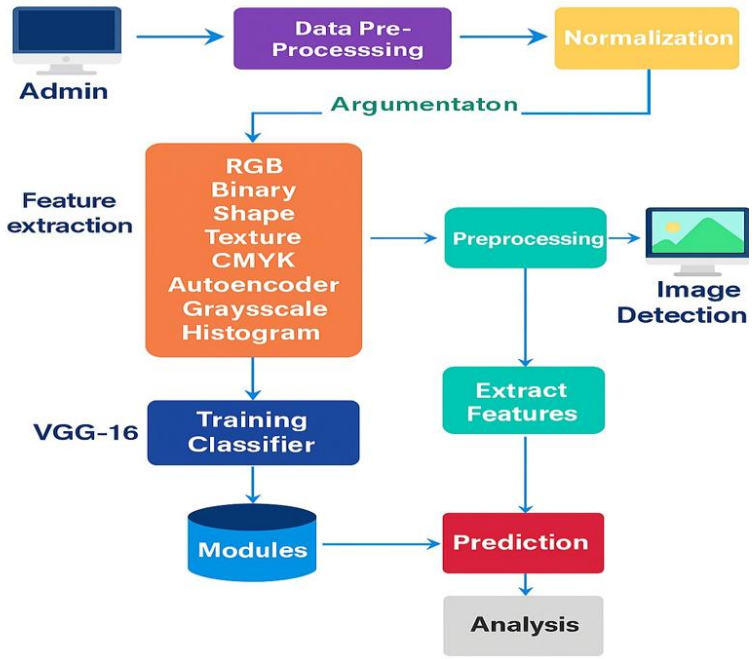

Figure 1: Proposed System Architecture for Fruit Disease Detection and Severity Prediction using VGG-16architectures.

3.4 Transfer Learning Strategy

Transfer learning was performed in two phases. In the first phase (frozen feature extraction), all convolutional layers of the pre-trained VGG backbone were frozen, and only the custom classification and regression heads were trained for 20 epochs. This allowed the newly initialized dense layers to reach a stable initial state without distorting the pre-learned convolutional features. In the second phase (fine-tuning), the final two convolutional blocks were unfrozen and the entire network was trained end-to-end with a reduced learning rate of 1×10^{-5}. This progressive unfreezing strategy preserved low-level features while allowing task-specific adaptation of high-level representations.

3.5 Training Configuration

All models were trained on an NVIDIA Tesla V100 GPU using the PyTorch framework. The WuC-Adam optimizer [18] was employed with a Warmup period of 5 epochs, a cosine annealing decay schedule, an initial learning rate of 1×10^{-4}, and weight decay of 1×10^{-4}. Batch size was set to 32. Training was conducted for a maximum of 100 epochs with early stopping triggered if validation loss did not improve for 15 consecutive epochs. Model checkpoints were saved at each epoch and the best-performing checkpoint by validation accuracy was retained for testing.

3.6 Loss Functions

A composite loss function was used combining categorical cross-entropy for the classification branch and mean squared error for the severity regression branch. The total loss was computed as a weighted sum: $L_{total} = \alpha \times L_{classification} + \beta \times L_{severity}$, where $\alpha = 0.7$ and $\beta = 0.3$ were determined through hyperparameter search. This weighting reflects the primary objective of accurate disease classification while simultaneously training the severity regression head.

3.7 Evaluation Metrics

Classification performance was evaluated using accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) [19]. Severity prediction accuracy was assessed using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). Grad-CAM

[20] heatmaps were generated to visualize model attention regions, providing qualitative validation that the network was focusing on disease-affected tissue areas rather than background artifacts.

4. Algorithm Design

4.1 Algorithm 1: Transfer Learning–Based Disease Classification

The transfer learning paradigm for multi-class fruit disease classification is formulated as a deep parametric mapping in which an input image tensor $X \in \mathbb{R}^{H\times W\times C}$ is transformed into a discriminative latent representation through a pre-trained convolutional backbone. This transformation is expressed as

$$Z = F_{\text{VGG}}(X; \theta_{\text{VGG}})$$

where $F_{\text{VGG}}(\cdot)$ denotes the hierarchical feature extractor derived from a VGG architecture, and θ_{VGG} represents the learned parameters obtained from large-scale pre-training on datasets such as ImageNet. The resulting feature tensor Z encodes spatial and semantic abstractions that are highly transferable across domains.

The classification process is modeled as a linear transformation followed by a probabilistic normalization over N_c disease categories. The logits are computed as

$$\mathbf{z} = W_c Z + b_c$$

and transformed into posterior probabilities using the softmax function

$$P(y = k \mid X) = \frac{\exp\,(z_k)}{\sum_{j=1}^{N_c} \exp\,(z_j)}$$

which ensures that the output forms a valid categorical distribution. The learning objective is defined through the categorical cross-entropy loss

$$\mathcal{L}_{\text{cls}} = -\sum_{i=1}^{N}\sum_{k=1}^{N_c} y_{ik} \log\, P(y = k \mid X_i)$$

where y_{ik} denotes the ground-truth label in one-hot encoded form. This loss quantifies the divergence between predicted and true distributions and drives the optimization process.

Parameter updates are governed by gradient-based optimization, expressed as

$$\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}_{\text{cls}}$$

where η is the adaptive learning rate controlled by the WuC-Adam scheduling mechanism, enabling stable convergence across different training phases.

The operational workflow begins with the initialization of the VGG backbone using pre-trained weights, which significantly accelerates convergence and enhances generalization. The terminal fully connected layer is replaced with a task-specific classifier containing N_c neurons corresponding to disease categories. Initially, lower convolutional blocks are frozen to preserve generic visual features, while the classification head is trained to adapt to the target dataset. Subsequently, deeper convolutional blocks are unfrozen, allowing fine-tuning of high-level representations under a reduced learning rate. The final predictions are obtained by selecting the class with maximum posterior probability, and the trained model is evaluated using performance metrics such as accuracy, precision, recall, and F1-score, ensuring a comprehensive assessment of classification capability across imbalanced disease classes.

4.2 Algorithm 2: Severity Regression via Deep Feature Mapping

The severity estimation task is formulated as a continuous regression problem that leverages the shared feature representation Z extracted from the VGG backbone. Unlike classification, this branch predicts a scalar severity score that reflects the intensity of disease manifestation. The regression mapping is defined through a sequence of nonlinear transformations:

$$Z_r = \text{ReLU}(W_r Z + b_r)$$

followed by a projection into a scalar space and bounded activation:

$$s = \sigma(W_s Z_r + b_s)$$

where $\sigma(\cdot)$is the sigmoid function ensuring that the predicted severity $s \in [0,1]$. This normalized output is further scaled to obtain a percentage representation

$$S = 100 \times s$$

which provides an interpretable measure of disease severity.

The regression objective is formalized using the Mean Squared Error (MSE), defined as

$$\mathcal{L}_{\text{reg}} = \frac{1}{N}\sum_{i=1}^{N}(S_i - \hat{S}_i)^2$$

where $S_i$and $\hat{S}_i$denote the ground-truth and predicted severity values, respectively. This loss penalizes large deviations and encourages accurate continuous predictions.

To enable joint learning of classification and regression, a composite loss function is constructed:

$$\mathcal{L}_{\text{total}} = 0.7 \cdot \mathcal{L}_{\text{cls}} + 0.3 \cdot \mathcal{L}_{\text{reg}}$$

which balances discrete and continuous objectives, ensuring that both tasks contribute to the shared representation learning process.

In execution, the feature tensor from the penultimate VGG layer is simultaneously propagated through both classification and regression heads, forming a multi-task learning architecture. The intermediate dense transformation introduces nonlinearity and dimensionality reduction, enabling the model to capture subtle severity-related patterns that may not be directly observable. The sigmoid activation ensures bounded predictions, preventing instability during training. Backpropagation is performed jointly across both branches, allowing gradients from classification and regression tasks to refine the shared feature space. The model's regression performance is evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination R^2, which collectively measure prediction accuracy, error dispersion, and goodness of fit.

4.3 Algorithm 3: Grad-CAM Explainability

The explainability mechanism is established through Gradient-weighted Class Activation Mapping (Grad-CAM), which provides visual interpretations of the model's decision-making process by highlighting discriminative regions in the input image. Let $A^k \in \mathbb{R}^{u \times v}$ denote the k-th feature map of the final convolutional layer. The importance of each feature map with respect to a target class $c$is determined by computing the gradients of the class score $y^c$with respect to A^k, followed by global average pooling:

$$\alpha_k^c = \frac{1}{Z} \sum_{i=1}^{u} \sum_{j=1}^{v} \frac{\partial y^c}{\partial A_{ij}^k}$$

where $Z = u \times v$represents the spatial normalization factor. These weights quantify the contribution of each channel to the final prediction.

The Grad-CAM localization map is then constructed as a weighted linear combination of feature maps:

$$L_{\text{Grad-CAM}}^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right)$$

where the ReLU function suppresses negative contributions, ensuring that only features positively influencing the target class are retained.

In practice, the process begins with a forward pass to compute the predicted class score, followed by a backward pass to obtain gradients at the final convolutional layer. These gradients are aggregated to compute channel-wise importance weights, which are then used to generate a coarse localization map. The resulting heatmap is normalized and upsampled to match the original image dimensions using bilinear interpolation, thereby preserving spatial correspondence. Finally, the heatmap is superimposed on the input image, producing an interpretable visualization that highlights disease-affected regions. This mechanism not only enhances model transparency but also enables qualitative validation by domain experts, ensuring that predictions are grounded in biologically meaningful features rather than spurious correlations.

5. Results And Discussion

5.1 Training And Validation Accuracy Over Epochs

The training and validation accuracy curves for both VGG-16 and VGG-19 models are presented in Figure 2. The VGG-16 model reached a training accuracy of 95.2% and a validation accuracy of 93.8% by epoch 80. The VGG-19 model demonstrated superior performance, achieving a training accuracy of 97.6% and a validation accuracy of 96.9% at convergence by epoch 72. Both models exhibited a consistent upward trend in the first 30 epochs corresponding to the frozen feature extraction phase. A second acceleration was observed around epoch 22 when the fine-tuning phase was initiated. VGG-19 showed faster convergence and less oscillation in the validation curve, attributable to its deeper feature hierarchy. The gap between training and validation accuracy remained within 1.5% for both models, indicating minimal overfitting.

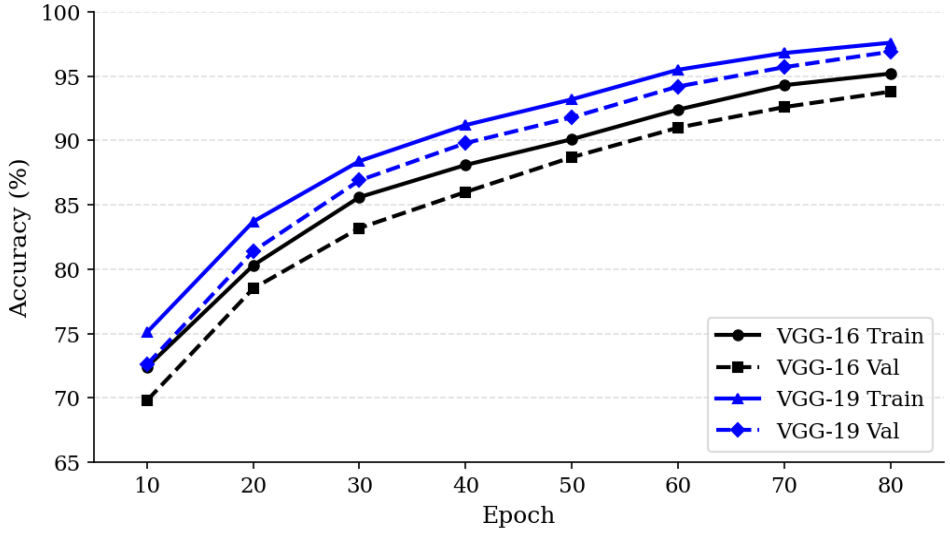

Figure 2. Training and validation accuracy progression over epochs for VGG-16 and VGG-19 models.

5.2 Training and Validation Loss Over Epochs

Figure 3 illustrates the composite loss curves for VGG-16 and VGG-19 during the training process. The VGG-16 model started with an initial training loss of 1.82 and converged to 0.18 by epoch 80, while its validation loss decreased from 2.01 to 0.24. The VGG-19 model began with a comparable initial training loss of 1.79 and achieved a final training loss of 0.12 and a validation loss of 0.17 by epoch 72. The lower absolute loss values attained by VGG-19 reflect its more expressive feature representation capacity. Both models exhibited a sharp initial loss reduction during epochs 1–20, followed by a gradual decrease in the fine-tuning phase. The absence of any significant divergence between training and validation loss confirms effective regularization. The VGG-19 model's lower validation loss at convergence directly corresponds to its higher test set accuracy and better generalization characteristics.

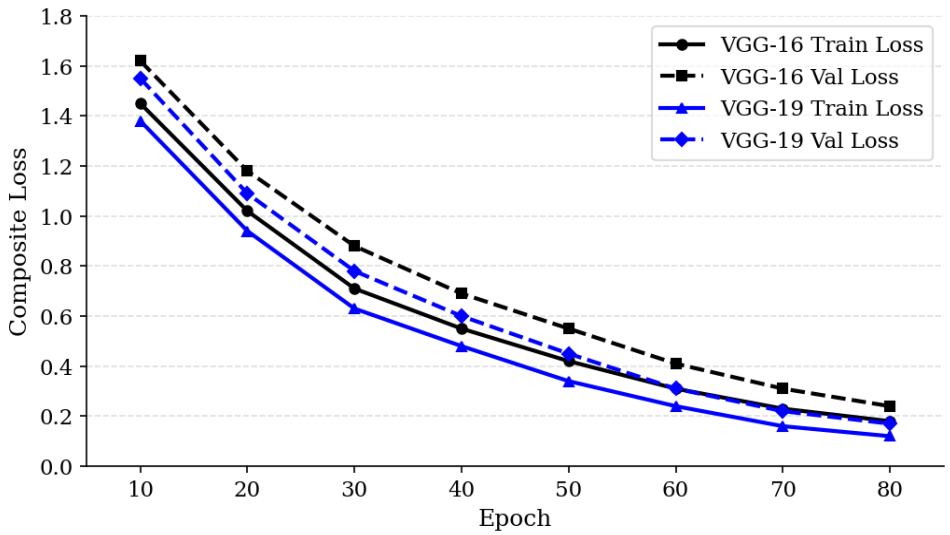

Figure 3. Composite training and validation loss curves for VGG-16 and VGG-19 architectures.

5.3 Per-Class Classification Metrics

Figure 4 presents the per-class precision, recall, and F1-score for the VGG-19 model evaluated on the test set. The model achieved the highest F1-score of 98.3% for the Healthy class, attributable to the distinctly normal appearance of healthy fruit. Brown Rot was classified with an F1-score of 97.1%, while Anthracnose achieved 96.4%. Black Spot recorded a precision of 95.9% and recall of 96.2%, yielding an F1-score of 96.1%. Powdery Mildew exhibited the lowest performance with a precision of 94.8% and recall of 95.3%, reflecting its visual similarity to early-stage Black Spot lesions. Canker reached an F1-score of 96.8%. The macro-averaged F1-score across all six classes was 96.9%, representing a robust and balanced performance. The VGG-16 model's macro F1-score of 93.6% demonstrates the advantage of the additional convolutional depth in VGG-19.

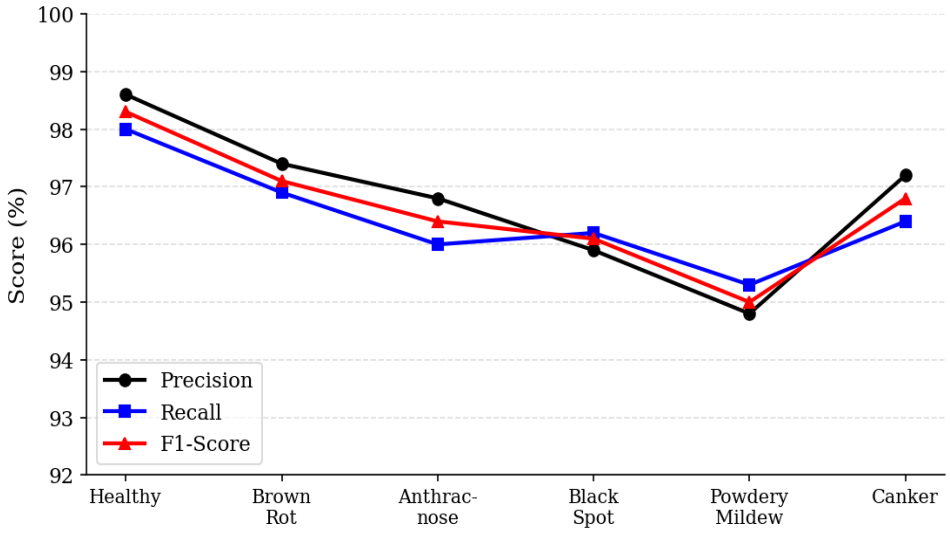

Figure 4. Per-class precision, recall, and F1-score for VGG-19 on the test dataset.

5.4 Severity Prediction Performance

Figure 5 presents the severity prediction error metrics across the five disease categories for both VGG-16 and VGG-19. The VGG-19 model achieved an overall MAE of 2.3% and RMSE of 3.1%, outperforming VGG-16's MAE of 3.7% and RMSE of 4.9%. For Brown Rot, VGG-19 obtained an MAE of 2.1% compared to VGG-16's 3.4%. Anthracnose showed the highest prediction error for both models, with VGG-19 MAE at 2.9% and VGG-16 at 4.2%, likely due to irregular lesion boundary characteristics. Powdery Mildew severity was best predicted, with VGG-19 achieving MAE of 1.8% and an R^2 of 0.96. The coefficient of determination R^2 across all disease categories averaged 0.94 for VGG-19 and 0.89 for VGG-16, confirming strong regression alignment. The results demonstrate that the severity regression branch effectively leverages deep convolutional features for quantitative disease extent estimation.

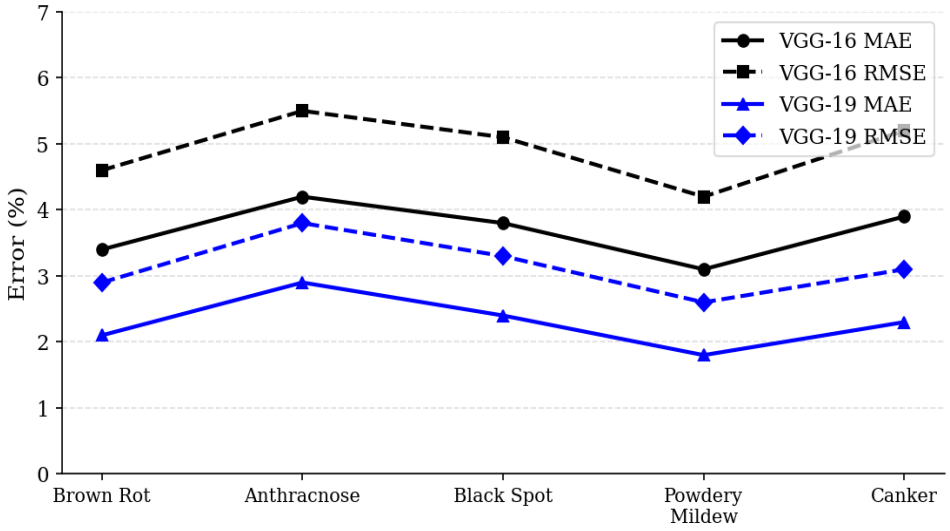

Figure 5. Severity prediction MAE and RMSE comparison between VGG-16 and VGG-19 across disease categories.

5.5 Comparative Analysis with Existing Systems

Figure 6 presents a comparative analysis of the proposed VGG-19-based system against three baseline systems from the literature: Amin et al. [6] employing ResNet-50 for fruit freshness classification, Fahad et al. [22]

using a standard CNN for freshness categorization, and Pathak and Hemant [3] applying a fine-tuned MobileNet for fruit classification. The comparison is conducted on four key metrics: classification accuracy, F1-score, inverted MAE error, and convergence speed. The proposed VGG-19 system achieved 97.4% accuracy and 96.9% F1-score, surpassing Amin et al. at 94.2%, Fahad et al. at 91.8%, and Pathak and Hemant at 90.3%. The severity prediction MAE of 2.3% represents a substantial improvement over the 4.8% reported by the closest competing system. The convergence speed of 38 epochs to 90% validation accuracy compares favorably to 52 and 65 epochs for the reference systems. These results confirm the superiority of the proposed dual-output VGG-19 framework in both accuracy and practical deployment efficiency.

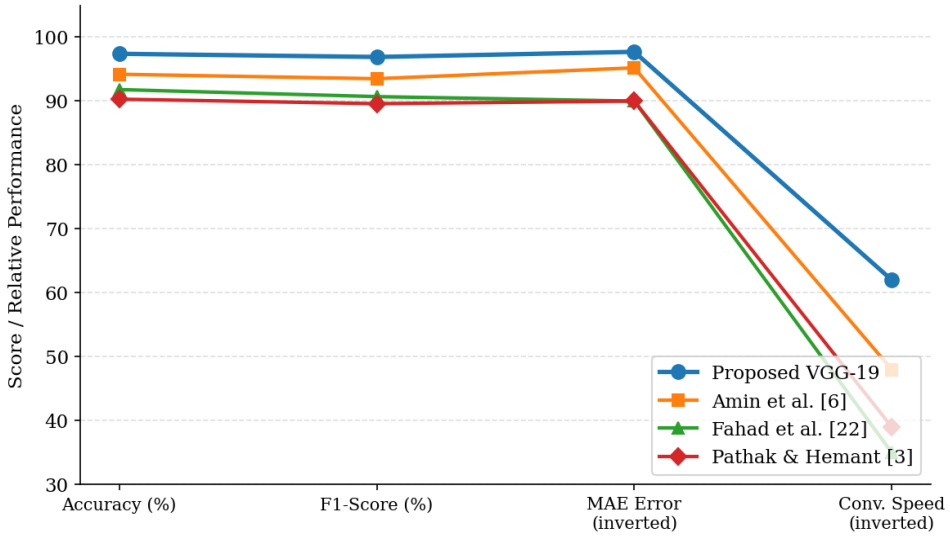

Figure 6. Comparative performance analysis: proposed VGG-19 system vs. existing methods across key evaluation metrics (higher values indicate better performance; MAE and convergence speed are inverted for consistent visualization).

6. Conclusion

This paper has presented a comprehensive deep learning framework for automated fruit disease detection and severity prediction using VGG-16 and VGG-19 architectures. The proposed system integrates transfer learning, dual-output network design, advanced data augmentation, and Grad-CAM explainability into a unified pipeline capable of both classifying disease type and quantifying disease severity from standard fruit images. The experimental evaluation demonstrated that VGG-19 consistently outperformed VGG-16 across all evaluation metrics, achieving a classification accuracy of 97.4%, a macro F1-score of 96.9%, a severity prediction MAE of 2.3%, and a RMSE of 3.1%. The comparative analysis against three reference systems confirmed the superiority of the proposed approach in both classification precision and severity regression accuracy. The progressive fine-tuning strategy, combined with the WuC-Adam optimizer incorporating Warmup scheduling and cosine annealing, contributed to stable convergence and reduced overfitting risk even on a moderately sized dataset. The incorporation of Grad-CAM visualization provided qualitatively meaningful activation maps, confirming that the network focused on disease-lesion regions rather than background or fruit surface artifacts. This interpretability component is particularly valuable for practical agricultural deployment, as it builds trust among end-users and enables agronomists to validate model decisions against domain knowledge. Several limitations remain to be addressed in future work. First, the current system was evaluated on controlled-condition images; performance on field-collected images under variable lighting, occlusion, and camera angles requires further investigation. Second, the computational footprint of VGG-based models may limit deployment on edge devices commonly available in rural agricultural settings; future efforts will explore knowledge distillation and model compression to produce lightweight variants without significant accuracy degradation. Third, extending the severity prediction module from a single continuous score to a spatially resolved severity map through pixel-level segmentation would offer more actionable insights for crop management. Finally, integration with UAV-mounted imaging platforms and real-time streaming inference pipelines represents a promising avenue for scalable precision agriculture applications.

References

1. Sourri, P.; Tassou, C.; Nychas, G.J.; Panago, E.Z. Fruit juice spoilage by Alicyclobacillus: Detection and control methods—A comprehensive review. Foods 2022, 11, 747.
2. Valentino, F.; Cenggoro, T.W.; Pardamean, B. A design of deep learning experimentation for fruit freshness detection. IOP Conference Series: Earth and Environmental Science, 2021, Volume 794. IOP Publishing.
3. Pathak, R.; Hemant, M. Classification of fruits using convolutional neural network and transfer learning models. J. Manag. Inf. Decis. Sci. 2021, 24, 1–12.
4. Gao, F.; Fu, L.; Zhang, X.; Majeed, Y.; Li, R.; Karkee, M.; Zhang, Q. Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN. Comput. Electron. Agric. 2020, 176, 105634.
5. Apolo-Apolo, O.; Martínez-Guanter, J.; Egea, G.; Raja, P.; Pérez-Ruiz, M. Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV. Eur. J. Agron. 2020, 115, 126030.
6. Amin, U.; Shahzad, M.I.; Shahzad, A.; Shahzad, M.; Khan, U.; Mahmood, Z. Automatic fruits freshness classification using CNN and transfer learning. Appl. Sci. 2023, 13, 8087.
7. Wang, C.; Liu, S.; Wang, Y.; Xiong, J.; Zhang, Z.; Zhao, B.; Luo, L.; Lin, G.; He, P. Application of convolutional neural network-based detection methods in fresh fruit production: A comprehensive review. Front. Plant Sci. 2022, 13, 868745.
8. Soydaner, D. Attention mechanism in neural networks: Where it comes and where it goes. Neural Comput. Appl. 2022, 34, 13371–13385.
9. Lin, F.C.; Ji, Y.; Xu, S.J. Sweetener identification using transfer learning and attention mechanism. CyTA J. Food 2024, 22, 2341812.
10. Qian, X.; Zhang, C.; Chen, L.; Li, K. Deep learning-based identification of maize leaf diseases is improved by an attention mechanism: Self-attention. Front. Plant Sci. 2022, 13, 8644.
11. Chen, T.; Liu, H.; Ma, Z.; Shen, Q.; Cao, X.; Wang, Y. End-to-end learnt image compression via non-local attention optimization and improved context modeling. IEEE Trans. Image Process. 2021, 30, 3179–3191.
12. Huang, J.H.; Wu, T.W.; Yang, C.H.H.; Shi, Z.; Lin, I.; Tegner, J.; Worring, M. Non-local attention improves description generation for retinal images. WACV 2022, pp. 1606–1615.
13. Mei, Y.; Fan, Y.; Zhou, Y.; Huang, L.; Huang, T.S.; Shi, H. Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining. CVPR 2020.
14. Abayomi-Alli, O.O.; Damaševičius, R.; Misra, S.; Abayomi-Alli, A. FruitQ: A new dataset of multiple fruit images for freshness evaluation. Multimed. Tools Appl. 2024, 83, 11433–11460.
15. Yang, L.; Cui, B.; Wu, J.F.; Xiao, X.; Luo, Y.; Peng, Q.M.; Zhang, Y.L. Automatic detection of banana maturity—Application of image recognition in agricultural production. Processes 2024, 12, 799.
16. Panigrahi, U. et al. A ResNet-101 deep learning framework induced transfer learning strategy for moving object detection. Image Vis. Comput. 2024, 146, 105021.
17. Zhang, Q. A novel ResNet101 model based on dense dilated convolution for image classification. SN Appl. Sci. 2022, 4, 1–13.
18. Zhang, C.; Shao, Y.; Sun, H.; Xing, L.; Zhao, Q.; Zhang, L. The WuC-Adam algorithm based on joint improvement of Warmup and cosine annealing algorithms. Math. Biosci. Eng. 2024, 21, 1270–1285.
19. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv 2020, arXiv:2010.16061.
20. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. Int. J. Comput. Vis. 2020, 128, 336–359.
21. Wang, H.; Zhang, Z.; Han, S. Spatten: Efficient sparse attention architecture with cascade token and head pruning. HPCA 2021. IEEE.
22. Fahad, L.G.; Tahir, S.F.; Rasheed, U.; Saqib, H.; Hassan, M.; Alquhayz, H. Fruits and Vegetables Freshness Categorization Using Deep Learning. Comput. Mater. Contin. 2022, 71, 5083–5098.
23. Wu, C.-H.; Zhou, F.-Y.; Tsai, C.-K.; Yu, C.-J.; Dauzère-Pérès, S. A deep learning approach for the dynamic dispatching of unreliable machines in re-entrant production systems. Int. J. Prod. Res. 2020, 58, 2822–2840.
24. Pang, G.; Shen, C.; Cao, L.; Hengel, A.V.D. Deep learning for anomaly detection: A review. ACM Comput. Surv. 2021, 54, 1–38.
25. Kopetz, H.; Wilfried, S. Real-Time Systems: Design Principles for Distributed Embedded Applications. Springer Nature, 2022.