



A Hybrid CNN Ensemble Approach with Feature-Level Fusion for Apple Leaf Disease Classification

Susovan Dutta¹, Snehasish Das¹, Parthajit Banik¹, Md Kaif¹

¹Dept. of Computer Science and Engineering, Narula Institute of Technology, Kolkata, India —

duttasusovan4@gmail.com, sid29082001@gmail.com, parthajitbanik6@gmail.com, md.kaif.cse@gmail.com

Abstract—The deterioration of apple leaves through disease imposes a tremendous impact on the quality of fruit production from apple orchards. Therefore, swift and accurate identification of these diseases is necessary for effective fruit management. Traditional methods for identifying apple leaf diseases utilize expert humans who inspect the leaves visually. However, visual inspections are labor-intensive, highly subjective, and cannot be conducted effectively in larger orchards. The rapid advancement of deep learning research is facilitating automated identification of apple leaf diseases from images of contaminated leaves using Convolutional Neural Networks (CNNs) that are able to learn visual characteristics such as spots, lesions, colour changes, or texture differences on apple tree leaves. However, when using a single CNN model to accomplish this task, one may encounter difficulties due to environmental variability, similarities between classes of diseases, and generalisation across variable field conditions. The current work proposes a hybrid ensemble CNN model for the analysis of apple leaf disease by providing a solution to these challenges. The ensemble model consists of EfficientNet-B0, DenseNet121, and ResNet50. Feature-level fusion of these models concurrently enables the extraction of strong and complementary feature representations from images for improved performance. The proposed hybrid CNN ensemble model achieved a classification accuracy of 99% when evaluated on a dataset of apple leaf disease images. Furthermore, the proposed model consistently outperformed backbone models such as EfficientNet-B0, VGG16, ResNet50, and MobileNetV2. Therefore, this proposed model has the potential to be an effective, reliable, scalable method to perform disease monitoring in apple production systems.

Keywords—Apple leaf disease Classification, Hybrid CNN Ensemble Model, EfficientNet-B0, DenseNet121, ResNet50

I. INTRODUCTION

The agricultural industry plays a vital role in sustaining the global population, and plant health directly impacts crop yield and food quality. Among various crops, apple cultivation is of significant economic importance worldwide, and apple leaf diseases directly affect fruit yield, quality, and long-term orchard productivity. Plant diseases, if not detected at an early stage, can result in severe economic losses and reduced agricultural productivity. Conventional plant disease diagnosis depends on manual visual inspection by trained experts, which is labor-intensive, time-consuming, and unsuitable for largescale or remote agricultural fields. Recent developments in deep learning (DL) and computer

vision have made it possible to use leaf images for automated detection of plant disease. For apple, leaf-based visual symptoms such as discoloration, lesions, and texture variations provide crucial indicators of disease presence. CNNs are particularly effective in learning hierarchical and discriminative features such as color variations, lesion patterns, texture irregularities, and shape distortions directly from raw images [1]. When compared to conventional image processing and machine learning (ML) methods, DL-based approaches have continuously shown better performance [2], [3]. Large-scale image datasets and improved computational resources have further accelerated the adoption of CNN-based models for plant disease classification, including apple leaf disease recognition [4], [5]. Studies have shown that

deep architectures can effectively handle complex visual patterns under controlled conditions [6], [7]. However, deploying these models for apple leaf disease detection in realworld agricultural environments remains challenging due to variations in lighting, background clutter, and disease similarity [8], [9].

Various CNN architectures such as VGGNet [14], ResNet [13], MobileNet [11], and EfficientNet [12] have been explored for detection of plant disease. With differing degrees of success, these architectures have also been used to classify apple leaf diseases. While deeper models achieve high accuracy, they often require high computational resources, limiting their usability in resource-constrained environments [10]. Lightweight models improve efficiency but may fail to capture sufficient discriminative features for visually similar apple leaf diseases [16], [17]. To address these issues, hybrid DL models have emerged as an effective solution by combining multiple CNN architectures to enhance feature diversity and robustness. Hybrid models leverage complementary feature representations extracted from different networks, improving classification performance under complex conditions. Motivated by these observations, this work propounds a hybrid CNN ensemble framework that integrates EfficientNet-B0, DenseNet121, and ResNet50 for accurate and efficient apple leaf disease classification.

The organization of the subsequent sections of this paper is as follows: Section 2 contains a summary of previous research on this topic; Section 3 describes the proposed work; Section 4 provides an analysis of the experimental findings; Section 5 concludes the paper.

II. RELATED WORK

The development of advanced plant disease recognition and classification was initially made possible through classical procedures of image processing that included the use of visual stimuli including colour histograms, texture elements and shape characteristics [3], [19]. These types of classical approaches required expert knowledge and showed limited robustness under varying environmental conditions. The introduction of DL systems that utilize CNN's for automated identification and recognition of plant disease using computergenerated models has improved upon previous methods of classification. CNNs have achieved accuracy rates comparable to horizon data sets listed in the PlantVillage Data Set by Mohanty et al. [4], particularly when generating systems for the identification of plant diseases in multiple plant varieties. The research of Ferentinos et al. [2] has shown that DL systems using CNN architectures are superior performers when compared to performance levels from standard classification methodologies. Sladojevic et al. [5] have indicated that DL networks that utilize an end to end learning model, exceed the functional output levels for traditionally

designed identification systems using neural networks.

Several studies analyzed the influence of image quality and background complexity on model performance. Barbedo et al. [6], [15] reported that uncontrolled field conditions significantly affect classification accuracy. In a comparative study of optimised CNN architectures, Too et al. [7] showed how different networks performed differently. Lightweight and efficient CNN architectures have been proposed to improve deployability. Howard et al. [11] introduced MobileNet for mobile vision applications. On the other hand, Tan and Le [12] proposed EfficientNet with compound scaling to balance accuracy and efficiency. Deep residual learning using ResNet was introduced by He et al. [13]. Simonyan and Zisserman [14] explored very deep CNNs. Real-time and mobile-based plant disease detection systems were investigated by Jiang et al. [16] and Picon et al. [17], demonstrating the feasibility of on-device inference. Chouhan et al. [18] proposed dataset development and augmentation strategies to improve model generalization. DL approaches were also applied to specific crops such as cassava [20], cucumber [19], citrus [21] and apple leaves [16]. Recent studies explored advanced architectures and alternative learning paradigms. Dosovitskiy et al. [22] introduced Vision Transformers for image recognition tasks. Severity estimation using DL was addressed by Wang et al. [23]. Li et al. [24] questioned the necessity of very deep CNNs and highlighted the need for more robust and efficient models.

Several studies have specifically focused on apple leaf disease detection due to the economic importance of apple cultivation and the visual distinctiveness of leaf symptoms. Jiang et al. [27] applied CNN-based approaches to classify apple leaf diseases such as Apple Scab (AS), Black Rot (BR), and Cedar Apple Rust (CAR) using image datasets captured under both controlled and field conditions. These studies highlight the effectiveness of DL for apple leaf disease classification, while also reporting challenges related to background variation, disease similarity, and model generalization. Recent studies have focused on apple leaf disease classification due to its economic importance and the distinct visual symptoms of affected leaves. In order to classify major apple leaf diseases with competitive accuracy, Liu et al. [26] proposed deep CNNs. Wang et al. [27] investigated severity estimation for apple leaf diseases using DL, highlighting challenges in varied lighting and background conditions.

Despite significant progress in DL-based plant disease classification, most existing studies rely on single backbone CNN architectures. Such models often struggle to distinguish visually similar apple leaf diseases and fail to maintain robustness under real-world variations in lighting, background, and symptoms appearance. Hybrid and ensemble-based approaches for apple leaf disease classification remain relatively underexplored. This research gap motivates the proposed hybrid CNN framework that integrates

EfficientNet-B0, DenseNet121, and ResNet50 to achieve improved accuracy, robustness, and practical applicability for apple leaf disease classification.

III. PROPOSED METHODOLOGY

The proposed plant disease classification approach highlights a pipeline consisting of dataset collection, data preprocessing, hybrid CNN ensemble model design, system implementation, and performance evaluation. The overall methodology is designed to leverage complementary feature representations extracted from multiple DL architectures. Fig. 1 presents the overall workflow of the proposed apple disease classification model, illustrating each stage from data collection and preprocessing to hybrid CNN ensemble model-based classification and performance evaluation.

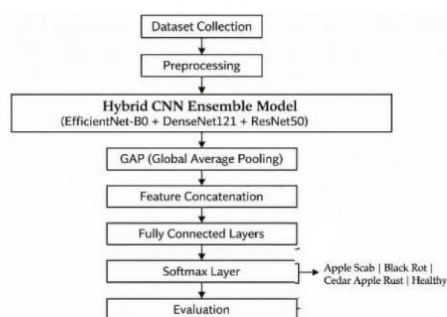


Fig. 1. Proposed workflow of the hybrid CNN ensemble model for apple leaf disease classification

A. Data Collection

This study's dataset [28] is gathered from Kaggle. This dataset contains four classes: AS, BR, CAR, and Healthy leaves. Of the 12,000 images in the dataset, 2,400 were used for validation and 9,600 were used for training. The dataset offers visually varied and well-labeled samples that are appropriate for analyzing DL-based apple leaf disease classification models.

B. Data Preprocessing

To ensure compatibility across the hybrid network branches, all input images were resized to 224×224 pixel resolution. To stabilize training, pixel values were normalized to the range of [0, 1]. Class labels were converted into one-hot encoded vectors to match the requirements of multi-class classification. During training, data augmentation techniques such as random rotation, vertical and horizontal flipping, zooming, and brightness adjustment were used to increase generalization and decrease overfitting. These augmentations simulate real-world variations commonly observed in field conditions.

C. Hybrid CNN Ensemble Model Design

The proposed hybrid CNN ensemble model integrates three pretrained CNN architectures:

EfficientNet-B0, DenseNet121, and ResNet50. Each backbone is initialized with ImageNet weights and operates in parallel, receiving the same input image. The convolutional bases are used as feature extractors to capture diverse and complementary visual representations. A Global Average Pooling (GAP) layer is used for each backbone in order to extract compact feature vectors and reduce spatial dimensions. The resulting feature vectors from all three networks are then concatenated to form a unified feature representation. This feature-level fusion allows the model to integrate residual, deep, and fine-grained features that were acquired by various architectures. A Fully Connected Classification Head is used to process the Fused Feature Vector. This head consists of the following:

- A dense layer with 512 neurons and ReLU activation
- A dense layer with 256 neurons and ReLU activation
- An output softmax layer with 4 Neurons corresponding

to the 4 target classes The hybrid architecture have increased the capacity of the fusion network and increased the robustness of the model against visually similar disease classes.

D. Hybrid System Implementation

The hybrid CNN ensemble model with Baseline Models were implemented using the Tensorflow and Keras Frameworks. The Adam Optimizer was used to train the model and the loss function is defined as:

$$Loss = - \sum_{j=1}^n y_j \log(\hat{y}_j) \quad (1)$$

where n represents the cardinality of the classes, y_j denotes the ground-truth label, and $\hat{y}_j$ is the predicted probability for class j. The training process was conducted in two phases. In the first phase, all backbone networks were frozen to train only the classification head. In the second phase, selective fine-tuning was performed by unfreezing the top layers of each backbone to further enhance performance.

E. Performance Measures

The performance of the proposed hybrid CNN ensemble model and comparison models was assessed using standard classification measures, including overall accuracy, precision, recall, F1-score, and class-wise accuracy.

IV. RESULTS AND DISCUSSION

A. Analysis of Confusion Matrices

The confusion matrices in Fig. 2 display the class-wise performance of the evaluated models for the apple leaf disease categories. The hybrid CNN ensemble model achieves the most prominent diagonal

dominance among all matrices. Combining EfficientNet-B0, DenseNet121, and ResNet50 uses complimentary feature representations, as evidenced by the considerable reduction in off-diagonal values. Particularly for visually comparable categories, the ensemble efficiently reduces class overlap and enhances resilience. It exhibits superior class discrimination and the lowest misclassification rate, confirming the efficacy of heterogeneous CNN ensemble learning, while the majority of errors across all models occur between classes with similar visual features. The highly concentrated diagonal values of EfficientNetB0 indicate strong correct classification across all classes.

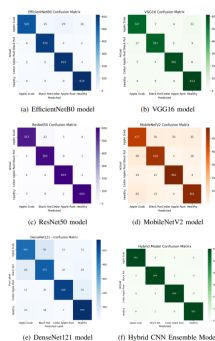


Fig. 2. Confusion matrices for the five models: (a) EfficientNetB0 (b) VGG16 (c) ResNet50 (d) MobileNetV2 (e) DenseNet121 (f) Hybrid CNN Ensemble Model (EfficientNetB0 + DenseNet121 + ResNet50)

Misclassifications are rare and mostly happen across fruit groups that are visually similar, indicating that the model successfully extracts discriminative properties. While VGG16 performs well, its off-diagonal values are somewhat greater than those of EfficientNet-B0. The limitations of deeper yet parameter-heavy classic CNN designs are reflected in some inter-class confusion, especially across classes with similar colour and texture patterns. Due to residual learning, ResNet50 shows better class separation. Misclassification rates are lower than VGG16, and the diagonal dominance is evident. By lowering vanishing gradient effects and enhancing classification consistency, the skip connections aid in maintaining significant features. In comparison, MobileNetV2 displays a greater number of off-diagonal entries. Its lightweight architecture sacrifices some representational capacity despite being computationally efficient, which increases confusion among closely related classes. DenseNet121 achieves superior

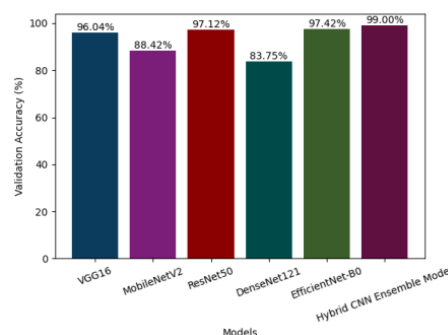


Fig. 3. Comparison of Validation Accuracy of six models

performance than VGG16 and MobileNetV2, with substantial diagonal concentration. Dense connections enhance gradient flow and feature reuse, increasing classification accuracy.

B. Analysis of Accuracy, Precision, Recall and F1-score

The hybrid CNN ensemble model consistently shows high performance across all apple leaf disease classes. The performance of the suggested hybrid CNN ensemble model was compared to a number of baseline architectures, including EfficientNet-B0, ResNet50, VGG16, MobileNetV2 and DenseNet121, in order to assess its efficacy, which is demonstrated in Fig. 3. For diseases such as AS, BR, and CAR, the model achieves near-perfect results against the earlier mentioned performance metrics, indicating minimal misclassification among visually similar classes. Healthy apple leaves are also accurately distinguished, further confirming the model's robustness. The class-wise evaluation highlights the effectiveness of combining features from multiple backbone networks, allowing the proposed model to capture subtle differences in color patterns, texture and lesion shapes specific to each apple leaf disease. These results underscore the suitability of the proposed framework for reliable disease diagnosis in real-world apple orchards. Table I summarizes the class-wise precision, recall, F1-score, and accuracy achieved by the EfficientNetB0, MobileNetV2, VGG16, DenseNet121, ResNet50 and the proposed hybrid CNN ensemble model. It clearly establishes that the proposed approach achieves the highest validation accuracy (99%) and consistently superior classwise precision, recall, and F1-scores for apple leaf disease classification compared to all baseline models.

C. Analysis of Accuracy and Loss Curves

The training and validation accuracy and loss curves as in Fig. 4 provide a comparison of the hybrid CNN against individual baseline models. In our study, the hybrid CNN ensemble model exhibits faster convergence, higher final accuracy, and lower validation loss compared to single back-

TABLE II — PERFORMANCE METRICS

Model	Class	Precision	Recall	F1-score	Accuracy
-------	-------	-----------	--------	----------	----------

Model	Class	Precision	Recall	F1-score	Accuracy
EfficientNetB0	AS	0.97	0.95	0.96	0.95
	BR	0.97	0.99	0.98	0.99
	CAR	0.97	0.98	0.98	0.98
	Healthy	0.98	0.98	0.98	0.98
VGG16	AS	0.94	0.93	0.95	0.93
	BR	0.94	0.97	0.95	0.97
	CAR	0.97	0.97	0.97	0.93
	Healthy	0.96	0.97	0.96	0.97
MobileNetV2	AS	0.81	0.8	0.81	0.8
	BR	0.85	0.89	0.87	0.89
	CAR	0.94	0.93	0.94	0.93
	Healthy	0.9	0.9	0.9	0.9
ResNet50	AS	0.95	0.96	0.95	0.96
	BR	0.98	0.97	0.97	0.97
	CAR	0.95	1	0.97	1
	Healthy	0.95	0.97	0.95	0.97
DenseNet121	AS	0.76	0.81	0.79	0.81
	BR	0.81	0.74	0.75	0.74
	CAR	0.92	0.88	0.9	0.88
	Healthy	0.96	0.91	0.93	0.91
Proposed Hybrid CNN ensemble model	BR	0.99	0.99	0.99	0.99
	CAR	1	0.99	1	0.99
	Healthy	0.99	0.99	0.99	0.99

bone networks such as EfficientNet-B0, ResNet50, VGG16, MobileNetV2, and DenseNet121. The curves indicate stable training behavior without overfitting, reflecting the model's ability to generalize across all disease classes, including visually similar ones like AS and CAR. In contrast, some individual models show fluctuations in validation accuracy and slower convergence, highlighting the benefits of feature-level fusion in capturing complementary patterns from multiple backbone networks. Overall, the comparative curves reinforce the superior performance and robustness of the proposed hybrid CNN for reliable apple leaf disease diagnosis.

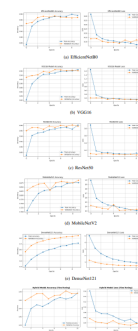


Fig. 4. Training and validation accuracy and loss curves of the six models: (a) EfficientNetB0 (b) VGG16 (c) ResNet50 (d) MobileNetV2 (f) DenseNet121 (g) Hybrid CNN Ensemble Model (EfficientNetB0 + DenseNet121 + ResNet50)

V. CONCLUSION

The paper highlighted a hybrid CNN ensemble model specifically for apple leaf disease classification by integrating EfficientNet-B0, DenseNet121, and ResNet50 using featurelevel fusion. The proposed hybrid CNN model effectively combined complementary feature representations from multiple backbone networks, resulting in enhanced robustness and improved classification performance for apple leaf disease categories such as AS, BR, and CAR. Experimental results on the apple leaf dataset demonstrated that the hybrid CNN ensemble model achieved an overall validation accuracy of approximately 99%, outperforming individual CNN architectures including EfficientNet-B0, VGG16, MobileNetV2, ResNet50, DenseNet121, and a custom CNN. The model consistently delivered high precision, recall, and F1-scores across all apple leaf disease classes, with negligible misclassification among visually similar diseases. The strong generalization capability and stable training behavior of the proposed approach indicate its suitability for practical apple orchard applications and realworld agricultural disease management. Despite the increased architectural complexity, the hybrid CNN ensemble model maintained computational efficiency due to the use of pretrained lightweight and residual backbones. Future work will focus on extending the framework to larger and more diverse real-field apple leaf datasets, incorporating multispectral and hyperspectral imagery, and optimizing the hybrid architecture for real-time deployment on edge devices and mobile platforms for rapid apple leaf disease diagnosis.

VI. REFERENCES

- [1] Saleem, M. H., Potgieter, J., & Arif, K. M. (2019). Plant disease detection and classification by deep learning. *Plants*, 8(11), 468.
- [2] Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and electronics in agriculture*, 145, 311-318.

- [3] Sankaran, S., Mishra, A., Ehsani, R., & Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Computers and electronics in agriculture*, 72(1), 1-13.
- [4] Mohanty, S. P., Hughes, D. P., & Salathe, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in plant science*, 7, 1419.
- [5] Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational intelligence and neuroscience*, 2016(1), 3289801.
- [6] Barbedo, J. G. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosystems engineering*, 172, 84-91.
- [7] Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture*, 161, 272-279.
- [8] Boulent, J., Foucher, S., Theau, J., & St-Charles, P. L. (2019). Convolutional neural networks for the automatic identification of plant diseases. *Frontiers in plant science*, 10, 941.
- [9] Lu, J., Hu, J., Zhao, G., Mei, F., & Zhang, C. (2017). An in-field automatic wheat disease diagnosis system. *Computers and electronics in agriculture*, 142, 369-379.
- [10] Kamilaris, A., & Prenafeta-Boldu, F. X. (2018). Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, 147, 70-90.
- [11] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- [12] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.
- [13] 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).
- [14] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [15] Barbedo, J. G. A. (2019). Plant disease identification from individual lesions and spots using deep learning. *Biosystems engineering*, 180, 96- 107.
- [16] Castillo-Martínez, M. A., Gallegos-Funes, F. J., Carvajal-Gómez, B. E., Urriolagoitia-Sosa, G., & Rosales-Silva, A. J. (2020). Color index based thresholding method for background and foreground segmentation of plant images. *Computers and Electronics in Agriculture*, 178, 105783.
- [17] Singh, V. K., Singh, B. P., Kisi, O., & Kushwaha, D. P. (2018). Spatial and multi-depth temporal soil temperature assessment by assimilating satellite imagery, artificial intelligence and regression based models in arid area. *Computers and Electronics in Agriculture*, 150, 205-219.
- [18] Nadafa, R. A., Hatturea, S. M., Bonala, V. M., & Naikb, S. P. (2020). Home security against human intrusion using Raspberry Pi. *Procedia Computer Science*, 167, 1811-1820.
- [19] Zhang, S., Wu, X., You, Z., & Zhang, L. (2017). Leaf image based cucumber disease recognition using sparse representation classification. *Computers and electronics in agriculture*, 134, 135-141.
- [20] Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., & Hughes, D. P. (2017). Deep learning for image-based cassava disease detection. *Frontiers in plant science*, 8, 1852.
- [21] Ji, W., Zhang, J., Xu, B., Tang, C., & Zhao, D. (2021). Grasping mode analysis and adaptive impedance control for apple harvesting robotic grippers. *Computers and Electronics in Agriculture*, 186, 106210.
- [22] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Hounsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- [23] Wang, G., Sun, Y., & Wang, J. (2017). Automatic image-based plant disease severity estimation using deep learning. *Computational intelligence and neuroscience*, 2017(1), 2917536.
- [24] Li, Y., Nie, J., & Chao, X. (2020). Do we really need deep CNN for plant diseases identification?. *Computers and Electronics in Agriculture*, 178, 105803.
- [25] Jiang, P., Chen, Y., Liu, B., He, D., & Liang, C. (2019). Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. *Ieee Access*, 7, 59069-59080.
- [26] Liu, B., Zhang, Y., He, D., & Li, Y. (2017). Identification of apple leaf diseases based on deep convolutional neural networks. *Symmetry*, 10(1), 11.
- [27] Wang, G., Sun, Y., & Wang, J. (2017). Automatic image-based plant disease severity estimation using deep learning. *Computational intelligence and neuroscience*, 2017(1), 2917536.
- [28] rm1000, "Augmented Apple Disease Detection Dataset," Kaggle. [Online]. Available: <https://www.kaggle.com/datasets/rm1000/augmented-apple-disease-detection-dataset>. [Accessed: Jan. 11, 2026]