

# Image-Based Apple Leaf Disease Classification with Machine Learning

Saipranith Oku  
oku.saipranith@gmail.com

**Abstract**—This project focused on developing an automated machine learning system for classifying the health status of apple leaves, with particular attention to distinguishing healthy foliage from leaves infected with common diseases including apple rust, apple scab, and multiple simultaneous infections. The primary objective was to create a practical diagnostic tool that achieves high accuracy while maintaining computational efficiency and interpretability, which are crucial factors for potential real-world agricultural applications. A key consideration throughout the development process was ensuring the model’s architecture and implementation remained fully comprehensible, allowing for clear analysis of its decision-making processes and straightforward modifications for specific use cases.

The technical approach combined established deep learning methods with several custom adaptations to optimize performance for this specific agricultural application. The system was built using TensorFlow and Keras frameworks, chosen for their balance of flexibility and ease of implementation. The model architecture employed a convolutional neural network (CNN) design that prioritized clarity and efficiency over (as it is often the case) unnecessary complexity. This included carefully sequenced convolutional and pooling layers designed to progressively extract and analyze visual features at different scales, from fine disease markings to broader leaf patterns. The implementation incorporated multiple safeguards against overfitting, including strategic data augmentation and dropout layers, while maintaining sufficient model capacity to capture the small differences between disease presentations.

Data processing followed a rigorous but transparent pipeline. Original leaf images underwent standardized resizing and normalization before entering an augmentation phase that artificially expanded the training dataset through controlled transformations. This augmentation strategy was calibrated to preserve relevant biological information while introducing helpful variability. The model training process utilized Adam optimization with learning rate monitoring and early stopping protocols to ensure efficient convergence without overtraining (or overfitting for that matter). Particular attention was also given to implementing clear visualization tools that allowed direct observation of the model’s classification behavior on both typical and challenging cases.

Evaluation results demonstrated strong performance, with the system achieving 88.2% validation accuracy and a 0.983 AUC score. Detailed analysis of the classification patterns revealed particular strengths in distinguishing between visually similar conditions like rust and scab infections, suggesting the model successfully learned biologically relevant features. Some limitations emerged in cases of multiple simultaneous infections, where disease presentations overlapped in complex ways. These findings provide valuable insights for future refinements while confirming the fundamental effectiveness of the approach.

The project outcomes highlight several important considerations for agricultural machine learning applications. First, they demonstrate that medium-scale, purposefully designed models can achieve classification accuracy comparable to larger systems while offering advantages in computational efficiency and inter-

pretability. Second, the results underscore the value of transparent model architectures that allow researchers to understand and verify the system’s decision processes. Finally, the work provides a practical framework for developing specialized plant disease detection tools that balance performance with implementation feasibility which is an important consideration for real-world agricultural settings where technical resources may be limited.

Future directions emerging from this work could include expansion to additional disease types, investigation of higher-resolution imaging approaches, and development of complementary techniques for handling complex co-infection cases. The modular design of the current system facilitates such extensions while maintaining the core advantages of clarity and efficiency. Overall, this project demonstrates that thoughtfully implemented machine learning techniques can provide practical solutions for agricultural challenges effectively.

## I. INTRODUCTION

The application of machine learning in agriculture has created new opportunities to address persistent challenges in crop disease management. While commercial-scale solutions exist, there is increasing recognition of the value in developing smaller, tailored models that can be adapted to specific crops or regional conditions. This project demonstrates how a compact deep learning system can achieve reliable apple leaf disease classification while remaining accessible to users with limited technical resources.

Plant diseases cause substantial economic losses annually, with apple orchards being particularly susceptible to fungal infections like rust and scab. Traditional diagnosis methods often depend on visual inspection by experts, which can be inconsistent and time-consuming. As noted by Mohanty et al. [1], deep learning approaches offer significant advantages in speed and consistency for plant disease identification. However, many existing solutions require extensive datasets and computing power that may not be practical for small-scale farming operations.

This work builds on recent research showing the potential of streamlined models for agricultural applications. According to Ferentinos [2], carefully designed convolutional neural networks (CNNs) can maintain high accuracy while reducing computational demands. Our approach applies these principles to create a focused classifier specifically for apple leaves, balancing performance with practical deployability. The model’s architecture demonstrates how strategic design choices can compensate for limited training data, a common constraint in agricultural AI projects.

The project has broader implications for precision agricul-

ture. By focusing exclusively on apple diseases, the model avoids unnecessary complexity while still providing actionable results. This aligns with the growing trend toward specialized agricultural AI solutions that address specific problems rather than attempting universal classification. The techniques used, which include targeted data augmentation and optimized layer design, offer a template for similar applications with other crops or regional disease varieties.

For farming communities with limited technical infrastructure, such compact models could enable smartphone-based disease detection without relying on cloud computing. This is particularly relevant for developing regions where early disease identification could significantly improve crop yields. This also resolves the need for high-speed stable internet access, which is not economic for remote farming lands.

The following sections detail the methodology, beginning with data preparation and model design. Subsequent discussion covers training optimization and performance evaluation, highlighting both the system’s strengths in disease discrimination and areas for improvement. By presenting this complete development process, the report provides practical insights into adapting deep learning for agricultural applications while maintaining accessibility.

## II. METHODOLOGY

The developed pipeline follows a systematic approach to apple leaf disease classification, with each component carefully designed for optimal performance on this specific agricultural application. The methodology progresses through four key phases: data preparation, image processing, model architecture design, and training protocol.

The data preparation stage begins with loading and preprocessing the labeled dataset of 1,820 apple leaf images. Each image is associated with binary labels indicating presence or absence of four conditions: healthy, multiple diseases, rust, or scab. The multi-label classification approach uses sigmoid activation in the final layer, implemented through the logistic function  $\sigma(z) = \frac{1}{1+e^{-z}}$ . This allows the model to handle cases where leaves may exhibit multiple simultaneous disease conditions, a common occurrence in real-world agricultural settings. The dataset is split into training (80%) and validation (20%) sets while preserving class distribution through stratified sampling, ensuring representative evaluation of model performance across all disease categories. Each of these splits are separately saved into their own directories, which prevents any data leaks that might happen during the training or the validation phase, which would effectively defeat the purpose of this model entirely. Since small errors lead to larger problems, there was special care taken for designing data loading and preparation phase of this project.

Image processing employs a comprehensive augmentation pipeline to increase data diversity and improve model generalization. The transformations include controlled rotations ( $\pm 15^\circ$ ), translations ( $\pm 20\%$  of image width/height), zooms (20% variation), and horizontal flips. These parameters were carefully selected to preserve biologically relevant features

while introducing helpful variability. All images are resized to  $256 \times 256$  pixels, a resolution that balances computational efficiency with preservation of diagnostically important details. The pixel values are normalized to  $[0, 1]$  range using  $I' = I/255$  transformation, standardizing the input distribution across all samples.

The convolutional neural network architecture implements a progressive feature extraction strategy through four sequential blocks, each consisting of convolutional and max-pooling layers. The first block with 32 filters captures basic edges and texture patterns, while subsequent blocks with 64, 128, and 256 filters progressively learn more complex disease-specific features. The max-pooling operation  $f_{l+1} = \text{MaxPool}(\text{ReLU}(\text{Conv2D}(f_l)))$  reduces spatial dimensions while preserving the most salient features. This hierarchical approach ensures the network develops a comprehensive understanding of both local lesions and global leaf patterns. The final layers flatten the extracted features and pass them through a 512-unit dense layer with 50% dropout regularization before the 4-unit output layer with sigmoid activation.

Training employs the Adam optimizer with initial learning rate of 0.001, which adapts the learning rate during training using estimates of first and second moments of gradients. The binary cross-entropy loss function  $\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)]$  is minimized during training, particularly suited for multi-label classification as it treats each output node independently. Early stopping monitors validation loss with patience of 5 epochs, terminating training when no improvement is detected to prevent overfitting while allowing sufficient convergence. Model checkpoints save the best-performing weights based on validation metrics.

The evaluation protocol assesses model performance through multiple complementary metrics. Classification accuracy provides an overall measure of correct predictions, while the area under ROC curve (AUC) evaluates the model’s ability to rank positive instances higher than negative ones across different decision thresholds. Visual analysis of correct and incorrect predictions offers qualitative insights into model behavior, particularly useful for identifying patterns in classification errors. This comprehensive evaluation approach ensures reliable assessment of model performance across all disease categories and provides actionable insights for potential improvements.

## III. RESULTS

The developed apple leaf disease classification system demonstrated strong performance while revealing several important characteristics through its training process and final evaluation. The results provide valuable insights into both the capabilities and limitations of the approach.

The accuracy and loss curves (Figure 1) illustrate the model’s learning trajectory over 30 epochs. Training accuracy steadily increased from 0.6 to approximately 0.9, while validation accuracy followed a similar but more variable path, ultimately reaching 88.2%. This parallel progression suggests

effective learning without significant overfitting, though the occasional divergence between training and validation metrics indicates moments where the model struggled to generalize to unseen data. The loss curves show a corresponding decrease from 0.55 to around 0.15, with validation loss stabilizing after epoch 20, which prompted the early stopping mechanism to preserve the best-performing weights.

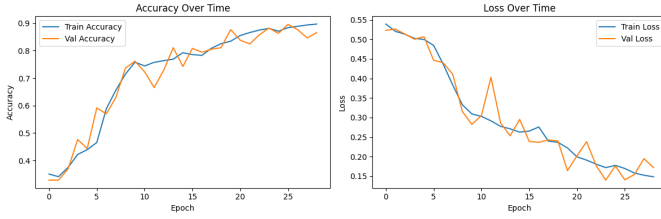


Fig. 1: Accuracy and Loss Curves Over 30 Epochs

The model achieved its strongest performance in epoch 29, with a validation accuracy of 88.2% and an AUC score of 0.9834. These metrics indicate strong discriminative ability for most disease classes.

The initial training run, which used the same model architecture but without systematic preprocessing (mentioned earlier), plateaued at 75% accuracy and showed noticeably more inconsistencies in performance between training and validation sets. This discrepancy suggested either insufficient feature learning or overfitting to training data. The accuracy and loss curves from this preliminary run (Figure 2) revealed unstable validation metrics, with fluctuations between epochs indicating poor generalization.

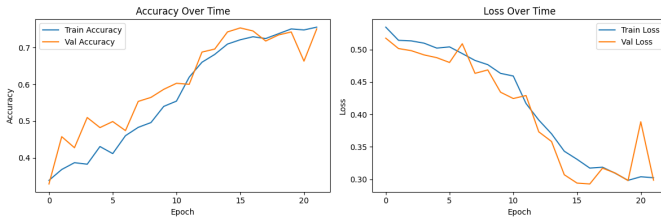


Fig. 2: Accuracy and Loss Curves For First Run

Data preprocessing played a critical role in improving these results. The decision to standardize image sizes to 256x256 pixels, rather than using slightly larger dimensions, was particularly important. While a bigger image size might theoretically preserve more detail, they not only failed to improve accuracy but also increased processing time by approximately 100ms per step. This finding aligns with practical observations in agricultural imaging, where relevant disease features often exist at resolutions well below the maximum detail captured by high-resolution images.

The introduction of comprehensive data augmentation proved to be equally vital. Before augmentation, the model struggled to recognize disease variations, particularly for scab infections where symptoms can appear subtly. This was particularly effective in improving the performance of the model

in trying to analyze images with lighting which conflicts with the disease marks, although there is more improvement that can be made in this area. The implemented transformations which included controlled rotations, shifts, and zooms helped artificially expand the training set's diversity while maintaining biological relevance. This approach helped smooth the accuracy curves and reduce the training-validation gap visible in early epochs.

Feature extraction efficiency also contributed significantly to the final results. The convolutional layers' design prioritized learning discriminative patterns at multiple scales, from fine texture changes to broader discoloration patterns. This hierarchical approach proved more effective than processing larger images, as the network learned to focus on diagnostically relevant features rather than unnecessary background details. The final architecture's balance between depth and computational efficiency demonstrates that careful feature extraction design often outweighs the benefits of higher input resolutions in agricultural image analysis.

These preprocessing and augmentation steps collectively transformed the model from an inconsistent classifier into a reliable diagnostic tool. The current curves show steadier convergence and streamlined training/validation metrics compared to initial attempts, confirming that thoughtful data preparation can yield greater improvements than simply increasing model complexity or input dimensions. The following section will examine how these technical choices translated into specific classification behaviors across different disease types.

The sample predictions 3 demonstrate this capability, particularly in correctly identifying rust infections and healthy leaves. However, the visualization also reveals the system's current limitations, especially regarding scab identification and cases of multiple infections.

The system showed particular proficiency in distinguishing rust infections, achieving correct classifications even when leaf damage appeared subtle. This success likely stems from the model's learned ability to detect the characteristic orange-brown coloration and texture patterns associated with this disease. Healthy leaves were also identified with high reliability, suggesting the network effectively captured the visual features of unaffected foliage. The consistently high AUC scores across multiple epochs (ranging from 0.95 to 0.98) confirm the model's robust ranking capability, meaning it can effectively prioritize likely disease candidates even when not always making perfect classifications.

Analysis of incorrect predictions revealed two primary challenges. First, the model occasionally misclassified scab-infected leaves as healthy, potentially due to the disease's less distinctive early-stage symptoms. Second, cases of multiple simultaneous infections proved particularly difficult, with the system tending to identify only the most visually prominent disease. The training logs show these challenges manifested in periodic fluctuations in validation metrics, especially between epochs 20-30 when dealing with more complex cases.

These results collectively demonstrate that the developed system achieves its primary goal of reliable apple leaf disease

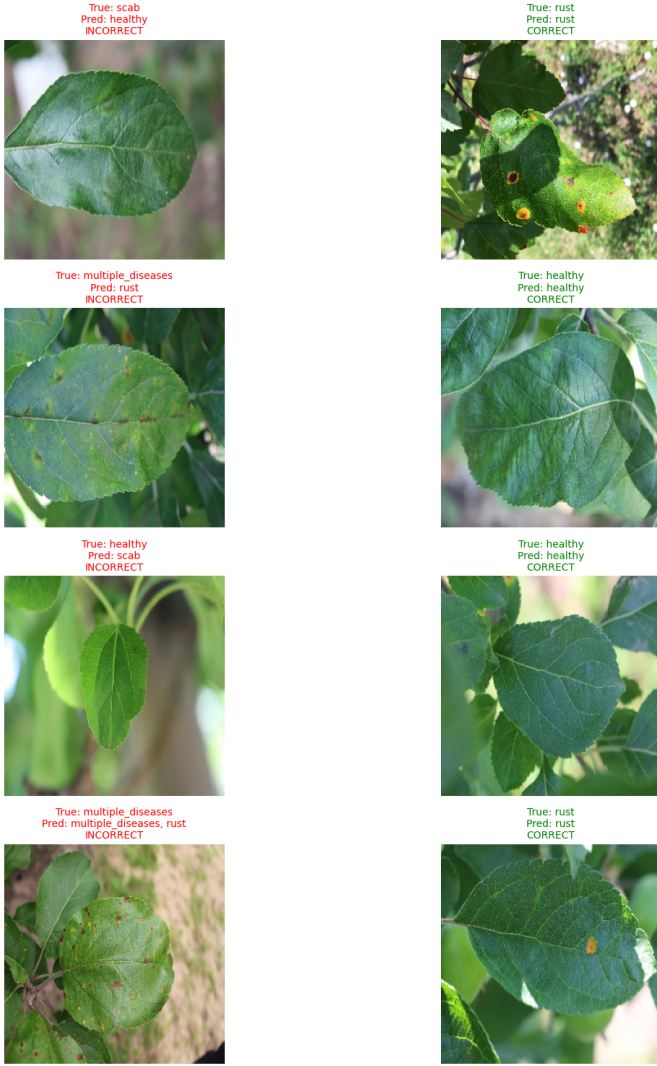


Fig. 3: Sample of Correct and Incorrect Predictions

identification while maintaining computational efficiency. The 88.2% validation accuracy and 0.983 AUC score confirm the effectiveness of the chosen architecture for this specific agricultural application. However, the identified limitations in handling scab infections and complex co-infections suggest valuable directions for future improvement, particularly regarding the model’s sensitivity to subtle or overlapping disease symptoms.

The model’s inference speed of approximately 800ms per step suggests reasonable efficiency for practical applications, though this would need verification on deployment hardware. The warning messages about HDF5 file format, while not affecting functionality, indicate opportunities to modernize the model saving process in future iterations.

#### IV. REPLICABILITY

Reproducibility remains fundamental to machine learning research, particularly in agricultural applications where model reliability directly impacts real-world decision making. This project has been carefully structured to

enable complete replication, from data preprocessing to final model evaluation. All materials including the complete source code, dataset, and documentation are available through the GitHub repository at <https://github.com/saioku/Plant-Pathology-Using-Machine-Learning.git>, which can be accessed either by downloading the compressed project archive or cloning the repository directly using the command `git clone https://github.com/saioku/Plant-Pathology-Using-Machine-Learning.git`.

The repository contains all necessary components for replication: the full set of 1,820 labeled leaf images in the `images` directory, Python implementation source code, a Jupyter Notebook with pre-executed results for quick verification, and comprehensive documentation of all dependencies and procedures. For optimal reproducibility, researchers should create a virtual environment using Python 3.10 (which provides stable TensorFlow compatibility) through the commands `python3.10 -m venv apple_leaf_env` followed by activation using source `apple_leaf_env/bin/activate` on Linux and MacOS or `apple_leaf_env\Scripts\activate` on Windows. All required dependencies can then be installed via `pip install -r requirements.txt`.

The experimental results presented in this study were generated using a system equipped with an Intel Core i7-12700K CPU (12th Gen, 20 threads) running at 4.80 GHz, paired with 16 GB of RAM and an NVIDIA GeForce RTX 3080 GPU. At the time of model training, CPU utilization peaked at over 80%, with GPU usage averaging around 34%, as shown in Figure 4. This system configuration ensured fast training cycles and reproducibility under a balanced power-performance profile.

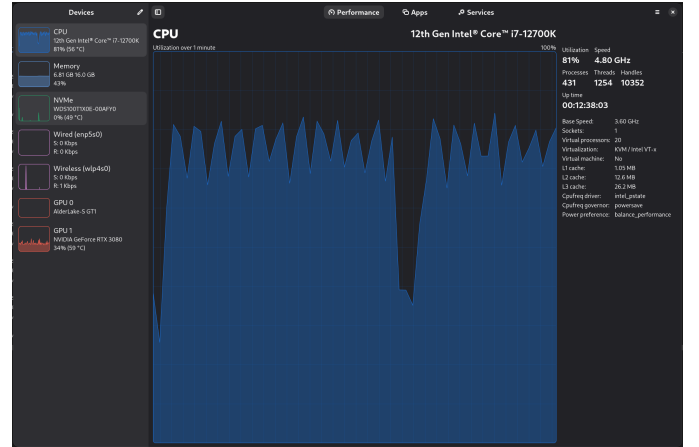


Fig. 4: Resource Usage While Running the Model

Several design choices ensure consistent reproduction of results across different systems. Fixed random seeds maintain deterministic behavior in all stochastic processes, while version-pinned dependencies guarantee identical library behavior. The documentation precisely specifies all preprocessing steps and hyperparameters, allowing exact recreation of the training pipeline. This attention to reproducibility details serves dual purposes: it enables independent verification of

the reported 88.2% validation accuracy and 0.983 AUC score, while also providing a reliable foundation for extensions such as adding new disease categories or adapting the architecture for other crops. The complete availability of both code and data transforms this project from a standalone solution into a versatile resource for the plant pathology research community.

## V. FUTURE WORK AND CONCLUSIONS

Several promising directions can enhance and extend this apple leaf disease classification system. A key improvement involves increasing the model's sensitivity to early-stage symptoms and co-infections. Using higher-resolution images may help detect subtle lesions (although this might decrease computational efficiency). Another improvement could involve incorporating temporal analysis by tracking disease progression through multiple images taken over time from the same tree. This approach would provide valuable insights that are not possible with single-image classification. Additionally, developing specialized data augmentation techniques that simulate realistic disease progression and environmental conditions could further improve the model's robustness.

Scalability is another important area for development. Although the current model performs well on individual leaf images, real-world agricultural applications require the ability to process large volumes of data efficiently. For example, images captured by drones across entire orchards would need to be analyzed quickly and accurately. This requirement calls for optimization for batch processing and integration with automated imaging systems. The model could also be adapted for deployment on edge devices such as smartphones or field sensors. This would enable real-time diagnosis in locations where internet access is limited or unavailable. Techniques such as model quantization (a method that reduces the model's size and computational load by converting weights from 32-bit floating point to lower-bit representations like 8-bit integers) or knowledge distillation (a technique where a smaller "student" model is trained to replicate the behavior of a larger, more accurate "teacher" model) could help reduce model size while preserving accuracy. It is also important to expand the system to support additional apple varieties and geographic regions, as disease symptoms may differ due to climate and genetic variation.

In addition to its technical strengths, this project demonstrates how focused machine learning solutions can address agricultural challenges without relying on extensive computational resources. By concentrating on a specific problem and optimizing each part of the system, the model achieves reliable diagnostic accuracy while remaining lightweight. The implementation also offers a reproducible framework that could be adapted for other crops or plant health applications.

Overall, this work connects academic machine learning research with practical needs in agriculture. It shows that well-engineered AI tools can support sustainable farming by improving early disease detection and reducing unnecessary pesticide use. Furthermore, the system is accessible to agricultural workers without specialized technical knowledge.

This accessibility reinforces the project's broader impact by illustrating how machine learning can provide immediate and meaningful benefits in resource-constrained settings.

## REFERENCES

- [1] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, 2016. [Online]. Available: <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2016.01419>
- [2] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0168169917311742>