International Conference on Engineering, Applied Science And Technology

Application of Deep Learning with Image Preprocessing for Strawberry Leaf Disease Classification

Muhamad Havidz Alkausar[1*]
[1]*Universitas Nusa Putra, Sukabhumi, Jawa Barat 43152, Indonesia*

ARTICLE INFO

Keywords:
Deep Learning
Convolutional Neural Networks (CNN)
Transfer Learning
Image Preprocessing
Plant Disease Classification

ABSTRACT

This project developed a deep learning-based strawberry leaf disease classification system using Convolutional Neural Networks (CNN). The goal was to overcome the limitations of manual methods, namely time-consuming and error-prone visual inspection. By applying transfer learning techniques and image preprocessing (such as resizing and normalization), the model was trained to classify leaf images into disease categories such as Leaf Spot, Powdery Mildew, or healthy leaves. Model evaluation demonstrated an overall accuracy of 63% on the test data. While accuracy varied by class,some classes demonstrated strong performance (e.g.,Blossom Blight with 90% accuracy and 100% recall). Confusion matrices were also used to analyze specific classification errors. The main challenges faced were potential overfitting and limited data. However, this project demonstrated the viability of AI approaches for precision agriculture, paving the way for the development of better models in the future.

Introduction

Strawberry leaf diseases pose a serious threat to farmers, reducing the quality and quantity of crops and resulting in significant economic losses. Rapid and accurate disease identification is essential for implementing appropriate management strategies. This project aims to develop an automated system capable of classifying strawberry leaf diseases based on images, thereby assisting farmers and agricultural experts in monitoring plant health and taking early preventive or therapeutic measures.

This project focuses on the limitations and inefficiencies of traditional methods for detecting strawberry leaf diseases. Current methods, visual inspection by farmers or experts, are typically time-consuming, prone to subjective errors, and require specialized expertise. This makes disease detection ineffective and less timely [1] . This problem is particularly pressing because accurate and rapid disease identification is key to efficient disease control and preventing crop losses. By developing an automated detection system, farmers can conduct more extensive field monitoring and detect diseases early, before they spread to other parts of the crop and cause significant damage. Artificial intelligence-based approaches offer significant advantages in terms of speed, efficiency, and accessibility over conventional methods. This approach can help maintain food security [2] , particularly in strawberry production, reduce the negative economic impacts on the agricultural sector, and increase overall productivity [3] .

Project Relationship with Computer Vision Concepts This project deeply applies various core concepts from the computer vision course , including:

- Image Classification: The main objective of this project is to classify strawberry leaf images into disease categories such as Powdery Mildew, Leaf Spot, Angular Leaf Spot, or healthy leaves. This is one of the basic applications in the field of computer vision. [4]
- Convolutional Neural Networks (CNN): This project will most likely use CNN as the core framework for processing and analyzing visual data in the form of strawberry leaf images, because CNN is specifically designed to recognize patterns and features in images. [5]
- Image Pre-processing : Based on the code you provided, the image pre-processing step is crucial. This process includes several techniques to prepare the data before it goes into the model, such as rescaling (resizing the image to the same dimensions), normalization (setting pixel values

* Corresponding author.
E-mail address: muhamad.havidz_ti18@nusaputra.ac.id

to a specific range), and possibly data augmentation to enrich the dataset. The code you uploaded demonstrates this pre-processing step. [6]

- Transfer Learning (Possibly): If the strawberry data used is limited, it is likely that transfer learning techniques will be applied. This technique uses a model that has been trained on a large image dataset (such as ImageNet) as a base, then re-adapted to the task of classifying strawberry leaf diseases. [7]
- Fine-Tuning (Possibly): As a continuation of transfer learning, fine-tuning is the process of retraining the top layers of the model to make it more suitable and specific for the strawberry leaf disease dataset. [8]
- Regularization (Dropout) To prevent overfitting , technique regularization like Dropout will likely be implemented. This technique works by randomly turning off some neurons during the training process , so that the model is forced to learn to recognize more important and powerful features [9]
- Evaluation Metrics: Model performance will be assessed using standard metrics in computer classification tasks. vision , such as accuracy , precision, recall, and F1 score.
- In addition, the Confusion Matrix will also used For analyze performance detailed model for each class. [10]

System Diagram and Design

System flow for leaf disease classification project based computer Vision usually follows the following pattern, which is also seen from the presence of stages pre - processing.

Research flowchart:

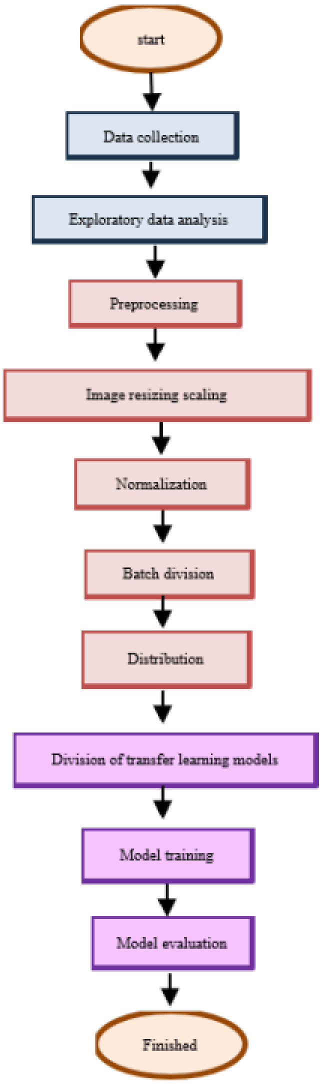

Figure 1. Research flowchart for strawberry leaf disease classification using deep learning

- Data collection: Collecting strawberry leaf image dataset from secondary sources like Kaggle. This dataset contains image of healthy strawberry leaves and those infected with various types of diseases .
- Exploratory Data Analysis (EDA): Recognition process data For understand characteristic features dataset , such as class distribution and existing image variations
- Pre-Processing: This is the stage where your code will come into play. It includes:
- Size Adjustment Image (Scaling/Resizing Image): Ensure all images have the same size , for example 224x224 pixels, so that the model can receive input consistently.
  - Normalization : Sets the intensity value image pixels to fit a certain scale range , such as 0 to 1, thus facilitating the model learning process .

  - Data Division : Splitting the dataset into three parts, namely data for training, data for validation, and data for testing.
  - Batch Division: Grouping data in size certain , such as 32, so that the training process can occur more efficiently.
- Building - Transfer Learning : Building stage structure models , usually using a pre-trained CNN architecture , such as Efficient Net Res Net VGG, or others, as a base and utilize the weights that have been trained
- Model Training: The process of teaching a model with training data over several epochs so that the model can learn patterns to classify diseases.
- Model Evaluation: Evaluate the performance of a trained model using various metrics on previously unseen test data.

The main technologies, frameworks, tools used in project

- Technology/Model (Example):
  - Convolutional Neural Networks (CNN): Used as the backbone of model architecture due to its excellent ability to recognize visual patterns.
  - Transfer Learning: Most likely used to leverage knowledge from models that have been trained on larger data sets . This
  - is very help in overcoming the limited amount of training data for strawberry leaf diseases and enabling better models in generalization to smaller data.
- Frameworks/Tools (Examples):
  TensorFlow /Keras: Framework deep general learning used For building , training, an evaluating CNN models. Pre – code pre - processing what you provide is likely part of a workflow within this framework.
  - Python: The programming language used during project .
  - Open CV /PIL (Pillow): A library used for image processing , such as resizing images , normalization, and so on, in stage pre - processing
  - Numpy: A library used to perform numerical calculations, often used in image data manipulation
  - Optimizer (e.g. Adam): The algorithm used during model training to adjust model weights to improve model performance
  - Activation Function (e.g. ReLU, Softmax): Mathematical functions used in model neurons to add non - linearity (ReLU in hidden layers) and generate class probabilities (Softmax in final layers).

System Testing and Evaluation Testing Methods and Process:
The system is tested using data that is different from the data training . The main objective of this testing is to evaluate the model's capabilities in classifying new and previously unseen data, as well as verifying the extent to which the model can generalize its knowledge. The evaluation method used is based on on standard quantitative metrics for classification tasks, namely accuracy , precision, recall, F1 score , and confusion matrix (for visual and detailed evaluation of class performance).

Test Results:
During the phase training , the model achieved a validation accuracy of 90–95% with a low loss rate. When tested with test data, the model showed mixed performance, but still provides important insights .
The following are the classification results obtained:

	precision	recall	f1-score	support
0	0.75	0.80	0.78	147
1	0.52	0.42	0.46	36
2	0.90	1.00	0.95	62
3	0.53	0.94	0.68	145
4	0.59	0.38	0.46	159
5	0.17	0.07	0.10	43
6	0.67	0.49	0.56	151
accuracy			0.63	743
macro avg	0.59	0.59	0.57	743
weighted avg	0.62	0.63	0.61	743

Figure 2. Classification report of the strawberry leaf disease classification model

Class Label Mapping: Based on the dataset used, the numeric labels above have the following mapping:
0: Disease Name for Label 0 : Angular Leafspot
1: Disease Name for Label 1: Anthracnose Fruit Rot
2: Disease Name for Label 2: Blossom Blight
3: Disease Name for Label 3: Gray Mold
4: Disease Name for Label 4: Leaf Spot
5: Disease Name for Label 5: Powdery Mildew Fruit
6: Disease Name for Label 6: Powdery Mildew Leaf

Interpretation of Results:

1. Accuracy: Measures the percentage of correct predictions overall. The model's accuracy reached 63%, meaning 63% of the model's predictions were correct.
2. Accuracy: This is the ratio of correct positive predictions to total positive predictions for each class. High accuracy indicates a low false positive rate. It can be

seen that class 2 has the highest accuracy (0.90), while class 5 has the lowest accuracy (0.17).

3. Recall: Measures the percentage of positive data correctly predicted for each class. High recall indicates that the model can detect a large proportion of positive cases (and has a low false negative rate). Class 2 achieved perfect recall (1.00), while class 5 achieved only a very low recall (0.07).
4. F1-Score: It is the harmonic mean of precision and recall which is used to evaluate overall performance, especially when there is class imbalance.
5. Support: Shows the actual number of examples in each class in the test data.

Confusion Matrix:

A visual confusion matrix shows the number of correct and incorrect predictions for each class, helping to identify which classes may be confusing the model. The main diagonal of the matrix shows the number of correct predictions, while values outside the diagonal indicate misclassifications.

Confusion Matrix:

A visual confusion matrix displays the number of correct and incorrect predictions for each class, making it easier to identify classes that may be confusing the model. The main diagonal of the matrix contains the number of correct predictions, while values outside the diagonal indicate incorrect classifications

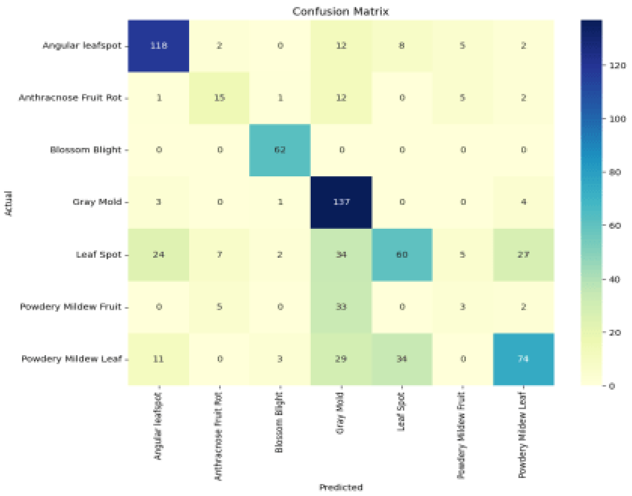

Figure 3. Confusion matrix of the proposed deep learning model for strawberry leaf disease classification

Interpretation of Confusion Matrix:

From the confusion matrix above, we can see how the model performed in predicting each class. The model successfully predicted 118 images as "Angular Leafspot" and the results were correct, so the number of correct predictions for this class is quite high. However, there were several misclassifications, such as 24 images that were actually "Angular Leafspot" but were predicted as "Leaf Spot", and 12 images of "Angular Leafspot" that were incorrectly categorized as "Gray Mold". The "Blossom Blight" class showed very good performance with 62 correct predictions and no significant misclassifications.

Sample Data Visualization

Below are some example images from the data used, complete with information about their shape, pixel value range, and index labels. These images provide an overview of the visual variation in the data and the possible forms of disease on strawberry

Figure 4. Sample dataset image labeled as Anthracnose Fruit Rot

Shape: (224, 224, 3)
Min pixel value: -1.168
Max pixel value: 3.499
Label indices: [2]

Angular leafspot

Typical characteristics of anthracnose in strawberries:

1. Black, rotting spots appear on the surface of the fruit.
2. The fruit becomes soft and watery in the infected area.
3. This disease can spread rapidly in humid and warm conditions.
4. It can also attack the stems and stolons of the plant.

Figure 5. Sample dataset image labeled as Blossom Blight

Shape: (224, 224, 3)
Min pixel value: -1.047
Max pixel value: 3.011
Label indices: [1]

Botrytis cinerea (Gray Mold)
General characteristics of this disease:

1. Occurs on flowers or young fruit.
2. Flowers or fruit turn brown, rot, and are often covered with gray mold (especially when damp).
3. Can spread rapidly in wet and humid conditions.
4. One of the flowers in the picture shows signs of necrosis (dead tissue) with dry edges, which is typical of Botrytis infection.

Figure 6. Fungal Rot on Strawberry Leaf
Image 2:

Shape: (224, 224, 3)
Min pixel value: -1.052
Max pixel value: 2,071
Label indices: [4]

Anthacnose fruit rot Characteristics:
Symptoms Description Small, dark purple or dark brown spots Circular or irregular in shape May have darker edges Sometimes the center of the spots turns pale or dry Attaches primarily older leaves But can spread to younger leaves If severe, leaves turn yellow, wilt, and fall off Inhibits photosynthesis and fruit growth

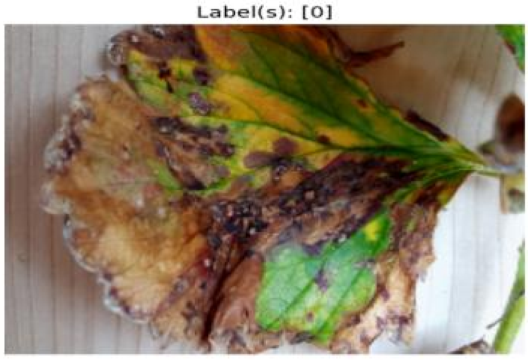

Figure 7. Diseased Strawberry Leaf with Rot
Image 10:

Shape: (224, 224, 3)
Min pixel value: -1.603
Max pixel value: 3.502
Label indices: [0]

Blossom Blight Cause:
Diplocarpon earlianum, Typical Characteristics of Leaf Scorch: Symptoms Explanation Small, dark purple spots that coalesce to form large lesions Looks like burns Leaves turn yellow and then dry out completely Plant becomes stressed Leaf edges curl and die Often affects older leaves first No white center of spots (unlike regular leaf spot) This distinguishes it from Mycosphaerella

Figure 8. Strawberry leaf with fungal symptoms

Image 9:
Shape: (224, 224, 3)
Min pixel value: -1.393
Max pixel value: 1.571
Label indices: [0]

Leaf spot Cause:
Peronospora spp. or Plasmopara spp. ⚠ General Characteristics of Downy Mildew on Strawberries: Symptoms Description Dark green or purple spots on the underside of the leaves Usually moist and spreading between the leaf veins The upper leaves may appear yellowed/stretched But the underside is

moldy The underside appears to have a fine dew-like coating Pathogen sporulation (water mold) Leaves may dry out and fall off If left untreated

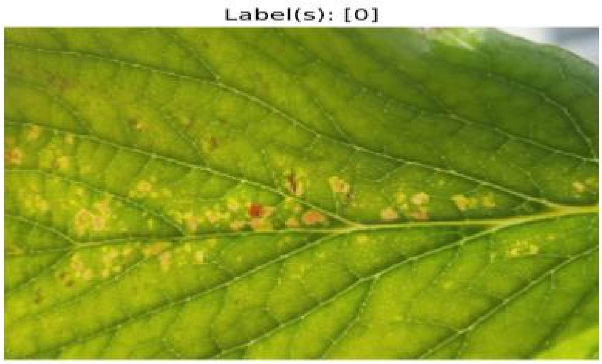

Figure 9. Yellow Strawberry Leaf

Powdery Mildew Fruit, Characteristics Visible in the Image: Symptoms Possible Causes Irregular yellow spots between leaf veins Symptoms of a virus or downy mildew No purple edges or circular spots Not fungal leaf spot Yellow color tends to spread from the main leaf vein Common in viral chlorosis or early downy mildew Leaf texture is intact, no holes Still in the early stages of infection

Figure 10. Strawberry Leaf Morphology

Shape: (224, 224, 3)
Min pixel value: -0.400
Max pixel value: 2.688
Label indices: [6]

Healthy Plants Characteristics of Healthy Leaves: Fresh, even green color No spots, chlorosis (yellowing), or necrosis (tissue damage) Leaves are upright and strong, not curled Leaf surfaces are smooth and free of mold or mildew Leaf edges are even and not drying out

Challenges and Limitations

Potential for Overfitting: One of the challenges identified during the training process is the risk of overfitting. This can be seen from the fact that the loss on the validation data begins to slow down and tends to stabilize after a few epochs, despite slight variations, while the loss on the training data continues to decrease significantly. This discrepancy between the training and validation curves indicates that the model is starting to memorize patterns in the training data too specifically.

Data Limitations: The availability of a complete, diverse, and accurately labeled strawberry leaf image dataset can be a barrier. Data sets with limited variation (for example, only created under the same lighting conditions or backgrounds) can make it difficult for models to recognize images acquired directly in the field under non-standard conditions.

Conclusion and Future Work

Key Findings: This study demonstrates that a deep learning model optimized using transfer learning (supported by effective image processing techniques, such as those in your code) effectively supports strawberry leaf disease classification. While the overall accuracy on the test data reached 63%, some classes performed exceptionally well (e.g., class 2 with an F1 score of 0.95), while others still have room for improvement. The resulting classification report and confusion matrix can provide a deeper understanding of the model's strengths and weaknesses.

Key Contributions: This project makes a significant contribution to the field of precision agriculture by demonstrating the effectiveness of deep learning models in accurately and efficiently diagnosing plant diseases. This research has produced a promising model, enabling strawberry farmers to make timely decisions and potentially increase their yields. Possible Future Improvements and Developments

References

[1] A. Herbert, P. Sitohang, TI Hermanto, and CD Lestari, “Strawberry Plants Using the Convolutional Neural Network Architect Method,” JITET (Journal of Inform. and Tech. Elektro Ter. , vol. 12, no. 3, pp. 4146–4154, 2024, [Online]. Available: http://dx.doi.org/10.23960/jitet.v12i3$1.5274

[2] TA Nasution, Yeni Herdiyeni, Wisnu Ananta Kusuma, Budi Tjahjono, and Iskandar Zulkarnaen Siregar, “Artificial Intelligence for Monitoring Pests and Diseases in Eucalyptus Plants: Systematic Literature Review,” J. Comput. Sci. and Agri-Informatics , vol. 10, no. 2, pp. 224–237, 2023, doi: 10.29244/jika.10.2.224-237.

[3] W. Nengsih, "Processing of Products from Strawberry Farmers and MSMEs in Lebakmuncang Tourism Village," Ikra-Ith Abdimas , vol. 9, no. 2, pp. 108–116, 2025, doi: 10.37817/ikra-ithabdimas.v9i2.4073.

[4] D. Nurani, IL Yanuar, and AD Putra, "Classification of Disease Types in Rice Leaf Images Using the Convolution Neural Network Algorithm," J. Technol. Inf. and Comput. , vol. 8, no. 3, pp. 525–530, 2022, doi: 10.36002/jutik.v8i3.2074.

[5] Alfitriana Riska, Purnawansyah, H. Darwis, and W. Astuti, "Comparative Study of Combination of GMI, HSV, KNN, and CNN on Herbal Leaf Classification," Indones. J. Comput. Sci. , vol. 12, no. 3, pp. 1201–1215, 2023, doi: 10.33022/ijcs.v12i3.3210.

[6] Rio Subandi, Herman, and Anton Yudhana, "Pre-Processing in Medical Image Classification of Pneumonia," Decod. J. Educ. Technol. Inf. , vol. 4, no. 1, pp. 86–93, 2023, doi: 10.51454/decode.v4i1.198.

[7] D. Lizard Sambawo Dimara, B. Rahmat, and H. Maulana, "Identification of Rice Leaf Diseases Using the Mobilenet Transfer Learning Method - Support Vector Machine," JATI (Jurnal Mhs. Tek. Inform. , vol. 9, no. 3, pp. 5413–5420, 2025, doi: 10.36040/jati.v9i3.14245.

[8] V. Salim, A. Abdullah, and PY Utami, "Image Classification of Plant Diseases on Paprika Leaves with Transfer Learning Method Using DenseNet-201," Indones. J. Comput. Sci. , vol. 13, no. 2, pp. 3001–3014, 2024, doi: 10.33022/ijcs.v13i2.3746.

[9] CN Syahputri and MS Hasibuan, "Optimizing Decision Tree Classification Using Pruning Techniques to Reduce Overfitting," JSiI (Journal of Information Systems) , vol. 11, no. 2, pp. 87–96, 2024, doi: 10.30656/jsii.v11i2.9161.

[10] N. Puspita Sari, "Analysis of CNN Algorithm Performance in Deep Learning-Based Medical Image Classification," J. Comput. , vol. 2, no. 2, pp. 87–92, 2024.