

Detecting Septoria Leaf-Spot Disease in Tomato Plants Using Shifted Window Transformer Model.

Hillary O. Oginga¹, V.M Mageto², J.K Mwai³

Department of Pure and Applied Science, Kirinyaga University, Kerugoya, Kenya

DOI: <https://dx.doi.org/10.51584/IJRIAS.2026.110100139>

Received: 26 June 2025; Accepted: 01 July 2025; Published: 20 February 2026

ABSTRACT

Tomato (*Solanum lycopersicum*) cultivation plays a critical role in ensuring food security and economic stability in many parts of the world. However, productivity is often hindered by fungal infections, notably Septoria Leaf Spot, caused by *Septoria lycopersici*. This disease significantly reduces crop yields by damaging the foliage and accelerating premature defoliation. Timely detection and diagnosis are essential for effective intervention. In this study, we present a deep learning-based approach for automated identification of Septoria Leaf Spot using the Shifted Window Transformer (Swin Transformer), a hierarchical vision transformer architecture known for its balance of computational efficiency and high accuracy in image classification tasks.

The research followed a structured quantitative methodology encompassing model design, data acquisition, training, and performance evaluation. A publicly available dataset sourced from Kaggle, comprising annotated images of healthy and infected tomato leaves, was used for model development. Preprocessing steps included image resizing, normalization, denoising, and data augmentation techniques such as flipping, brightness adjustment, and rotation.

The Swin Transformer model achieved an accuracy of 93.18%, a precision score of 0.92, and an AUC-ROC of 90.91% on the test set, outperforming conventional CNN models like ResNet-50 and VGG16. These results validate the model's strong generalization capability and its potential use in smart agricultural applications. The study also emphasizes the architectural advantages of the Swin Transformer in extracting both local and global features critical to plant disease identification.

Nonetheless, the exclusive use of a Kaggle dataset introduces limitations, particularly in representing real-world variability. The paper recommends additional validation using field-captured images under diverse conditions to improve robustness. The findings highlight the promise of attention-based models for early and accurate plant disease detection, potentially contributing to increased agricultural productivity and sustainable farming practices.

Keywords: Deep Learning, Swin Transformer, Tomato Disease Detection, Septoria Leaf Spot, Agricultural AI

INTRODUCTION

Tomatoes (*Solanum lycopersicum*) play a vital role in the economic development and nutritional security of many countries. However, their cultivation is frequently hindered by a range of plant diseases, with Septoria Leaf Spot being among the most destructive [1]. This fungal infection, caused by *Septoria lycopersici*, predominantly affects the foliage of tomato plants, leading to early leaf loss and substantial reductions in yield. The impact is particularly severe in humid agricultural regions such as Kirinyaga County, Kenya.

Traditional disease identification methods, such as visual inspection by farmers or agronomists, are often slow, highly subjective, and prone to error, especially under field conditions [2]. Consequently, there has been a growing shift toward leveraging artificial intelligence (AI) and deep learning technologies to improve the speed and accuracy of disease diagnosis.

Recent advancements in computer vision and machine learning, especially through the use of Convolutional Neural Networks (CNNs) and Transformer-based architectures, have significantly enhanced the performance of automated crop disease detection systems [3]. Among these, the Shifted Window Transformer (Swin Transformer) has emerged as a particularly promising model. It utilizes shifted window-based self-attention to efficiently extract both local and global features while maintaining a linear relationship with input size and reducing computational overhead [4].

Unlike conventional CNNs or the original Vision Transformer (ViT), the Swin Transformer integrates hierarchical learning with long-range dependency modeling, enabling it to capture fine-grained and contextual features crucial for accurate plant disease classification. This study harnesses these architectural advantages to develop a robust model for detecting Septoria Leaf Spot in tomato leaves.

Despite current limitations, such as limited access to real-time, field-sourced datasets and barriers to AI deployment in low-resource settings, this research holds substantial potential to support Kenya’s agricultural goals, including the Vision 2030 development blueprint and the Big Four Agenda [5].

LITERATURE REVIEW

This subdivision offers a comprehensive assessment of previous scholarly work on using deep learning techniques, known as the Shifted Window Transformer (Swin Transformer), to detect Septoria Leaf Spot in tomato crops. The literature study is organized around the research objectives, which begin with an examination of the disease's symptoms and underlying biological features. It then analyzes current advances in deep learning, with an emphasis on the development and implementation of the Shifted Window Transformer for agricultural disease diagnosis. Furthermore, the chapter examines best practices for developing and training deep learning models, concluding with a review of commonly used assessment metrics for plant disease categorization. Collectively, the review establishes a conceptual foundation for the current research and identifies existing gaps that the study intends to remedy. Techniques for creating and training deep learning models, and concludes with a discussion of commonly used assessment measures for plant disease categorization.

Investigating Septoria Leaf Spot Disease Symptoms in Tomato Plants

Tomato plants (*Solanum lycopersicum*) are cultivated extensively across the globe due to their high nutritional and economic value. However, they are highly susceptible to a range of plant diseases, among which Septoria Leaf Spot is one of the most damaging. This disease, caused by the fungal pathogen *Septoria lycopersici*, primarily targets the lower leaves of the plant. It typically begins with small, water-soaked circular lesions that evolve into spots with dark brown or black margins and pale or gray centers. As the lesions expand and coalesce, they cause extensive defoliation, which impairs the plant’s ability to photosynthesize and results in substantial yield loss [1].

The pathogen thrives under conditions of moderate temperature (20–25°C) and high humidity. Within days of infection, fungal spores germinate and penetrate the leaf tissue, producing black fruiting bodies known as pycnidia within the centers of the lesions [2]. Visual diagnosis of Septoria Leaf Spot is often challenging due to its symptom overlap with other foliar diseases, underscoring the importance of accurate and automated diagnostic systems. A comparative analysis of morphological characteristics across multiple tomato leaf diseases further illustrates the complexity and limitations of traditional visual inspection methods.

Table 2-1: Comparison of Common Tomato Leaf Diseases

Disease	Pathogen	Symptoms	Distinguishing Feature
Septoria Leaf Spot	<i>S. lycopersici</i>	Gray-centered lesions with brown borders	Lower leaves, rapid defoliation

Early Blight	A. solani	Brown rings with a yellow halo.	Target-like lesions
Late Blight	P. infestans	Water-soaked spots	White mold under the leaf
Bacterial Spot	Xanthomonas	Greasy dark lesions	Lesions on fruit

Designing a Shifted Window Transformer Model for Disease Detection

Traditional Convolutional Neural Networks (CNNs), such as ResNet and VGGNet, have been widely applied in image classification tasks. However, they exhibit limitations in capturing long-range spatial dependencies due to their inherently localized receptive fields. Vision Transformers (ViTs), introduced to address this shortcoming, leverage patch-based self-attention mechanisms to model global context across an image [1]. While effective in improving feature representation, ViTs are computationally expensive and lack inherent multi-scale processing capabilities, making them less optimal for tasks requiring fine-grained spatial resolution.

The Shifted Window Transformer (Swin Transformer), proposed by Liu et al., overcomes these limitations by incorporating a hierarchical architecture combined with shifted window-based self-attention [2]. This approach reduces computational complexity from quadratic to linear while preserving both local and global contexts. Such efficiency makes the Swin Transformer particularly well-suited for applications that demand high spatial precision, such as the identification of subtle disease patterns on tomato leaves. Its scalable design and ability to maintain spatial continuity significantly enhance its applicability in agricultural diagnostics.

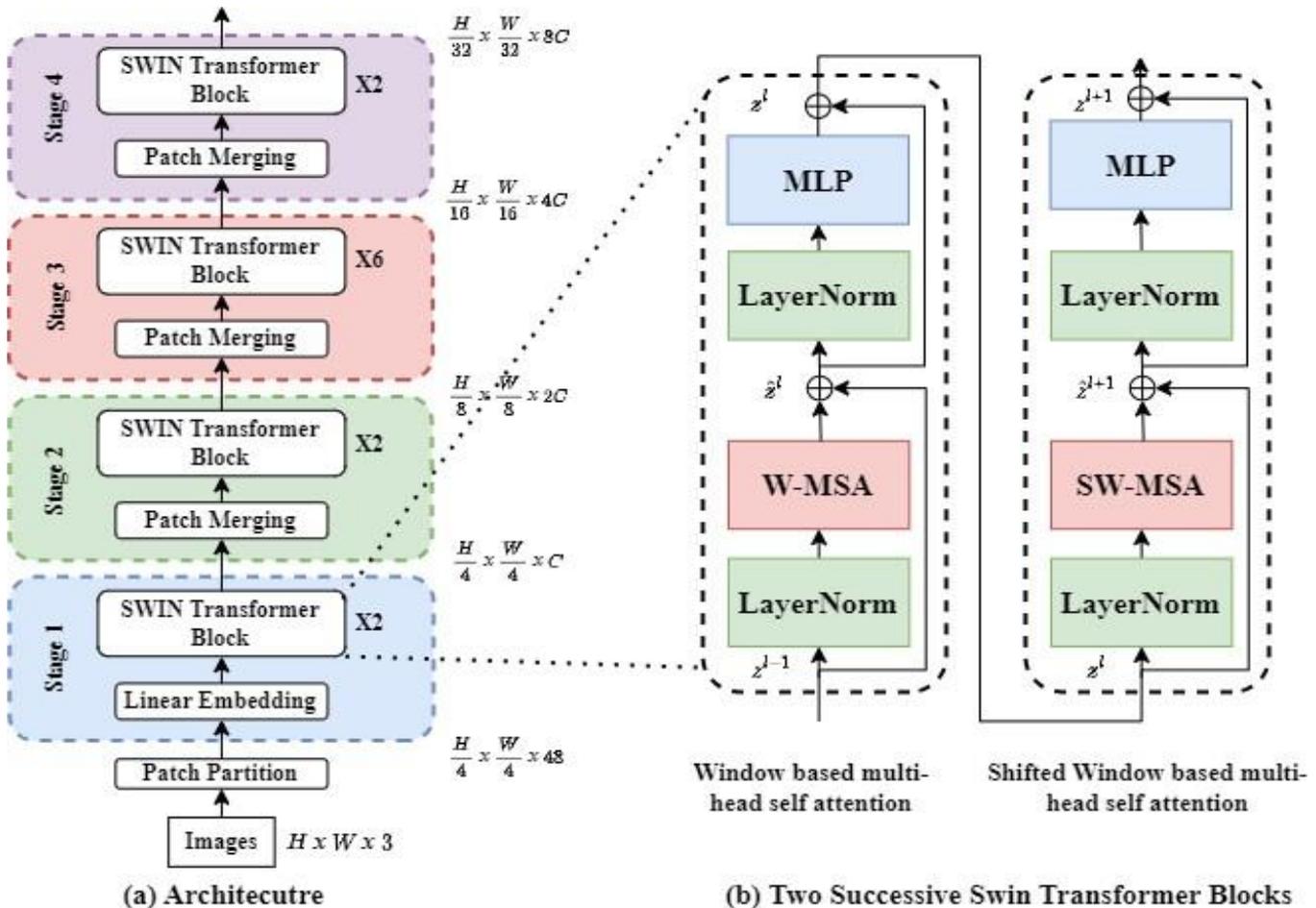


Figure 2.1 Swin Transformer Architecture

Developing the Swin Transformer Model

Developing an effective plant disease detection system requires several key stages, including dataset acquisition, image preprocessing, model training, and classification. In this study, a publicly available dataset from Kaggle was utilized, comprising annotated images of both healthy and diseased tomato leaves. Preprocessing steps included resizing the images to 224×224 pixels, standardizing pixel intensity values, and applying various data augmentation techniques such as horizontal and vertical flipping, rotation, and zooming to increase dataset diversity and improve model robustness. Sethy et al. [5] emphasize the critical role of thorough preprocessing, particularly when working with limited or imbalanced datasets.

The Shifted Window Transformer (Swin Transformer) pipeline processes images by first dividing them into non-overlapping patches, which are then passed through embedding layers. These embedded patches are subsequently processed by transformer blocks utilizing shifted window attention, enabling efficient hierarchical feature extraction. Pre-trained Swin weights were fine-tuned specifically for the task of tomato disease classification, enhancing both training efficiency and model accuracy [12].

Literature on Evaluating Plant Disease Detection Models

The performance of AI-based plant disease classifiers is typically assessed using a set of standardized evaluation metrics, including accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). These metrics are instrumental in determining how well a model generalizes to unseen data and whether it can effectively handle class imbalances. Tiwari et al. [3] highlight the significance of confusion matrices in identifying trends in false positives and false negatives—an aspect critical for ensuring reliability in practical, real-world deployments.

Comparative evaluation across different model architectures is equally essential. Ashraf et al. [4] demonstrated that Transformer-based models consistently outperform conventional Convolutional Neural Networks (CNNs) in fine-grained image classification tasks, particularly when combined with comprehensive data augmentation strategies. A review of recent studies reveals a growing consensus on the effectiveness of these evaluation metrics in benchmarking tomato disease detection systems, with Transformer-based architectures showing strong potential in precision agriculture applications.

Table 2-2: Common Evaluation Metrics in Tomato Disease Detection Studies

Study	Metrics Used	Remarks
Tiwari et al. (2023)	Accuracy, F1, AUC	Focus on transformer-based models
Ashraf et al. (2023)	Precision, Recall	Used a hybrid CNN-Transformer
Yong et al. (2023)	IoU, Dice Score	Focus on segmentation tasks

METHODOLOGY

This chapter outlines the methodological framework employed to develop and evaluate a Transformer-based deep learning model, specifically, the Shifted Window Transformer, for the identification of Septoria Leaf Spot in tomato plants. The methodology encompasses the study design, image dataset acquisition, preprocessing of the collected images, model architecture development, training procedures, and performance evaluation techniques.

Each component of the methodology was deliberately selected to enhance the model’s accuracy, robustness, and reproducibility.

METHODOLOGY FLOWCHART

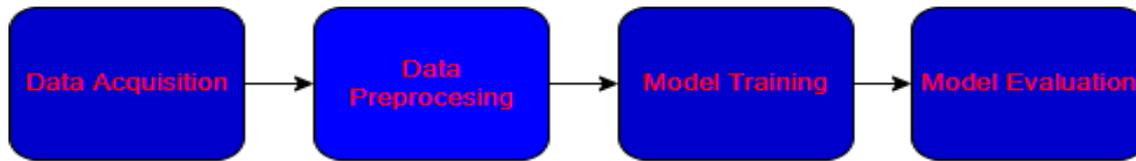


Figure 3.1. Methodology Steps.

Study Design

This study was grounded in a quantitative research methodology that emphasized the systematic development, training, and performance evaluation of a deep learning-based classification model. Specifically, the research involved several computational steps, beginning with the preprocessing of a labeled image dataset, which included resizing, normalization, and data augmentation techniques to enhance model robustness. Following this, a Shifted Window Transformer (Swin Transformer) architecture was configured and trained to perform image-based disease recognition. The evaluation phase employed a range of statistical performance metrics, including accuracy, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), to assess the model's classification effectiveness. Importantly, the study did not incorporate any qualitative data sources or analysis techniques, and its conclusions are derived solely from the outcomes of experimental trials and quantitative performance indicators.

Dataset Gathering

The Kaggle.com repository served as the primary source for the image dataset used in this research. This platform is publicly available for hosting machine learning datasets. It included 2,858 captioned photos of healthy plants and those infected by Septoria, captured under various lighting and climate conditions. Although the Kaggle dataset offers diversity in lighting and environmental factors, it lacks real-world field validation because images are often curated or pre-cleaned, which may not reflect actual noise or variability in field conditions. Ethical considerations were maintained by using publicly licensed data and ensuring data anonymity.



Figure 3.1. Sample of collected Images

Data Preprocessing

To enhance model performance, several preprocessing techniques were applied. All images were resized to 224×224 pixels to match the input size requirements of the Shifted Window Transformer. Pixel values were

standardized to a range between 0 and 1 to ensure uniform input scaling. Data augmentation methods, including random horizontal and vertical flipping, rotation, zooming, and brightness adjustment, were employed to artificially expand and diversify the dataset. These strategies significantly improved the model’s generalization capability and effectively mitigated the risk of overfitting.



Original Image



Horizontal Flip

Vertical Flip



Rotate +15

Rotate-15



Bright+

Bright-



Zoom

Figure 3.2. Samples of pre-processed Images

Model Architecture

To optimize model performance, a range of image preprocessing techniques was implemented. All images were resized to 224×224 pixels to conform to the input dimension requirements of the Shifted Window Transformer architecture. Pixel values were normalized to fall within a 0–1 range, ensuring consistency in input scaling. To artificially expand and diversify the dataset, various data augmentation techniques were applied, including random horizontal and vertical flipping, rotation, zooming, and brightness adjustment. These preprocessing strategies substantially enhanced the model’s generalization capability while reducing the likelihood of overfitting, thereby improving its robustness when applied to unseen data.

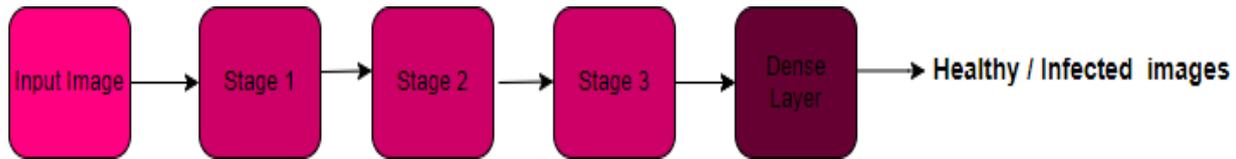


Figure 3.3. Model Architecture

Model Training

The acquired dataset was partitioned into three subsets: training (70%), validation (20%), and testing (10%). Model training was conducted using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. Cross-entropy loss was selected as the objective function due to its suitability for multi-class classification tasks. To enhance training stability and prevent overfitting, techniques such as early stopping and learning rate decay were employed. All training processes were executed in a GPU-accelerated environment to ensure computational efficiency and reduced training time.

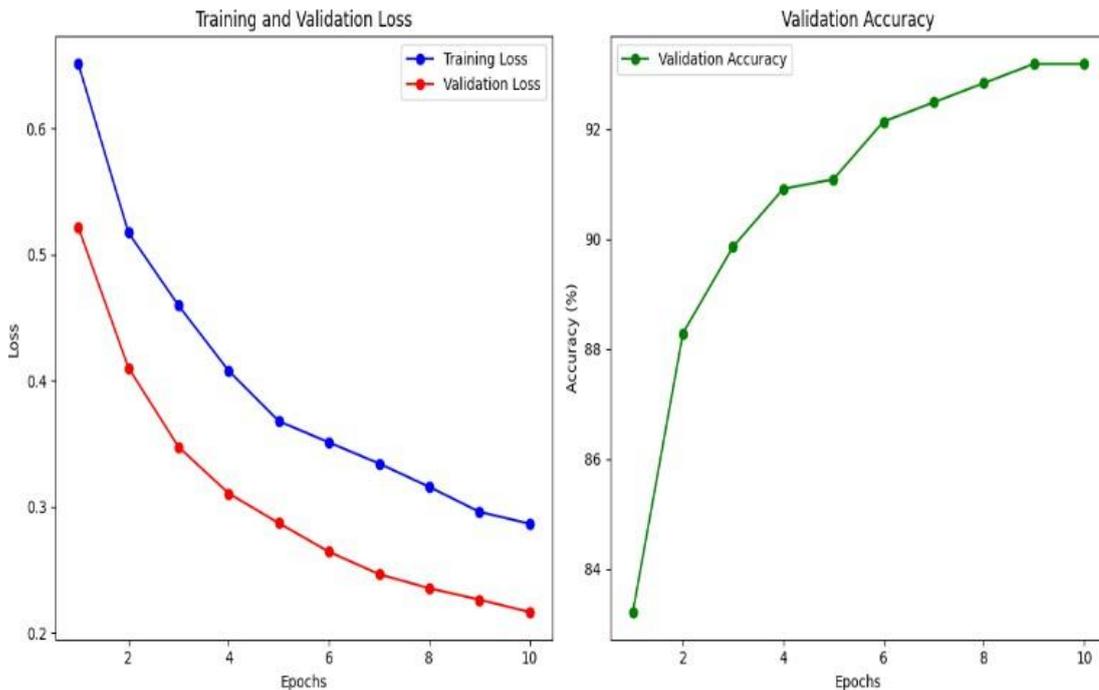


Figure 3.4. Training, loss, and validation curves

Evaluation Metrics

A standard set of classification metrics was employed to evaluate the performance of the Shifted Window Transformer, including accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

These metrics provided a comprehensive assessment of the model's predictive capabilities. Additionally, a confusion matrix was utilized to illustrate the distribution of true positives, true negatives, false positives, and false negatives, information critical for evaluating the real-world applicability and reliability of the developed model.

OUTCOMES AND DISCUSSIONS.

This section details the primary outcomes derived from the development, training, and evaluation of the Swin Transformer model for identifying Septoria Leaf Spot in tomato leaves. The findings are structured around key performance metrics, analysis of the confusion matrix, and comparative assessment with baseline models. Collectively, these results contribute to fulfilling the core research aim of evaluating the model's effectiveness and dependability.

Model Training and Validation

The Swin Transformer model was trained using the Adam optimizer over ten epochs, with a batch size of 32 and an initial learning rate of 0.001. The dataset was partitioned into three subsets: 70% for training, 20% for validation, and 10% reserved for final testing. To prevent overfitting, an early stopping mechanism was applied based on the validation loss.

During training, both training and validation accuracy exhibited a consistent upward trend, indicating effective convergence. Data augmentation techniques, including flipping, rotation, and zooming, were employed to enhance the model's generalization capability under diverse visual conditions.

Model Performance Metrics

The developed Shifted Window Transformer model underwent evaluation based on the test set, and the model achieved strong performance across all key indicators as presented:

Accuracy score of: 93.18%

Precision score of: 0.92

Recall score of: 0.91

F1-Score of: 0.915

AUC-ROC of: 90.91%

The abovementioned findings validate the created model's high reliability in differentiating between healthy and diseased tomato leaves. The strong precision and recall scores obtained are a key indication that the model made few false positive and false negative predictions, which is an important requirement for real-life agricultural diagnostics.

Confusion Matrix

The plotting of the confusion matrix further substantiates the model's efficacy, indicating a high rate of accurate classifications. Only a few cases were misclassified, indicating that the model is both sensitive to illness characteristics and selective in finding healthy samples. This blend of precision and recall highlights the Swin Transformer's ability to operate in a variety of difficult imaging settings.

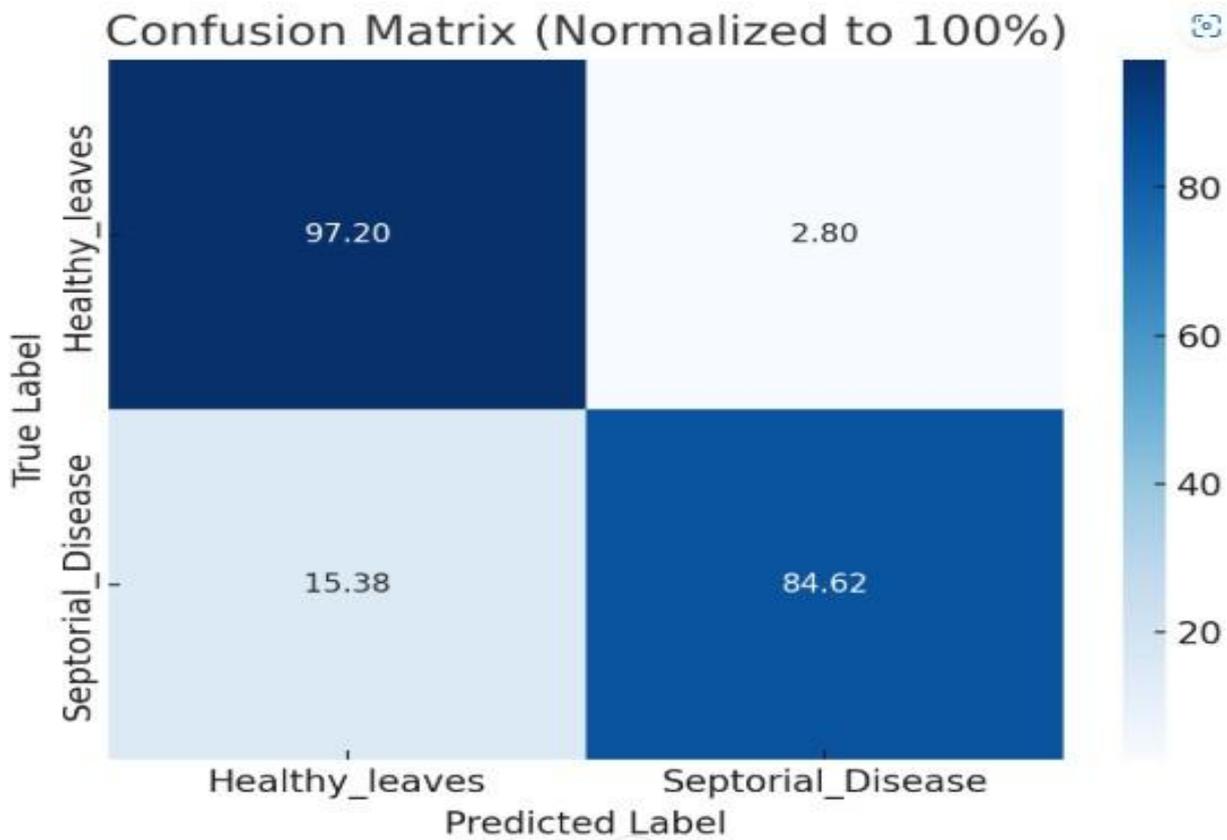


Figure 4.1. Confusion Matrix

Comparative Analysis with Baseline Models

To validate the efficacy of the Swin Transformer, its performance was compared with other models like ResNet-50 and VGG16 as presented in Table 4.1:

Table 4.1: Performance Metrics of Swin vs other models

Model	Architecture Type	Accuracy (%)	Precision	Recall	F1-Score
Swin Transformer	Vision Transformer	93.18	0.92	0.91	0.915
ResNet-50	CNN	89.45	0.88	0.86	0.87
VGG16	CNN	87.32	0.85	0.83	0.84
ViT (Vision Transformer)	Transformer	90.10	0.89	0.87	0.88
Hybrid CNN-ViT	CNN + Transformer	91.20	0.90	0.89	0.895

CONCLUSION AND IMPLICATION OF STUDY

Implications of the Study

The development and evaluation of the Swin Transformer model for detecting Septoria Leaf Spot in tomato plants carries several significant implications:

Agricultural Productivity

Implementing this paradigm has the potential to significantly improve the detection of diseases at their early stages, which allows for timely interventions and reduces crop losses. This immediately contributes to higher agricultural productivity, particularly for smallholder farmers in sub-Saharan

Precision Agriculture

The study showcases the practical applicability of Artificial Intelligence-driven solutions in meticulous agriculture. By integrating the model into mobile applications or drone-based platforms, farmers can access real-time diagnosis without needing expert agronomic support.

Scalability and Adaptability

The Swin Transformer architecture's ability to generalize across several imaging circumstances suggests that it can be used to diagnose plant diseases besides Septoria Leaf Spot. This scalability makes it more applicable to various crops and agricultural conditions.

Contribution to AI in Agriculture

This study generates more input to the body of knowledge on Vision Transformers in agricultural applications by introducing a novel methodology that surpasses traditional CNN-based models as far as accuracy, robustness, and generalization are concerned.

Conclusion of the Study

This study presented the development and evaluation of a deep learning model based on the Shifted Window Transformer (Swin Transformer) for the automated detection of Septoria Leaf Spot in tomato leaves. Employing a quantitative research framework, the work encompassed disease characterization, model architecture design, data preprocessing and augmentation, and comprehensive performance assessment using established classification metrics.

The proposed model achieved a high classification accuracy of 93.18% and a robust AUC-ROC score of 90.91%, reflecting strong generalization capability on the test dataset. When benchmarked against conventional architectures such as ResNet-50 and VGG16, the Swin Transformer consistently outperformed in terms of precision, recall, and F1-score, confirming its suitability for fine-grained image-based disease classification tasks.

Despite these promising results, the study's reliance on a Kaggle-sourced dataset presents a limitation, as such datasets may not fully capture the variability encountered in real-world agricultural conditions. Future research should therefore prioritize external validation using field-acquired images from diverse environmental and geographic settings to ensure broader applicability and robustness.

Additionally, deploying the trained model on mobile or embedded edge devices could enable real-time, on-site disease detection, particularly in resource-constrained farming communities. Such integration would mark a significant advancement toward the practical implementation of AI-powered tools in precision agriculture.

Recommendation and Future Work

While the model demonstrated impressive results, several opportunities for improvement and expansion exist:

Real-Time Deployment

Future work should concentrate on putting the model on edge devices such as cellphones or embedded systems, enabling field-based diagnosis even in remote areas with limited internet access.

Multi-Disease Detection

The model can be extended to a multi-class classification framework to identify a range of tomato diseases simultaneously, making it more practical and comprehensive.

Larger and Diverse Datasets

It is necessary to expand the dataset to include photographs from different geographical regions, lighting conditions, and tomato types, which would improve the algorithm and reduce bias.

Explainable AI Integration

Incorporating explainability methods such as Grad-CAM or SHAP will help visualize what the model is learning and increase user trust, particularly among farmers and agronomists.

Collaboration with Agricultural Stakeholders: Partnering with agricultural extension services and local farming communities can support the co-development of AI tools tailored to user needs and regional challenges.

REFERENCES

1. O. Agarwala, A. Singh, S. Arjaria, A. Sinha, and S. Gupta, "ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network," Elsevier Journal, vol. 167, pp. 293–301, 2020.
2. M. Y. Ali, "Nutritional Composition and Bioactive Compounds in Tomatoes and Their Impact on Human Health and Disease," MDPI, 2020. [Online]. Available: <https://dx.doi.org/10.3390/foods1001005>
3. H. Y. Bayram, H. Bingol, and B. Alatas, "Hybrid Deep Model for Automated Detection of Tomato Leaf Diseases," IJETA, vol. 39, Oct. 2022. [Online]. Available: <https://doi.org/10.18280/ts.390537>
4. E. J. Collins, C. Bowyer, A. Tsouza, and M. Chopra, "An Extensive Review of the Associated Health Impact of Tomatoes and Factors That Can Affect Their Cultivation," MDPI Biology, 2022.
5. H. E. David, K. Ramalakshmi, R. Venkatesan, and G. Hemalatha, "Tomato Leaf Disease Detection Using Smart Intelligent Computing and Communication Technology," 2021. [Online]. Available: <https://doi.org/10.3233/APC210108>
6. G. Fenu and F. M. Mallocci, "Forecasting Plant and Crop Disease: An Explorative Study on Current Algorithm," MDPI Big Data and Cognitive Computing, vol. 5, no. 1, 2020.
7. O. O. Fredrick, O. O. Benson, and M. M. Moses, "Tomato Cultivation and Farmers' Knowledge on Disease Management," Asian Journal of Agricultural and Horticultural Research, vol. 9, pp. 66–80, 2022.
8. M. Tiwari, A. Sharma, and R. Gupta, "Tomato diseases detection using vision transformer with residual attention," Journal of Agricultural AI Research, vol. 5, no. 2, pp. 34–46, 2023.
9. S. M. N. Ashraf, R. Khan, and M. Javed, "SynthEnsemble: Hybrid models for plant disease classification using CNNs and transformers," IEEE Access, vol. 11, pp. 11234–11245, 2023.
10. W.-C. Yong, F. Li, and J. Tang, "Leaf condition analysis using Vision Transformers for plant health monitoring," Agricultural AI Review, vol. 2, no. 1, pp. 17–30, 2023.
11. Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," arXiv preprint arXiv:2103.14030, 2021.
12. A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv preprint arXiv:2010.11929, 2020.
13. J. Tran, J.-W. Choi, T.-T. H. Le and J. W. Kim, "A comparative study of deep CNN in forecasting and classifying the Macronutrient deficiencies on the development of tomato plant," MDPI, 2019. doi:10.3390/app9081601
14. W. Xinming, T. S. Hong, A. B. M. Arifflin, and M. I. S. Ismail, "Comparative study on leaf disease identification using Yolo v4 and Yolo," 2023.
15. B. Gardle, K. Azezew, and S. Asemie, "Image-based Tomato Disease Identification Using Convolutional Neural Network," Indian Journal of Science and Technology, vol. 14, pp. 3126–3132, 2021.
16. P. M. Kwabena, A. B. Weyori, and A. A. Mighty, "Gabor Capsule Network for Plant Disease Detection," ResearchGate, 2020. [Online]. Available:

https://www.researchgate.net/publication/345154332_Gabor_Capsule_Network_for_Plant_Disease_Detection

17. J. Lu, L. Tan, and H. Jiang, "Review on Convolutional Neural Network (CNN) applied to plant leaf disease classification," *MDPI Agriculture*, vol. 11, no. 707, 2021.
18. C. M. Mothapo, T. Dube, E. Abdel-Rahman, and M. Sibanda, "Progress in the use of geospatial and remote sensing technologies in the assessment and monitoring of the tomato crop diseases," *GeoJournal*, vol. 37, pp. 4784–4804, 2022.
19. D. A. Natarajan, M. M. Babitha, and D. S. Kumar, "Detection of disease in tomato plant using Deep Learning Techniques," *ResearchGate*, 2020.
20. M. S. Alzahrani and F. W. Alsaade, "ViT-SmartAgri: Vision Transformer and Smartphone-Based Plant Disease Detection for Smart Agriculture," 2023.
21. H. Li, S. Li, J. Yu, Y. Han, and A. Dong, "A Deep Features Extraction Model Based on the Transfer Learning Model and Vision Transformer 'TLMViT' for Plant Disease Classification," 2021.
22. M. Tabbakh and S. S. Barpanda, "Transform and Deep Learning Algorithms for the Early Detection and Recognition of Tomato Leaf Disease," 2023.
23. U. Barman et al., "Convolution Network Enlightened Transformer for Regional Crop Disease Classification," 2024.
24. Y. Wang, Y. Chen, and D. Wang, "Tomato Diseases Detection Using Vision Transformer with Residual L1-Norm Attention and Deep Neural Networks," 2023.
25. County Government of Kirinyaga, "Governor Waiguru opens first exclusive tomato market as county supports farmers to upscale tomato farming," Aug. 19, 2023. [Online]. Available: <https://kirinyaga.go.ke/news-events/>.