SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I6P128

Volume 12 Issue 6, 352-363, June 2025
© 2025 Seventh Sense Research Group®

Original Article

A Hybrid Transformer-Based Approach for Arecanut Disease Prediction

Sangeetha Shibu[1], Jinu Raj R[2], Divya G S[3], Jincy Jesudasan[4]

[1,2,3,4]*Department of Computer Science and Engineering, Rajadhani Institute of Engineering and Technology, India.*

[1]*Corresponding Author : Sangeetha.shibu@outlook.com*

Received: 15 April 2025 Revised: 16 May 2025 Accepted: 18 June 2025 Published: 27 June 2025

Abstract *- Crop development is affected by several variables, such as climatic conditions, soil quality, and diseases, which significantly affect yield and productivity. Arecanut, or betel nut, is a tropical crop susceptible to various diseases affecting different parts of the plant, from root to fruit. Accurate and timely disease recognition is essential for maintaining crop health and enhancing agricultural productivity. Conventional disease identification techniques depend on manual examination, which consumes time and is susceptible to inaccuracies. Existing models often face challenges in extracting long range dependencies and global feature interactions, limiting the classification accuracies. This study presented a hybrid deep learning (DL) framework combining ResNet-50 and Swin transformer for better disease identification. The ResNet-50 model extracts hierarchical spatial features, while the Swin Transformer with shifted window self-attention captures global dependencies, enhancing classification by emphasising specific disease patterns. The framework is trained and examined using a dataset of Arecanut disease sourced from Kaggle with 11,063 images across nine disease categories. Findings demonstrated that the suggested framework attains a classification accuracy of 98.42%, outperforming conventional methods. The study highlights the effectiveness of incorporating transformer-based attention mechanisms in agricultural disease detection.*

Keywords *- Arecanut disease, Deep learning, Self-attention mechanism, ResNet-50, Swim transformer, Convolutional Neural Networks.*

1. Introduction

Agriculture is India's main occupation, making it the second-largest producer of agricultural products worldwide. India's economy heavily relies on agriculture, with farmers cultivating various crops to sustain livelihoods and meet market demands [1]. Among these, areca nuts are a commercially significant crop that grows widely in tropical regions. It plays an essential role in the livelihoods of thousands of farmers, particularly in states such as Kerala, Karnataka, Assam, and Maharashtra. The production of Arecanut is severely affected by numerous diseases that reduce both the yield and quality of the crops [2].

Arecanut crops are vulnerable to various diseases caused by fungal, bacterial and viral infections. Several biotic and abiotic stress factors contribute to disease outbreaks, impacting crop production. Some of the most common diseases include Mahali Koleroga, caused by Phytophthora species, which results in fruit rot and premature nut drop, and Stem Bleeding, a severe condition that weakens the tree structure and reduces its longevity. Other major diseases include Bud Borer, Stem Cracking, and Yellow Leaf Disease, each exhibiting unique symptoms that require close monitoring for early detection [3].

Traditional disease detection methods rely heavily on expert manual inspection, which is time-consuming, subjective, and prone to error—particularly in remote or large-scale plantations. Farmers and agricultural experts typically depend on visual observation to examine the severity of the diseases. However, such methods are inconsistent and inefficient, particularly for large-scale plantations.

Advancements in artificial intelligence (AI) and DL led to the emergence of automated disease diagnosis through computer vision (CV), offering improved precision and robustness. Integrating AI models in agriculture has significantly transformed crop health monitoring, allowing real-time analysis and accurate disease classification.

DL methodologies mainly convolutional neural networks (CNNs), illustrate striking capabilities in plant disease identification. While deep learning (DL) techniques have recently demonstrated strong performance in plant disease classification, most existing models either fail to fully capture complex spatial textures or cannot model long-range dependencies in visual data. The absence of global feature understanding results in misclassification, especially when diseases exhibit overlapping visual symptoms.

To overcome these problems, this study suggests a hybrid DL method that integrates ResNet-50 and Swin Transformer with an attention mechanism to enhance classification accuracy. ResNet-50 serves as the feature extractor, learning spatial patterns and disease-specific textures. Swin Transformer refines feature representations through shifted window self-attention, enabling the model to effectively process local and long-range dependencies. This study ensures a more robust and scalable approach for automated Arecanut disease identification. The important contributions of this research are outlined below:

- To create a hybrid DL framework integrating ResNet-50 and Swin Transformer to efficiently extract local spatial features and global contextual dependencies for Arecanut disease prediction.
- To evaluate the efficiency of the suggested framework in Arecanut disease detection by comparing its classification efficiency with existing approaches.

The subsequent portions of the study are presented in the following manner: A review of related works on Arecanut disease detection is outlined in Section 2. The details of the hybrid ResNet-50 and Swin Transformer model are explained in detail in Section 3. Also, Section 4 discusses the outcomes of the study and assesses the efficiency of the framework. The research is summarised in Section 5, along with ideas for upcoming research.

2. Related Works

Ghate et al. (2025) [4] employed the DL model to improve the consistency of Arecanut grading. A dataset of 2000 high-resolution images was collected and augmented with 8 CNN architectures evaluated, among which DenseNet 121 and Inception V3 attained an accuracy of 95.67% and 96%. Since these models showed efficiency, the dataset primarily contained frontal images, limiting the presentation of surface texture, color uniformity and hidden defects, which impacted the fine-grained classification accuracy. Pavan et al. (2025) [5] created an automated disease detection system for Arecanut diseases. The study utilised a trained ResNet 50 model with image preprocessing methods. The study demonstrated better accuracy in disease detection, offering farmers timely and actionable information for improved crop management. The study was limited due to its reliance on the quality of the image, where changes in the image's resolution, lighting conditions and capture angles affected the proficiency of the framework. Kumar et al. (2024) [6] focused on detecting and classifying diseases in Arecanut leaves using DL frameworks to offer an efficient alternative to traditional disease identification methods. A dataset of diseased Arecanut leaf images was gathered and split for training, validation and testing with ResNet, MobileNet and VGG Net, where VGG Net attained 92% accuracy, surpassing other models leading to its deployment in an Android application to help farmers in early detection and management. However, the model focuses only on Mahali (Koleroga), Stem Bleeding, and Yellow Leaf Disease, limiting its capability to detect other potential Arecanut diseases.

Naik and Rudra (2024) [7] performed a study on transfer learning-based categorisation of Arecanut X-ray images utilising both traditional CNN and quantum CNN (QCNN) approaches. The study was examined with 12 transfer learning (TL) models and found that QCNN surpassed CNN, attaining 97.72% accuracy. The study noted that the computational cost of quantum processing resulted in longer training times compared to conventional CNN techniques. Riza (2024) [8] developed an Arecanut disease detection application to help farmers identify diseases quickly and accurately using a machine learning (ML) model integrated into an Android system.

The study employed a CNN algorithm to process image data of 10 diseases and 32 symptoms using a teachable machine. The system attained an average accuracy, providing a user-friendly interface and fast detection process. The study lacks a discussion on latency, computational efficiency, and battery consumption, which impacts the real-time usability of Android devices in remote areas.

Chikkalingaiah et al. (2024) [9] created a DL-based approach to help Arecanut farmers estimate yield by segmenting Arecanut bunches from images. The study employed a U-Net squared model for segmentation and a modified YOLOv3 model for counting the nuts, attaining an accuracy of 88% on training and 85% on validation accuracy for segmentation, where the YOLO attained 94.7% accuracy in yield estimation. Results demonstrated that models perform well, allowing an effective solution for Arecanut yield estimation.

However, the U-Net squared model increases memory, and computational costs limit real-time deployment. Krishna et al. (2023) [10] studied the fruit rot disease (FRD) prediction scores in Arecanut, utilising previous meteorological data by applying DL frameworks. Meteorological and disease score data utilises a rule-based algorithm for training and evaluation. The Vanilla GRU framework, optimised and attained a minimum MSE of 0.0009 and an R2 score of 0.99, showed better performance with a low RMSE of 0.33. However, due to the limited dataset, there is a risk of overfitting, which limits the framework's capability to adapt to new weather patterns.

Patil et al. (2023) [11] aimed to classify the dehusked Arecanut into 5 categories employing a customised CNN and evaluated its performance with the standard AlexNet architecture. A dataset of 300 Arecanut was generated with a specialised instrumentation setup. The images were preprocessed before being given to the models for feature extraction. The customised CNN surpassed AlexNet, attaining

97.33% average accuracy and 97.48% F1 scores. AlexNet requires more memory, computation power, and storage, rendering it less appropriate for practical implementations. Hedge et al. (2023) [12] created a CNN-based system to detect and categorise diseases affecting Arecanuts, trunks and leaves. The framework was trained to employ a dataset of 1,100 images, which was split into 80:20 for training and validation. The suggested method effectively classified diseases and attained an accuracy of 93.05%, providing better and more accurate diagnoses for farmers. However, the study uses a limited dataset, which impacts the ability of the framework to generalise effectively.

Naik and Rudra (2023) [13] created an automated classification system for Arecanut quality assessment using X-ray imaging and DL. To enhance the efficiency of the framework, an adaptive genetic algorithm was employed to optimise the YOLOv5 architecture, utilising a custom-created dataset consisting of Arecanut X-ray images. The research attained a mAP of 97.84%, surpassing the other YONO models. However, the study faces a high computational demand for genetic algorithm optimisation. Mahaveen et al. (2023) [14] created a CNN-driven framework aimed at the late diagnosis of infections in the leaves of Arecanut, fruit and trunk to assist farmers in maintaining crop health. The study utilised datasets containing healthy images and diseased images of Arecanut samples. The framework attained 88% accuracy in determining diseases like Stem bleeding, Mahali and yellow leaf while providing preventive measures and corrective actions. However, its performance depends on the disease stage and image quality. Jenitta and Swetha (2023) [15] created a CNN-based system for detecting Arecanut diseases affecting leaves and trunk. The study utilised a dataset of 200 images, split into a 70:30 ratio for training and testing. The framework trained over 50 epochs attained 81.35% accuracy in identifying diseases like stem bleeding, mahali and yellow leaf spot. However, the model was trained on a limited dataset, which is insufficient across different environmental conditions.

Patil et al. (2023) [16] introduced a CNN-driven method for intelligent Arecanut assessment to reduce manual labor and improve efficiency in the segregation process. Utilising a dataset of Areca nuts cultivated in the Western Ghats region, the study performed a 10-fold cross-validation method with contrast enhancement and no data augmentation.

The custom CNN model attained the results with a standard deviation of 4.1% for cropped images with contrast enhancement, illustrating the feasibility of automated Arecanut segregation. In order to create a high-precision monitoring method for Areca yellow leaf disease (YLD), Xu et al. (2023) [17] used thermal infrared and multispectral data collected from an Areca orchard by an unmanned aerial vehicle (UAV). Ten vegetation indices and the Relief feature selection method were used in the study to generate 6 ML models. The random forest (RF) model attained 95.5% accuracy and an RMSE value of 0.049. Also, the study noted that the results were affected by the lighting conditions during imaging, which led to errors in multispectral data analysis. Balanagouda et al. (2023) [18] studied the efficiency of oomycete-specific fungicides in managing the FRD of Arecanut under different application timings and fruit development stages. The study employed generalised linear mixed models (GLMMs) to analyse FRD severity, occurrence, and cumulative fallen nut rate across two experimental approaches based on monsoon periods. The results showed that the application of fungicide reduced FRD by over 65%. Existing studies on Arecanut disease detection primarily employ DL and ML models, but they lack advanced feature extraction for handling complex disease patterns [5, 8, 12]. Many approaches suffer from dataset limitations, leading to challenges in generalising models across diverse environmental conditions [6]. Most datasets focus on specific diseases, missing other emerging diseases that could be critical for farmers [6]. Additionally, image quality variations contribute to inconsistencies in detection performance [5]. CNN-based models suffer from limited spatial dependencies and struggle with complex texture and fine-grained disease classification [4, 11]. High computational demand limits the feasibility of real-time disease detection on low-power agricultural devices [13]. The UAV-based multispectral analysis is sufficient, but it is affected by the environmental conditions [17]. To overcome these limitations, this study integrates ResNet 50 for feature extraction and Swin transformer for classification, ensuring higher precision in real-world applications.

3. Materials and Methods

Arecanut disease detection employs DL methods to examine and classify images of Arecanut plants, helping in early diagnosis and improved agricultural management. The images collected from the dataset undergo data preprocessing and augmentation, which involves resizing, normalisation and enhancement techniques to improve the efficiency of the framework. The preprocessed dataset is divided into separate subsets for model training and testing.

The hybrid categorisation framework consists of two main components: ResNet-50 captures the hierarchical spatial characteristics from the input images, and the Swin transformer refines these extracted features by employing self-attention mechanisms. A Fully Connected (FC) layer was utilised to process the features extracted, and classification was performed using a softmax function. Finally, the trained framework accurately classifies the input image into its respective disease category, ensuring precise disease identification and effective model validation. Figure 1 presents the systematic workflow of the suggested hybrid approach for Arecanut disease identification.

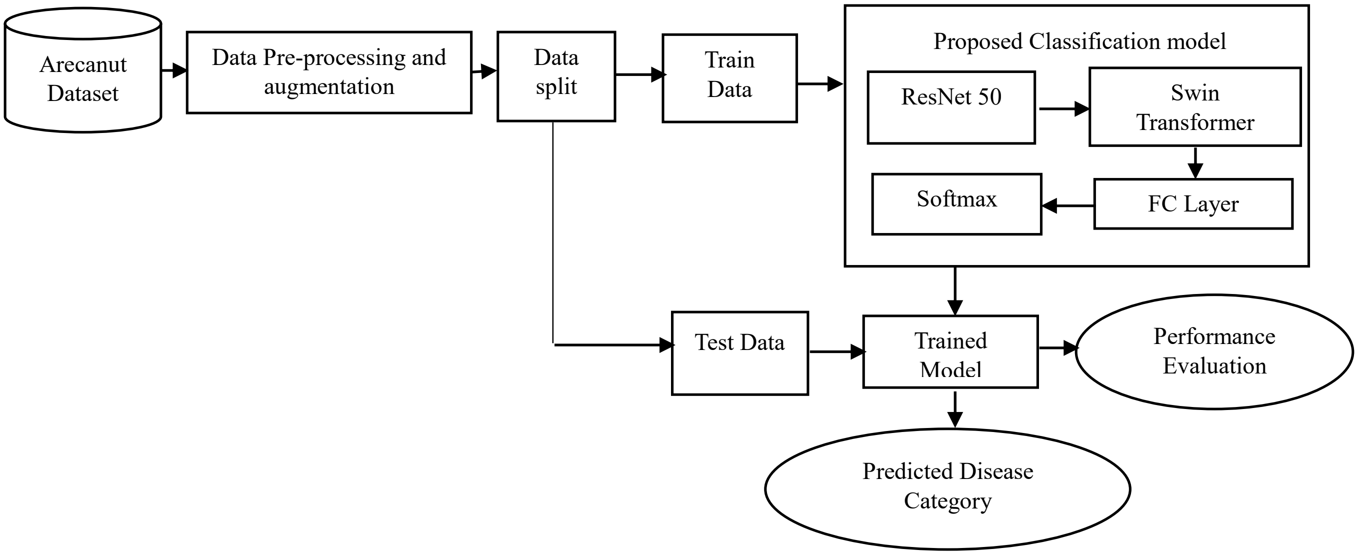
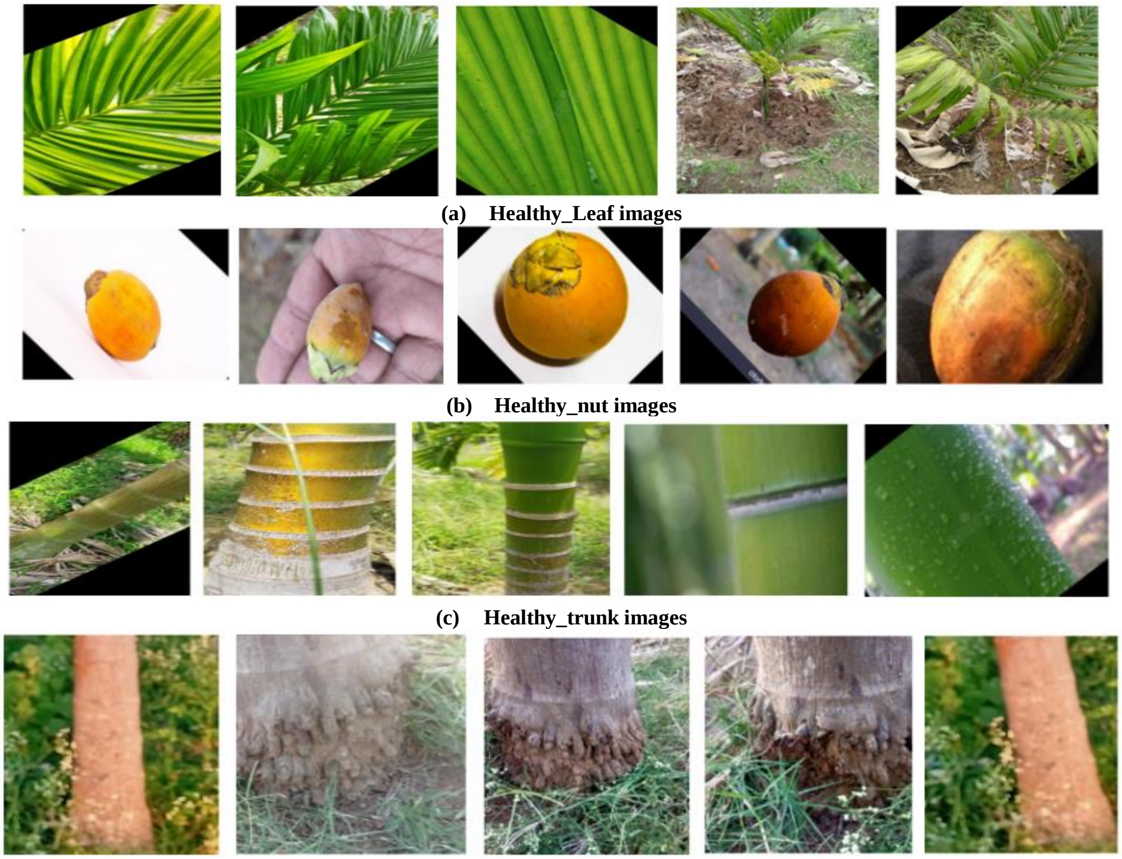


Fig. 1 Block diagram of the suggested hybrid framework

3.1. Dataset Description

The dataset was sourced from the Kaggle repository and provides a useful standard for identifying and categorising arecanut diseases [19]. The dataset consists of 11,063 images structured into two directories: training and testing. The training set includes 8,847 samples, and the testing set consists of 2,216 samples. Each directory consists of 9 different classes covering both healthy and diseased plant parts, whereas it includes Healthy_Leaf, Healthy_Nut, Healthy_Trunk, Healthy_Foot, Mahali_Koleroga, Stem_bleeding, bud_borer, stem cracking, and yellow leaf disease. Figure 2 (a) to (i) illustrates the sample images from each class.

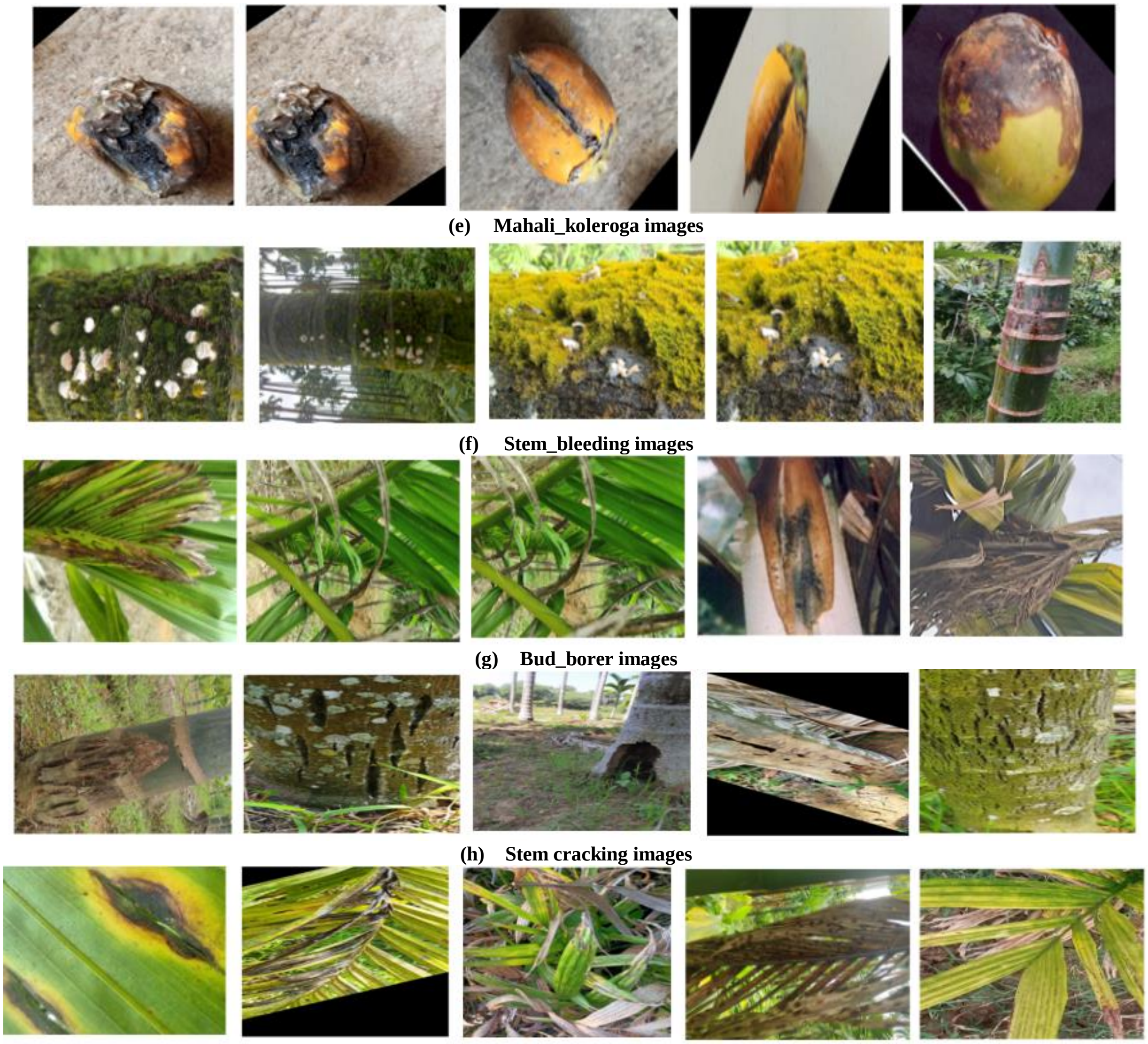


(a) Healthy_Leaf images

(b) Healthy_nut images

(c) Healthy_trunk images

(d) Healthy_foot images



(e) Mahali_koleroga images

(f) Stem_bleeding images

(g) Bud_borer images

(h) Stem cracking images

(i) yellow leaf disease images

Fig. 2 Sample images in the disease dataset

3.2. Data Preprocessing and Augmentation

Data preprocessing prepares raw images for training by standardising their format and improving their quality by applying image resizing, normalisation and class balancing. The dataset was standardised to maintain uniformity in feature representation by transforming its attributes to obtain a mean of 0 and a Standard Deviation (SD) of 1. The dataset was subsequently split into 80:20 ratios for training and testing. To improve the framework's generalisation capability and reduce overfitting, data augmentation methods like rotation, flipping, brightness modification, zooming, and noise addition were utilised.

3.3. Model Development

3.3.1. ResNet-50

ResNet-50 is a deep CNN formulated to solve vanishing gradients in deep networks. It is attained through residual learning, where the skip connections allow the gradients to flow directly through the network. Figure 3 presents the model architecture of ResNet-50, which comprises 50 layers, incorporating a convolutional layer, Batch Normalisation (BN), activation functions and Fully Connected layers (FC) [20]. It follows a structured design of one initial convolutional layer followed by four stages of residual blocks and a final classification layer.

Each residual block in ResNet 50 is based on the bottleneck design, where the three convolutional layers are used instead of two. The 1 x 1 convolutions reduce the dimensionality, while the 3 x 3 convolution captures the spatial information. The skip connection, illustrated in Figure 4, enables identity mapping for each block, which allows the network to focus on learning residual features.

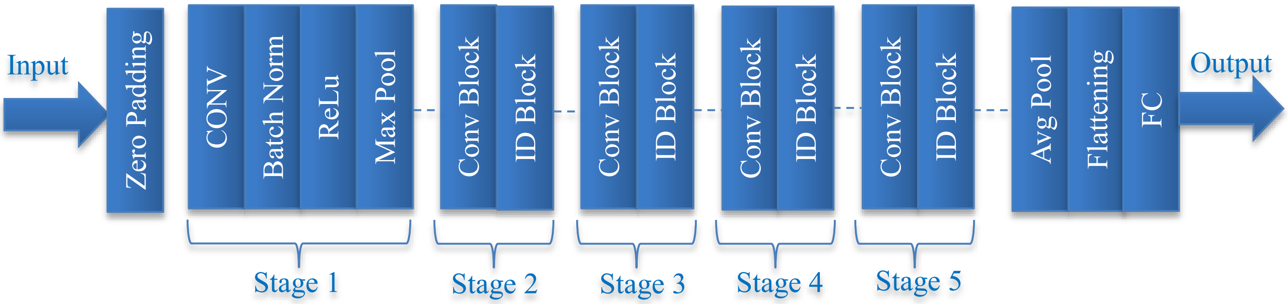


Fig. 3 ResNet 50 model architecture

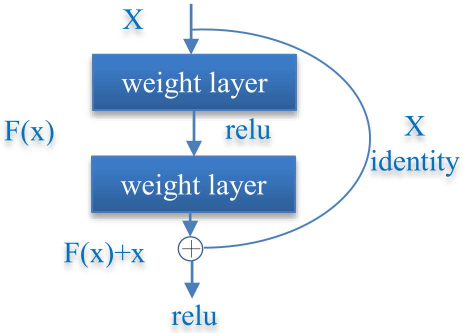


Fig. 4 Skip connection

The residual function is mathematically formulated, as shown in Equation (1).

$$y = F(x, W_i) + x \tag{1}$$

Where the input is represented by x , $F(x, W_i)$ It is the learned transformation function, and the sum y is the final output. This allows the model to learn complex hierarchical representations while maintaining gradient stability. ResNet - 50 ensures the edges, textures, color variations and disease symptoms are effectively captured from the input images. By utilising its residual connections, the model ensures stable learning even with the large dataset used in this study.

3.3.2. Swin Transformer

The Swin transformer is an advanced DL architecture designed for visual recognition tasks. It is an extension of the standard vision transformer that introduces hierarchical feature learning and local attention mechanisms. Similar to CNN, the Swin transformer processes images using patches by maintaining spatial relationships [21]. Figure 5 shows the basic architecture of the Swin transformer, and Figure 6 shows its computation process.

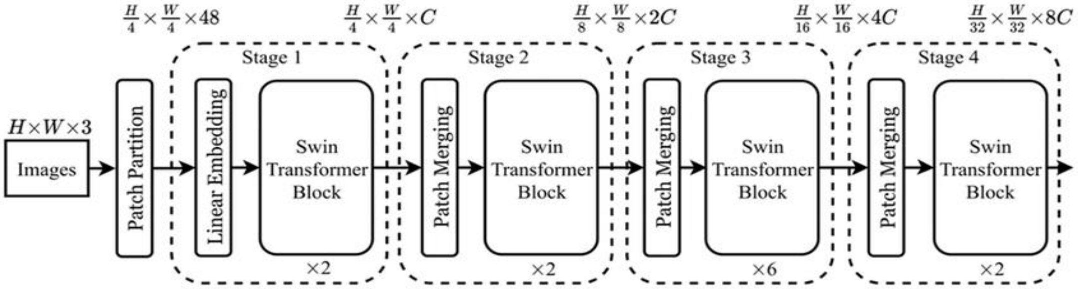


Fig. 5 Basic architecture of Swin transformer

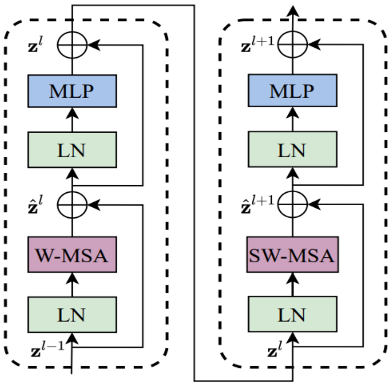


Fig. 6 Computation process of Swin transformer

The input feature map of size $H \times W \times C$ is reshaped into smaller overlapping windows. Each window holds $M \times M$ patches. This partitioning allows the Swin transformer to apply local self-attention inside each window separately rather than globally over the entire image. As shown in Equations (2), (3) and (4), the input feature map within each and every window is transformed to matrices query (Q), key (K), and value (V) using three projection matrices. $P_Q, P_K,$ and P_v.

$$Q = XP_Q \tag{2}$$

$$K = XP_K \tag{3}$$

$$V = XP_V \tag{4}$$

These projections are important for computing self-attention scores across different feature dimensions. Self-attention is calculated within each window using the scaled dot product attention formula presented in Equation (5).

$$Attention\ (Q,K,V) = Softmax\ \left(\frac{QK^T}{\sqrt{d}} + B\right)V \tag{5}$$

To reduce the impact of variance in the dot product, a scaling factor of $\sqrt{d}$ Is applied. Additionally, a relative positional bias, indicated by B is introduced before the softmax operation. By restricting attention to localised windows, the computational complexity associated with global attention is significantly reduced. Following the attention mechanism, the Multi-Layer Perceptron (MLP) layers improve the feature representations through FC transformations and nonlinear activation functions. Additionally, layer normalisation standardises the inputs at each layer, promoting stable training and ensuring a consistent gradient flow. This entire process is mathematically represented in Equation (6) and (7).

$$X_l = MLP\big(LN(X'_l)\big) + X'_l\ where\ l = 1,2, \ldots\ldots\ldots..L \tag{6}$$

$$X'_l = SW - MSA\ \big(LN(X_{l-1})\big) + X_{l-1}\ where\ l = 1,2, \ldots\ldots\ldots L \tag{7}$$

Where X_{l-1} serves as the input to the attention mechanism and $LN(X_{l-1})$ Stabilises the input. The Shifted Window Multi-Head Self-Attention ($SW - MSA$) processes the features within the local windowed regions to improve efficiency, while the residual connections ensure smooth information flow. The resulting feature representation X'_l undergoes further refinement, where layer normalisation $LN(X'_l)$ Improves training stability, the MLP applies fully connected transformation with nonlinear activations to refine features, and the residual connection maintains gradient flow, ultimately producing the refined output. X_l. Plant disease detection requires fine texture variations and subtle color differences, the ability of the Swin transformer to model local dependencies while preserving global context significantly improves classification performance. The model learns low-level patterns in the early stages and high-level patterns in deeper layers, making it efficient for detecting diseases.

3.3.3. Proposed Hybrid Model

The suggested hybrid framework for Arecanut disease recognition combines the capabilities of the ResNet-50-Swin transformer with an attention mechanism to improve classification accuracy. The framework processes input images of size 224 x 224 x 3 through ResNet-50, where the initial convolutional layers extract low-level spatial features while deeper layers capture complex textual patterns in diseased regions. A Global average pooling (GAP) is applied to decrease feature dimensionality, ensuring computational efficiency and preventing overfitting. After feature extraction, the Swin transformer encoder utilising $SW - MSA$ is employed to capture global dependencies and spatial correlations among various disease categories. The self-attention mechanism enables the framework to concentrate on the most significant areas of the image, improving disease localisation. The features that are extracted are refined by an FC layer with 128 units and a ReLU activation function. The disease is then predicted from the ten categories using the softmax classifier. The algorithm for the suggested hybrid framework is presented below.

```
Algorithm: Arecanut Disease Classification using Hybrid ResNet-50-Swin Transformer
Input: Arecanut disease images
Output: Arecanut disease classification model
Begin:
Load and preprocess data:
    1. Collect the dataset $D = \{(P_i, M_i)\}_{i=0}^{N-1}$, where $P_i$ Is the Arecanut disease image and $M_i \in \{0, 1, 2, …….,9\}$ (disease categories)
    2. Preprocess:
        • Resize: $P_i \rightarrow P_i' \in R^{150\times150}$
        • Normalise: $P_i' \rightarrow \frac{P_i' - \mu}{\sigma}$
        • Data Augmentation: $P_i' \rightarrow \{P_i''\}$ (random rotation, flipping, brightness adjustment, zooming, noise addition)
    3. Define ResNet-50 feature extractor:
       Input: $224 \times 224 \times 3$
       Block 1: Conv2D (64, (7,7), activation='relu')
                    MaxPooling2D (pool size= (3,3)
       Block 2: Residual Block (3× Bottleneck layers, 64)
       Block 3: Residual Block (4× Bottleneck layers, 128)
       Block 4: Residual Block (6× Bottleneck layers, 256)
```

```
        Block 5: Residual Block (3× Bottleneck layers, 512)
        Flatten ()
        Dense (512, activation=’relu')
        Dropout (0.5)
    4.  Define Swin Transformer Model:
        Input: Features from ResNet-50
        Patch Partitioning
        Transform patches into feature vectors
        Swin Transformer Blocks:
        Self-Attention ()
        Shifted window mechanism
        Layer Normalisation
        MLP
        Flatten ()
        Dense (512, activation=’relu')
        Dropout (0.5)
    5.  Concatenate (ResNet-50, Swin Transformer)
        Dense (512, activation=’relu')
        Dropout (0.5)
        Dense (10, activation= 'softmax')
    6.  Model Compilation and Training:
    a.  Compile each model P:
                optimizer=Adam ()
                learning rate= 0.001
                loss=sparse_ categorical _crossentropy
                metrics=[accuracy]
    b.  Train: P.fit ($P_{train}$ , $M_{train}$ ,validation_data= ($P_{val}$, $M_{val}$), batch size= (64), epochs= (20)
    7.  Model Evaluation:
    a.  Evaluate:
              metrics=P.evaluate($P_{test}$ , $M_{test}$), where metrics contain accuracy recall precision.
Save the Model:
End
```

Table. 1. Hyperparameters for the proposed hybrid model

Hyperparameters	Values
Learning rate	0.001
Activation Function	ReLU, Softmax
Optimizer	Adam
Loss Function	Categorical Crossentropy
Number of Epochs	20
Batch Size	64

3.4. Software and Hardware Setup

The proposed system was tested and trained on Google Colaboratory, utilising Python along with the Keras library, leveraging the platform's GPU acceleration, 12.75 GB of RAM and 68.50 GB of available storage. TensorFlow support within the Colab environment facilitates efficient DL computations in the system. The system configuration comprised a 64-bit Windows 10 operating system, an Intel Core i7-6850K processor running at 3.60GHz with 12 cores and an NVIDIA GeForce GTX 1080 Ti GPU equipped with 2760 MB memory, ensuring excellent computing performance. The framework's performance was assessed using predictions derived from the test dataset. Hyperparameters are predefined variables set before the learning process begins, controlling and optimising the model's training behavior to improve performance. Table 1 outlines the Hyperparameters of the proposed hybrid framework.

4. Results and Discussion

An accuracy plot is a visual representation that depicts the changes in training and validation accuracy over successive epochs, providing information about the model's learning progression. A loss plot visually depicts the variation in training and validation loss throughout each epoch, offering information into the model's convergence behavior, and the presence of overfitting. The efficiency of the suggested framework, based on accuracy and loss over 20 epochs, are presented in Figures 7 and 8. The training accuracy at the initial epoch starts approximately at 0.84, whereas the validation accuracy is about 0.85, indicating that the framework starts with low classification performance. As the training progresses, both the training and validation accuracy improve learning. By the final epoch, the training reaches around 0.96, and validation reaches around 0.98, demonstrating that the framework generalises effectively to unknown data with minimal overfitting.

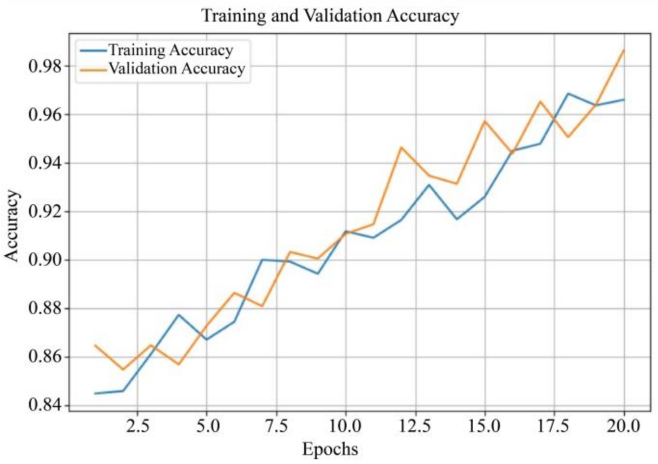


Fig. 7 Accuracy plot of the proposed model

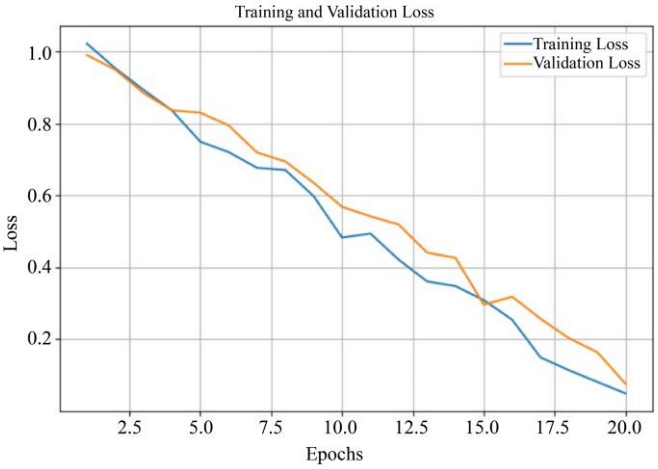


Fig. 8 Loss plot of the proposed model

Regarding the loss of the framework, at the initial epoch, the value of training loss is almost 1.0 and validation loss is also around 1.0, indicating a high error rate. As the epoch increases, the losses steadily decline, indicating effective optimisation. By the final epoch, the training loss is reduced to about 0.1, and the validation loss is at about 0.15, demonstrating that the model has effectively reduced the errors while maintaining a good generalisation capability.

Evaluation metrics examine the effectiveness of the proposed DL model by offering information about its predictive accuracy and classification performance. Metrics offer thorough information on the strength of the model and potential areas for improvement. These measures are valuable in overcoming problems such as overfitting, class imbalance and underfitting, hence maintaining the framework's resilience and reliability. The mathematical formulations for these metrics are presented in Equation (8) to (11).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

$$Precision = \frac{TP}{TP+FP} \tag{9}$$

$$Recall = \frac{TP}{TP+FN} \tag{10}$$

$$F1-score = 2 \times \frac{precision \times Recall}{Precision+Recall} \tag{11}$$

Where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

The overall performance of the proposed hybrid framework across multiple evaluation metrics is demonstrated in Figure 9. The framework attains 98.42% accuracy, reflecting its efficacy in making accurate predictions. A precision of 98.45% indicates its ability to accurately classify positive instances. The recall measured at 98.38% indicates the capability of the framework to extract actual positive cases, reducing the probability of missing instances. The F1 score of 98.40% balances the precision and recall, ensuring a stable and well-rounded classification performance. These metrics signify consistency and reliability across various evaluation criteria, making it highly effective for the given classification task.

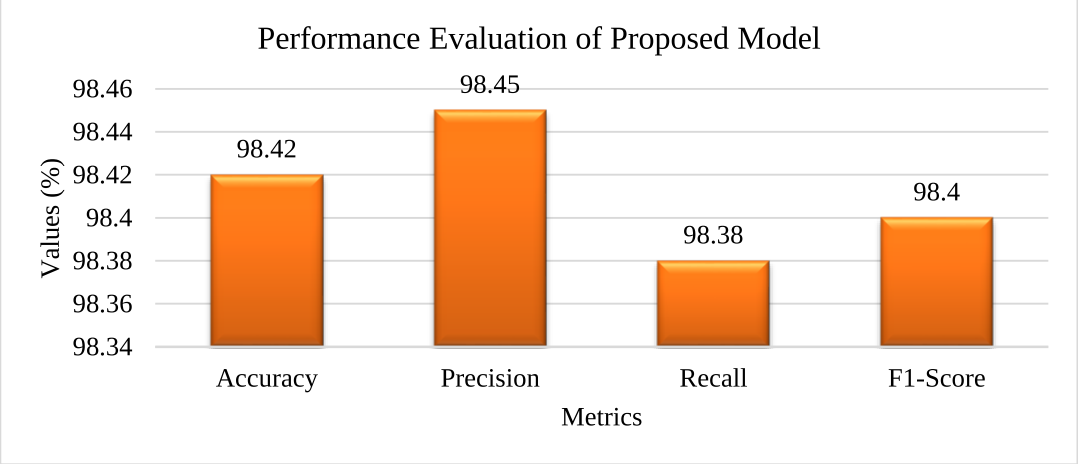


Fig. 9 Performance evaluation of the proposed hybrid model with an attention mechanism

The confusion matrix enables a complete assessment of the framework's categorisation by illustrating the relationship between actual and predicted labels. Figure 10 presents a detailed examination of the proposed hybrid framework across different categories of Arecanut disease conditions. The diagonal elements represent correctly classified instances, indicating strong classification accuracy for all categories. The model exhibits high accuracy in identifying Mahali Koleroga (627), Healthy Nut (465) and Yellow Leaf Disease (368) by the high values along the diagonal. Minimal misclassifications are observed in off-diagonal elements, signifying that the model effectively differentiates between healthy and diseased samples with minimal FP and FN.

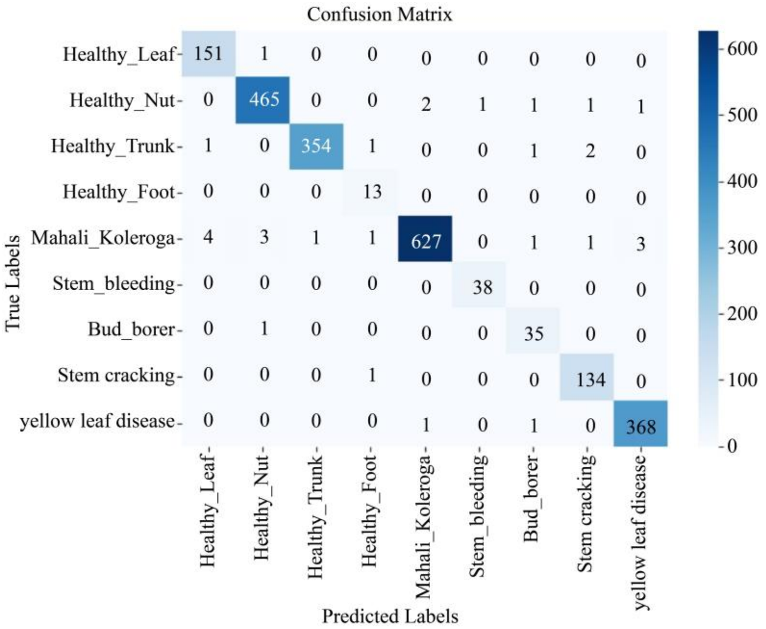
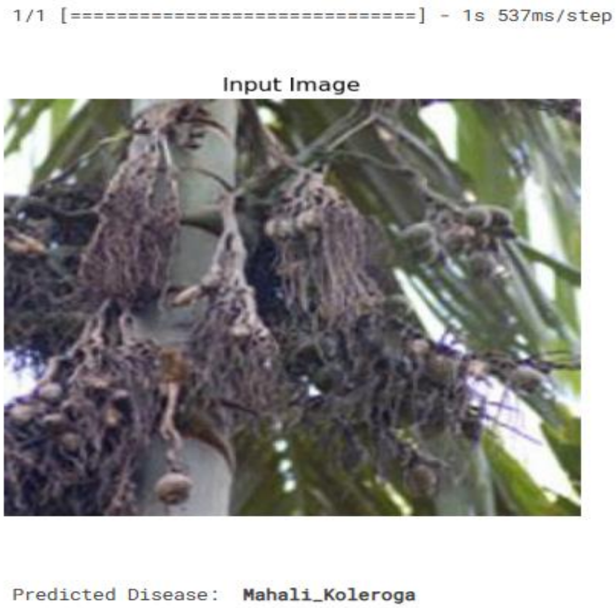


Fig. 10 Confusion matrix of the proposed hybrid model

An image selected at random from the Arecanut disease dataset is categorised by the proposed hybrid model, correctly identifying it as either a healthy or diseased category. This accurate classification, as illustrated in the corresponding Figure 11, highlights the reliability and effectiveness of the model in differentiating between different Arecanut disease conditions within the dataset.



```
1/1 [==============================] - 1s 537ms/step
```

```
Predicted Disease:  Mahali_Koleroga
```

Fig. 11 Predicted output

Table 2 provides a comparative overview of various models used for Arecanut disease classification, which indicates that the proposed hybrid model with attention mechanism attained an accuracy of 98.42%, outperforming all the other methods. The traditional CNN model shows 81.35% accuracy, which struggles with feature extraction limitations. At the same time, VGGNet (92%) and Inception V3 (96%) lacked the advanced attention mechanism. Although the RF (95.5%) and QCNN (97.72%) models perform well, their high computation cost makes them less practical. The proposed hybrid model excels in feature extraction and utilising transformer-based global attention, allowing precise classification across disease patterns with minimal misclassifications and making it the most effective solution. Figure 12 illustrates the accuracy of comparing the suggested hybrid framework and existing techniques.

Table 2. Accuracy comparison of the proposed hybrid model with existing approaches

Author & Ref	Model	Dataset	Accuracy (%)
Jenitta & Swetha [15]	CNN	Arecanut images	81.35
Kumar et al. [6]	VGGNet	Arecanut leaf images	92
Xu et al. [17]	RF	UAV and thermal infrared data	95.5
Ghate et al. [4]	Inception V3	Arecanut images	96
Naik & Rudra [7]	QCNN	Arecanut X-ray images	97.72
Proposed hybrid ResNet 50 + Swin Transformer Model		**Arecanut disease dataset**	**98.42**

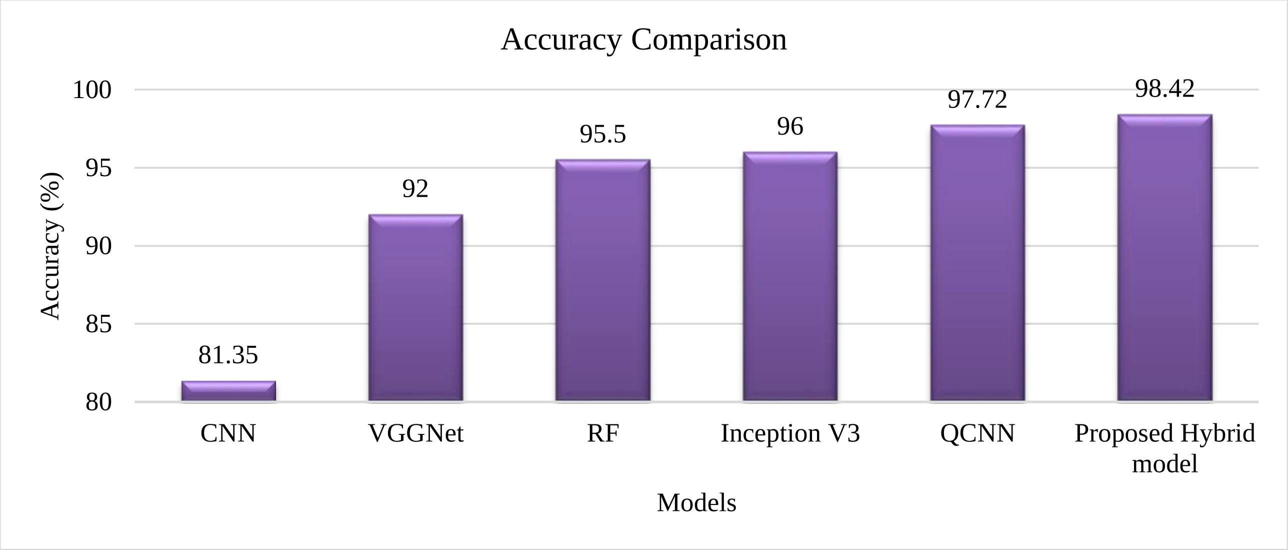


Fig. 12 Accuracy comparison of proposed hybrid model with existing approaches

5. Conclusion

This research suggests a hybrid DL framework integrating ResNet-50 and Swin Transformer for accurate and automated detection of Arecanut diseases. The dataset for the research is sourced from Kaggle and comprises 11,063 images covering both healthy and diseased plant parts. The suggested framework obtained an accuracy of 98.42%, surpassing conventional CNN-based models. Utilising ResNet-50 for feature extraction and Swin Transformer for capturing long-range dependencies, the model significantly improves classification performance. The study demonstrates that integrating transformer mechanisms into plant disease detection enhances robustness and generalisation, allowing it to be greatly effective for practical agricultural applications. In future work, current research can be expanded by incorporating multispectral or hyperspectral imaging to enhance disease identification, developing lightweight models for mobile-based applications, and combining the system into a smart agricultural framework for real-time disease monitoring.

Acknowledgements

The author sincerely thanks the supervisor for guidance and continuous support throughout the study.

References

[1] Shabari Shedthi Billadi et al., “Classification of Arecanut using Machine Learning Techniques,” *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, pp. 1-8, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Shuhan Lei et al., “Remote Sensing Detecting of Yellow Leaf Disease of Arecanut based on UAV Multisource Sensors,” *Remote Sensing*, vol. 13, no. 22, pp. 1-22, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] M.G. Anilkumar et al., “Detection of Diseases in Areca Nut using Convolutional Neural Networks,” *International Research Journal of Engineering and Technology*, vol. 8, no. 5, pp. 4282-4286, 2021. [Google Scholar] [Publisher Link]

[4] Dhanush Ghate et al., “Enhancing Arecanut Quality Grading: A Comparison of Custom CNNs and Transfer Learning Models,” *Researchsquare*, pp. 1-27, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[5] M. Pavan et al., “Areca Nut Disease Detection using Deep Learning,” *Journal of Computer Science*, vol. 18, no. 1, pp. 1-7, 2025. [Google Scholar] [Publisher Link]

[6] S. Anupama Kumar et al., “An Automated Deep Learning Model to Classify Diseases in Arecanut Plant,” *Indian Journal of Computer Science and Engineering*, vol. 15, no. 1, pp. 43-53, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Praveen M. Naik, and Bhawana Rudra, “Quantum-Inspired Arecanut X-Ray Image Classification using Transfer Learning,” *IET Quantum Communication*, vol. 5, no. 4, pp. 303-309, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Khairunnisa, and Ferdy Riza, “Android-Based Areca Plant Disease Detection Using Convolutional Neural Network (CNN) Algorithm,” *Install: Computer Journal*, vol. 16, no. 3, pp. 277-288, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Anitha Arekattedoddi Chikkalingaiah et al., “Segmentation and Yield Count of an Arecanut Bunch Using Deep Learning Techniques,” *IAES International Journal of Artificial Intelligence*, vol. 13, no. 1, pp. 542-553, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Rajashree Krishna, and K.V. Prema, “Constructing and Optimizing RNN Models to Predict Fruit Rot Disease Incidence in Areca Nut Crop based on Weather Parameters,” *IEEE Access*, vol. 11, pp. 110582-110595, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Sameer Patil, Aparajita Naik, and Jivan Parab, “Efficient Deep Learning Model for De-Husked Areca Nut Classification,” *Applied and Natural Science*, vol. 15, no. 4, pp. 1529-1540, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Ajit Hegde et al., “Identification and Categorization of Diseases in Arecanut: A Machine Learning Approach,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 3, pp. 1803-1810, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Praveen M. Naik, and Bhawana Rudra, “Classification of Arecanut X-Ray Images for Quality Assessment using Adaptive Genetic Algorithm and Deep Learning,” *IEEE Access*, vol. 11, pp. 127619-127636, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Namra Mahveen et al., “Enhancing Areca Nut Plant Wellness: Innovative Disease Detection using Deep Learning Algorithms,” *3rd Indian International Conference on Industrial Engineering and Operations Management*, New Delhi, India, pp. 1-8, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] P. Pallavi, K. Sowmya Nag, and S.V. Veenadevi, *Arecanut Disease Classification using CNN*, Advances in Intelligent Systems and Technologies, pp. 19-24, 2023. [CrossRef] [Publisher Link]

[16] Sameer Patil et al., “Optimizing Dehusked Arecanut Quality Segregation: CNN-based Approach with Contrast Enhancement and Data Augmentation,” *International Symposium on Smart Cities, Challenges, Technologies and Trends*, pp. 1-10, 2023. [Google Scholar] [Publisher Link]

[17] Dong Xu et al., “Areca Yellow Leaf Disease Severity Monitoring using UAV-based Multispectral and Thermal Infrared Imagery,” *Remote Sensing*, vol. 15, no. 12, pp. 1-19, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Patil Balanagouda et al., “Timing of Oomycete-Specific Fungicide Application Impacts the Efficacy against Fruit Rot Disease in Arecanut,” *Frontiers in Plant Science*, vol. 14, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] T. Paviraj, Arecanut Dataset, Kaggle. [Online]. Available: https://www.kaggle.com/datasets/tejpaviraj/arecanut/data

[20] Bishwas Mandal, Adaeze Okeukwu, and Yihong Theis, “Masked Face Recognition using ResNet-50,” *arXiv*, pp. 1-8, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[21] Z. Liu et al., “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, pp. 9992-10022, 2021. [CrossRef] [Google Scholar] [Publisher Link]