Proceedings of the International Conference on Inventive Computation Technologies (ICICT-2025)
DVD Part Number: CFP25F70-DVD; ISBN: 979-8-3315-1223-1

An Innovative Hybrid Feature Extraction Method for the Diagnosis of Coconut Leaf Diseases

S.Nithya priya
Research Scholar
Department of computer science
Vels Institute of Science, Technology & Advanced Studies
Chennai, Tamil Nadu, India.
priyasnithya475@gmail.com

L.Ramesh
Assistant professor
Department of Computer Applications (UG)
Vels Institute of Science, Technology & Advanced Studies
Chennai, Tamil Nadu, India.
lramesh.scs@vistas.ac.in

***Abstract*— In South India, cultivating coconuts is a major agricultural activity, but the trees are challenged by unfavorable weather conditions and environmental factors. These difficulties comprise pest infestations and a variety of leaf diseases. Because of the coconut trees' enormous foliage and shadowing, it might be challenging to identify and locate these problems. The study proposes a feature extraction strategy which integrates the feature vectors of deep learning (DL) and handcrafted features into a single, cohesive representation for detecting the coconut leaf diseases. The selected features were fed into the Machine learning (ML) classifier like Support Vector Machine (SVM), Random Forest (RF) and XGBoost. The experimental results show that the proposed technique outperforms with an accuracy of 97.4 % showing the importance of feature extraction techniques in identifying the coconut leaf diseases.**

Keywords— : ***Coconut leaf disease detection, Feature Extraction. Handcrafted features, Deep Features, Classifiers***

I. INTRODUCTION

Smart agricultural initiatives have grown vital in recent years to fulfill the world's growing food demands. The need for food is rising dramatically, despite the fact that the amount of land accessible for agriculture is decreasing daily. The growing demand for food production is also influenced by both natural and man-made events. The coconut (Cocos nucifera L.) is a palm plantation that benefits people from the fruit to the trunks. India is actually the world's third-largest producer of coconuts (Maray et al.,2022). Coconut crops are extremely valuable since they are used in a variety of industries for both food and non-edible parts. These crops are susceptible to diseases such as bud rot, bud root dropping, leaf yellowing, CCI caterpillars, grey leaf spot and stem bleeding(Karegowda et al., 2024). A family's livelihood that is entirely dependent on the coconut economy and the corresponding industrial units are negatively impacted by the diseases that influence the productivity of coconut plantations. Root (wilt) disease (RWD), stem bleeding, Ganoderma-Basal Stem Rot (BSR), leaf blight, leaf rot, and bud rot are among the common diseases that affect coconut trees. Termites, black-headed caterpillars, rhinoceros beetles, coconut eriophyid mites, and red palm weevils are among the pests that live in coconut trees.

. In this field, artificial intelligence (AI) methods like CNNs are quite popular since they provide the best alternatives for detecting diseases in coconut trees. Accurate and real-time plant disease detection methods aid in the creation of disease prevention plans and ensure adequate nutrition on a broad scale, which is further improved by small-scale, profitable crop protection(Bouthaina,et al., 2024). Computational intelligence or soft computing approaches combined with image processing techniques will assist farmers diagnose diseases more efficiently. A subset of feature engineering is feature extraction. In situations where raw data is meaningless, data scientists turn to feature extraction. Image files are a common use case for feature extraction, which converts unprocessed data into numerical features that ML systems may exploit. By extracting an object's shape or an image's redness value, data scientists can produce new features that are appropriate for ML applications.

Feature extraction improves ML models' precision and effectiveness by simplifying the training data. In order for models to concentrate on the most pertinent aspects, it first eliminates redundant data by removing noise and extraneous information. This immediately improves model accuracy since training on only the most important data keeps the model focused on its goal and prevents unnecessary information from having a detrimental effect. Moreover, feature extraction accelerates learning because models trained on streamlined datasets converge more quickly and produce predictions that are more correct. Lastly, by removing unnecessary data, cutting down on processing overhead, and guaranteeing that processing power is focused on important activities, it results in a more effective use of computational resources. Feature extraction is an important phase in maximizing ML performance because of these advantages. Further, the plant diseases can be identified by signs and symptoms which are taken from the plant and represented as features. Thus, feature extraction methods are essential in these types of algorithms(Kavitha et al., 2024).

Numerous research uses customized datasets, which are frequently obtained from small-scale collections or Kaggle. These datasets might be biased and have less generalizability if they don't accurately reflect real-world circumstances. The majority of studies is restricted to particular coconut types and environmental circumstances. To test these models across a variety of geographical regions and coconut species, more research is needed. Numerous models are frequently investigated in studies; however the benefits and disadvantages between DL (CNN, YOLO) and classical ML (SVM, RF, KNN) are not thoroughly compared. Hybrid models that combine the advantages of both strategies require further study. The main objectives of the study is to

- Propose a hybrid feature extraction and fusion strategy to combine handcrafted features like textures, color and shape and deep features extracted using a pre-trained model CNN-ResNet50
- Evaluate the performance of the proposed feature extraction technique by employing the ML classifiers like SVM, RF and XGBoost.

The remaining portion of the document is structured as follows. An overview of the relevant literature is provided in Section 2, and the suggested model is presented in Section 3. The performance validation of the suggested model is then covered in Section 4, and the work is concluded in Section 5..

II. LITERATURE STUDY

Plant disease identification has advanced significantly in recent years because of the use of ML and DL approaches. Subbaian et al., (2024) suggests a model to identify and locate infections and diseases in coconut leaves from images, the YOLOv4 algorithm was used. Advanced features including path-based aggregation, contextual feature selection, spatial pyramid pooling, and cross-stage partial connections are all included in the YOLOv4 model. These characteristics improve the model's capacity to effectively detect problems in coconut leaf images captured under many environmental circumstances, including insect infestations, leaf flaccidity, and leaf yellowing and drying.

Conventional ML models are implemented with GLCM, SIFT, and a combination of both. Four ML RF, LR, SVM, and KNNwere tested using a customized dataset from Kaggle. By combining GLCM+SIFT with SIFT alone, the RF was able to attain 100% accuracy. SVM reached 92.8% accuracy with SIFT alone, whereas LR and KNN had 100% accuracy. In comparison to GLCM characteristics, SIFT features turned out to be preferable. Farmers, legislators, and other agricultural stakeholders gain from this solution's rapid identification and categorization of coconut diseases, which promotes improved disease management and lowers crop damage Karegowda et al., 2024).

Arrahimi et al.,(2024) proposes a study with the use of PCA to modify the method of extraction features and compare it with no PCA for unhealthy and healthy types. The classification will then be carried out using SVM using different treatments, such as changing the features utilized and the volume of data that needs to be processed in order to conduct tests or experiments.

Popular segmentation techniques were used to segment a collection of carefully selected images of healthy and diseased coconut trees to quickly identify the irregular boundaries. To anticipate diseases and pest infections, a deep 2D CNN that has been specially constructed is trained. Additionally, using the inductive transfer learning technique, the most advanced Keras pre-trained CNN models VGG16, VGG19, InceptionV3, DenseNet201, MobileNet, Xception, InceptionResNetV2, and NASNetMobile were refined to categorize the images as either healthy or infected(Singh et al., 2021).

Barman et al., (2023) employs two Artificial Neural Network (ANN) architectures to classify the extracted data features of the damages into five different classes, such as Eriophyid Mite, Rhinoceros Beetle, Red Palm Weevil, Rugose_Spiraling_White_fly, and Rugose_in_Mature, with an average testing accuracy of nearly 100% each. The texture features of the damages are extracted using the GLCM and Gray Level Run Length Matrix (GLRLM) techniques. The SVM, Decision Tree (DT), and Naïve Bayes (NB) are also introduced for damage diagnosis to compare the results with the other ML approaches. The SVM method reports nearly 100% accuracy on the fuse characteristics of GLCM and GLRLM.

Sarvas et al., (2024) provides a thorough analysis of the use of DL models in smart agriculture for the identification and categorization of palm diseases. The goal of the study is to increase the results' generalizability and robustness while addressing the shortcomings noted in earlier research. With all models attaining excellent accuracy rates throughout training and validation, the trials demonstrate encouraging outcomes. The study offers insightful information for further investigation, directing the use of DL approaches to tackle important agricultural issues and enhance crop health monitoring systems. Combining different kinds of insect pests and plant diseases using different databases is another contribution. The model's robustness is increased by taking into account various disease categories and stages within the white scale group, resulting in a thorough categorization system.

Bouthiana et al., (2024) explores the most recent developments in the identification of palm tree diseases using deep learning (DL). CNNs, in particular, have been widely employed in agricultural research because of their powerful image processing capabilities..

III.METHODOLOGY

. The methodology integrates the DL features (CNN-based embeddings) with handcrafted features like color, texture, and shape to improve the detection accuracy of coconut leaf diseases. Data Acquisition (Image), Image preprocessing, feature extraction, feature fusion, classification, and evaluation are the main steps in the process. The workflow for the fused features coconut leaf disease prediction is depicted in figure 1.

Image Acquisition : Drones, agricultural imaging devices, or smartphone cameras are used to take images of the coconut leaves. It gathers samples of both healthy and diseased leaves under various environmental circumstances. Images are stored in a structured dataset that has labels, such as blight, wilt, leaf spot, and healthy.

Preprocessing Images: The images are downsized to their usual dimensions, which for DL models are 224 by 224 pixels. Unwanted noise is eliminated by the use of Gaussian filtering. To improve feature detection, normalization converts images to various color spaces (RGB, HSV, or CIELAB). Otsu's thresholding is used in segmentation to identify and isolate diseased areas(Sharma et al., (2024)..Data augmentation creates more training data by adjusting contrast, flipping, rotation, and zooming.

Hybrid Feature Extraction: In this step, an increased comprehensive feature set is created by extracting deep and handcrafted features.

Handcrafted Feature Extraction: This involves extracting features from images that are based on color, texture, and shape. The RGB, GSV, and CIELAB histograms are retrieved in color-based feature extraction to identify any discoloration in the diseased regions. Next, determine the color distribution by computing color moments such as mean, variance, and skewness. To extract texture-based information, the Gray level co-occurrence matrix (GLCM) is used for contrast, correlation, and homogeneity. Leaf surface microtexture investigation is done using the Local Binary Patterns (LBP). The multi-scale texture details are then captured using the wavelet transform. The leaf lesion region, perimeter, and aspect ratio are extracted using shape-based characteristics to identify different diseased symptoms(Faisal et al., 2024).

Deep Learning Feature Extraction: In DL-based feature extraction, pertinent features are automatically extracted from leaf images using a pretrained CNN. These deep features can be utilized for classification because they capture complex patterns such lesion forms, textural modifications, and disease-specific discoloration.

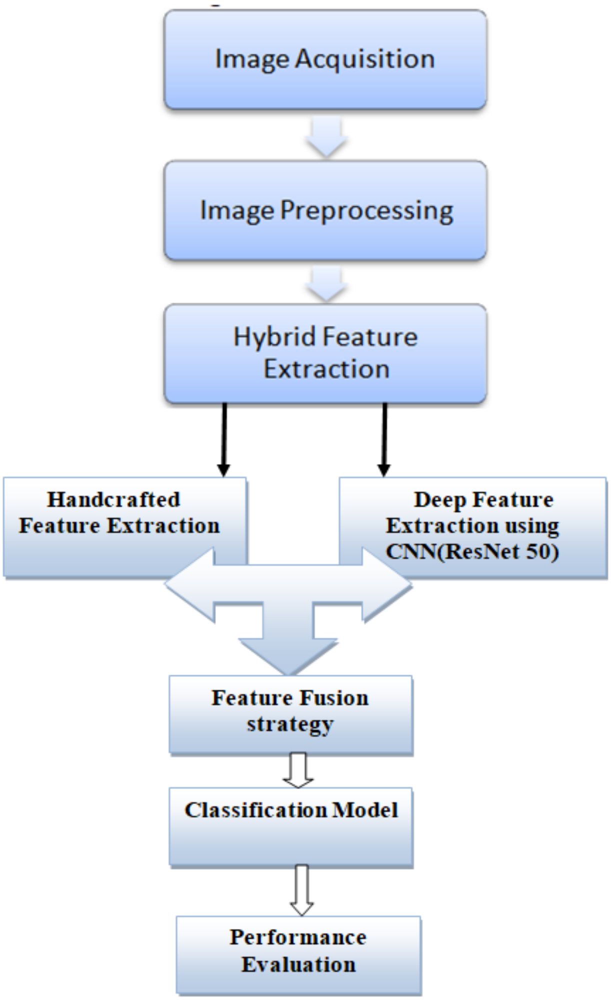

Figure 1: Workflow -DL based Coconut Leaf Disease Prediction

In this study, the huge datasets (such as ImageNet) are trained using a pretrained model called ResNet50(Tuhin et al., 2024).. This model does not require retraining the entire network, making it suitable for feature extraction or fine-tuning. ResNet50 uses residual connections to capture more complex and deeper patterns. Fully connected (FC) layers occur following a number of convolutional layers in CNN models. The FC layers produce high-dimensional feature vectors, whereas the convolutional layers retrieve spatial characteristics. The overall layers in this deep feature extraction strategy is depicted in figure 2

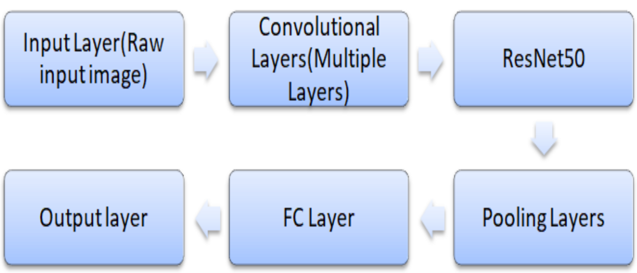

Figure 2 : Layers in Deep Feature Extraction Strategy

The Final Classification Layer must be removed to extract features. For classification, CNN models have a final softmax or dense layer (ImageNet has 1000 classes, for example). Since it only requires the deep features and not the classification output, this layer is eliminated. Next, extract features from either a convolutional layer or the FC layer. A compressed feature representation of the image can be found in the final FC layer. Alternatively, the Global Average Pooling (GAP) layer, which offers spatially aware feature embeddings, can be used to extract features. Multiple convolutional layers precede feature vector representation in CNN models. The size of the retrieved feature vector usually depends on the model used, and can be 512, 1024, or 2048. The overall layers in this deep feature extraction strategy is depicted in figure 2

Feature Fusion strategy : Feature Concatenation, a straightforward fusion technique, is used to merge deep and handcrafted features into a single feature vector

$$F_{final}=[F_{deep},F_{handcrafted}] \qquad (1)$$

Then used a Z-score Standardization to normalize characteristics.

Classification Model : The fused features are used to train machine learning classifiers like SVM, random Forest and XGBoost. As an alternative, optimize a fully connected neural network for the ultimate classification of diseases.

IV. Results and Discussions

The experiments are conducted in python with an appropriate dataset and results are observed and analysed.

A. Dataset Description

A specialized set of data used for the diagnosis and analysis of diseases and pest attacks affecting coconut plants is the "Coconut diseases and Pest Infestations" dataset. This dataset can be downloaded from Kaggle. Images showing coconut leaves, stems, and fruits affected with a variety of diseases, including bud rot, leaf blight, stem bleeding, and root wilts, as well as pest infestations from animals like rhinoceros beetles and red palm weevils, are commonly included in this dataset. The dataset may include photos as well as metadata, including geographic locations, environmental conditions, and expert-labeled remarks to improve classification accuracy (Kaggle)

B. Performance Evaluation

The performance is analyzed to evaluate the performance of the ML models, using the confusion matrix, accuracy, precision, recall, and F1-score(Alam et al., 2024). Table 1(a) and Table 1(b) shows the performance analysis of the ML classifiers like SVM, Random Forest, and XGBoost with and without proposed feature extraction technique using the assessment metrics: F1-score, Accuracy, Precision, and Recall.

Table 1(a) Performance Analysis: Hybrid feature Extraction

Classifier	Accuracy (%)		Precision (%)	
	With Feature Extraction	Without Feature Extraction	With Feature Extraction	Without Feature Extraction
SVM	92.5	87.5	91.2	86.1
Random Forest	93.8	90.2	93.7	90.4
XGBoost	97.4	93.1	95.8	92.3

Table 1(b) Performance Analysis : Hybrid feature Extraction

Classifier	Recall (%)		F1-Score (%)	
	With Feature Extraction	Without Feature Extraction	With Feature Extraction	Without Feature Extraction
SVM	92.2	86.5	92.0	86.0
Random Forest	94.6	90.1	94.4	89.7
XGBoost	96.5	92.5	96.7	92.6

XGBoost's robust gradient boosting framework allows it to achieve the highest accuracy of 96.3% and F1-score of 95.9%, making it the most effective. Random Forest does well, using many decision trees to get an accuracy of 94.5%. SVM offers strong precision and recall despite performing marginally worse (92.1% accuracy). The accuracy of SVM is the lowest, both with and without feature extraction (92.5% and 87.5%, respectively). Further, it routinely performs worse than RF and XGBoost in terms of precision, recall, and F1-score.Its accuracy increases from 90.2% to 93.8% after feature extraction, demonstrating an increase in efficiency. The performance of the ML methods (SVM,RF and XGBoost) in two scenarios with and without feature extraction is shown in Figure 3.

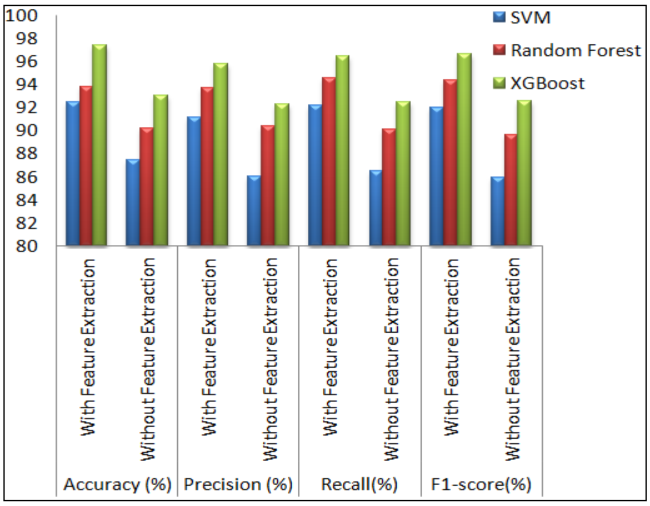

Figure 3. Performance Analysis- Proposed Feature Extraction method with ML models

In comparison to training them without feature extraction, nearly all classifiers that use feature extraction have higher accuracy, precision, recall, and F1-score. This implies that feature extraction facilitates in collecting more appropriate data and improves the classification accuracy. The most efficient classifier is XGBoost, particularly when feature extraction is used. Across all classifiers, feature extraction greatly improves model performance. SVM may not be the ideal option for this task because it performs least well

V . Conclusion

The study uses handcrafted features and deep features for the selecting the most pertinent features from the raw data. It uses CNN-ResNet50 for extracting the features and the performance of the proposed strategy is assessed by employing the classification algorithms like SVM, Rf and XGBoost to find the efficacy of the feature extraction method. The findings show that the XGBoost with the proposed feature extraction strategy outperforms with an accuracy of 97.4% than the contrasted methods. In future study the comparison among the various pretrained CNN models can be studied to provide valuable insights.

Moreover, instead of extracting information from a single layer, multiple layers will improve the accuracy of the model. The study is dependent on well captured images and this limits its applicability to real-world variances like lighting and shadow. In addition, the model has not been verified in actual field conditions and is not adaptive to new diseases without retraining. Future research should focus on developing feature extraction methods to further increase the model efficiency and ensuring practical applicability through dataset expansion, multimodal learning, and explainability.

References

[1] Maray, M., Albraikan, A. A., Alotaibi, S. S., Alabdan, R., Al Duhayyim, M., & Al-Azzawi, W. K. (2022). Artificial intelligence-enabled coconut tree disease detection and classification model for smart agriculture. Computers and Electrical Engineering, 104, 108399.

[2] A. G. Karegowda, V. D and Y. Harshalatha, "Coconut Disease Detection using Traditional ML Classifiers with GLCM and SIFT Features," *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*, Bengaluru, India, 2024, pp. 1-6, doi: 10.1109/CSITSS64042.2024.10816784.

[3] Subbaian, S., Balasubramanian, A., Marimuthu, M., Chandrasekaran, S., & Muthusaravanan, G. (2024). Detection of coconut leaf diseases using enhanced deep learning techniques. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-13.

[4] Arrahimi, A. R., Julianto, V., & Rahmanto, O. (2024). Machine learning to Detect Palm Oil Diseases Based on Leaf Extraction Features and Principal Component Analysis (PCA). Klik-Kumpulan Jurnal Ilmu Komputer, 11(1), 26-36.

[5] Singh, Piyush & Verma, Abhishek & Alex, John Sahaya Rani. (2021). Disease and pest infection detection in coconut tree through deep learning techniques. Computers and Electronics in Agriculture. 182. 105986. 10.1016/j.compag.2021.105986.

[6] Vishnoi, V.K., Kumar, K. & Kumar, B. A comprehensive study of feature extraction techniques for plant leaf disease detection. Multimedia Tools and Applications 81, 367–419 (2022). https://doi.org/10.1007/s11042-021-11375-0

[7] Barman, U., Pathak, C. & Mazumder, N.K. Comparative assessment of Pest damage identification of coconut plant using damage texture and color analysis. Multimedia Tools and Applications 82, 25083–25105 (2023). https://doi.org/10.1007/s11042-023-14369-2

[8] Savaş, S. (2024). Application of deep ensemble learning for palm disease detection in smart agriculture. Heliyon, 10(17).

[9] R. Bouthaina, R. Khaled, K. Okba and M. Abdelhak, "Palm Tree Diseases Detection Using Deep Learning: A Short Review," *2024 8th International Conference on Image and Signal Processing and their Applications (ISPA)*, Biskra, Algeria, 2024, pp. 1-8, doi: 10.1109/ISPA59904.2024.10536746.

[10] Kavitha, M. G., Raviprakash, M. L., Alsalami, Z., Nagendar, Y., & Ramachandra, A. C. (2024, April). COCOs Nucifera Tree Disease Prediction Using Faster Region-Convolutional Neural Network with ResNet-50. In 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE) (pp. 01-04). IEEE.

[11] Sharma, A., Trivedi, N. K., & Sharma, A. K. (2024). Disease categorization and early detection in coconut leaves. In Artificial Intelligence and Information Technologies (pp. 156-162). CRC Press.

[12] Faisal, M., Leu, J. S., Avian, C., Prakosa, S. W., & Köppen, M. (2024). DFNet: Dense fusion convolution neural network for plant leaf disease classification. Agronomy Journal, 116(3), 826-838.

[13] Tuhin, H. H., Kayum, M. A., Sarker, M. R. K. R., Siddiquee, S. M. T., & Pandey, B. (2024, May). Identifying The Coconut Leaf Disease Using Deep Transfer Learning Approach. In 2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST) (pp. 427-432). IEEE.

[14] https://www.kaggle.com/datasets/samitha96/coconutdiseases

[15] Alam, T. S., Jowthi, C. B., & Pathak, A. (2024). Comparing pre-trained models for efficient leaf disease detection: a study on custom CNN. Journal of Electrical Systems and Information Technology, 11(1), 12.