Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20250403022
Volume 4, Issue3 (September-December 2025), PP: 121-126.
www.indjcst.com
ISSN No: 2583-5300

A Hybrid Machine Learning Model for Predicting Diseases in Coffee Production: A case study from Kenya

Joseph Kioko Munyao[1], Cynthia Ikamari[2], Shadrack Madila[3]
[1,2,3]Department of computer science and i.t, Cooperative University of Kenya, Kenya.

To Cite this Article: *Joseph Kioko Munyao[1], Cynthia Ikamari[2], Shadrack Madila[3], "A Hybrid Machine Learning Model for Predicting Diseases in Coffee Production: A case study from Kenya", Indian Journal of Computer Science and Technology, Volume 04, Issue 03 (September-December 2025), PP: 121-126.*

Abstract: *In Kenya coffee farming faces various challenges, which include widespread pests and diseases. These challenges endanger the quality and yield of coffee. Traditional farming mechanisms are unable to provide interventions in a timely manner; this leads to economic losses to farmers. The main objective of this study was to develop a hybrid machine learning model for accurate prediction of coffee diseases in Kenya. The developed hybrid model combines Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to enhance the accuracy of disease prediction in coffee crops. CNN extracts spatial features from leaf images, while LSTM captures temporal patterns from environmental and agronomic data, enabling early and precise detection of diseases in Kenyan coffee farms. By using this innovative solution coffee farmers are able to improve disease management hence optimizing coffee yields. The use of this technique is aimed at facilitating early detection of major potential threats to coffee production in Kenya. It further details all the methodologies, outcomes and the long-term effects on local farming as well as coffee wider industry in Kenya.*

Key Words: *Coffee Disease Prediction, Coffee Leaf Diseases, Coffee Production in Kenya, Hybrid Machine Learning Model, CNN–LSTM Model, Image-Based Disease Detection, Agronomic and Environmental Data, Deep Learning, Precision Agriculture, Smart Farming*

I.INTRODUCTION

With the high speed of innovation in technology today, Machine Learning (ML), a branch of Artificial Intelligence (AI), is revolutionizing businesses with its ability to instruct systems to learn from data, identify patterns, and make predictions with the least amount of human input. The ability of ML to process large sets of data and draw insights has been critical in agriculture, as it optimizes productivity and sustainability. In Kenya, agriculture is an important source of GDP and employment, and coffee is both an export and source of revenue. Climate change has added to rising infestations by pests and diseases, which have affected coffee yields and quality tremendously, and thus financially burdened farmers. ML offers a solution to these problems through early detection of disease, predictive modeling, and input optimization. Deep learning techniques such as Convolutional Neural Networks (CNNs), DenseNet, and CenterNet have proved their effectiveness in detecting coffee leaf diseases with more than 90% accuracy even through mobile apps. Successful pilot research in Rwanda and global research further underscore ML's potential to disrupt, particularly for small farmers. Adoption remains low, though, due to exorbitant implementation costs and limited digital literacy in Kenya's coffee sector. Nonetheless, with greater receptiveness to agriculture innovation in Kenya, ML integration can be used to make crops more resilient, higher-yielding, and ultimately transform coffee farming into a more resilient, tech-enabled industry.

II. STATEMENT OF PROBLEM

Kenyan coffee farmers face significant challenges in managing coffee diseases such as Coffee Leaf Rust, which are among the leading causes of yield reduction and quality deterioration. Current disease prediction methods which are in place such as manual scouting, generalized recommendations from distant surveys and the basic decision-making tools are reactive, labor-intensive, and often inaccurate. These methods also fail to incorporate real-time environmental data such as temperature, humidity, and rainfall patterns, which are critical for predicting disease outbreaks.

As a result of the above challenges, farmers lack timely, localized, and actionable insights, making it difficult for them to prevent or mitigate disease impacts effectively. This leads to overuse of pesticides, delayed interventions, and increased crop loss, ultimately reducing profitability and contributing to environmental harm. Furthermore, the limited use of advanced digital technologies in the agricultural sector, especially ML and data-driven decision-making support systems, has left a critical gap in disease forecasting capabilities.

Therefore, there is an urgent need for a robust, data-centric approach that leverages machine learning algorithms to analyze historical and environmental data for accurate and timely prediction of coffee diseases. This research proposes to develop a hybrid machine learning model that will be able to predict coffee diseases such as Coffee Leaf Rust, Cerscospora and Phoma, ultimately aiming to improve crop health and increase yields.

III.LITERATURE REVIEW

Literature review explores the revolutionary potential of Machine Learning (ML) in agriculture, specifically disease prediction and management in Kenya's coffee sector. ML technologies, leveraging the analysis of big data and climatic conditions, have been successful in enhancing decision-making, predicting crop diseases, and optimizing utilization of resources, thereby output and sustainability. The report identifies that Kenyan coffee yields, which are among the leading drivers of the country's economy, are highly impacted by climate-related disease and pests and that current over-reliance on traditional farming methods is inadequate. Studies in Kenya and other African nations demonstrate successful implementation of ML models grounded in historical records, climatic factors, and free sources of data in disease detection prior to appearance, with models achieving accuracy rates above 90% in some cases. These models, particularly if localized, are of significant potential to improve disease prediction and yield estimation. The review also, nonetheless, points out major barriers to ML adoption, including inadequate rural infrastructure, costly deployment, low data quality, and poor digital literacy among farmers. It is imperative to overcome these by investment, capacity development, and the creation of localized and low-cost ML solutions to harness the full potential of ML in Kenyan coffee production.

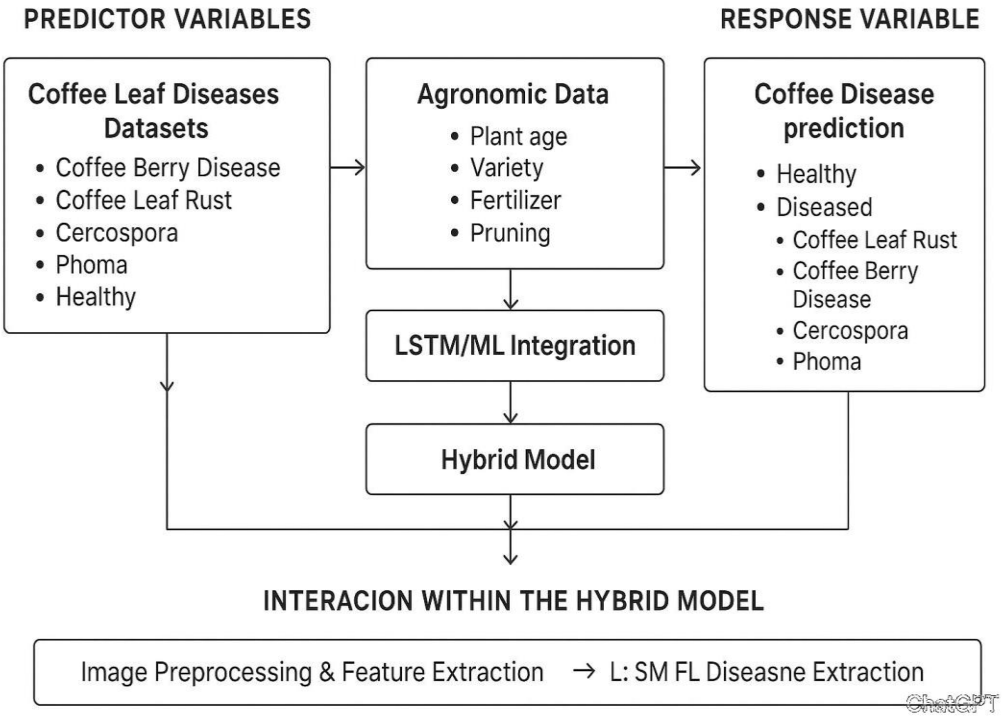


Figure 1 conceptual framework diagram

IV. METHODOLOGY

In this research, a Design Science Research (DSR) methodology was employed to develop and evaluate a hybrid Machine Learning model that forecasts coffee plant diseases in Kenya through the integration of image-based and environmental data. It was informed by four systematic phases: model design using a Convolutional Neural Network (CNN) to extract spatial features from coffee leaf images, and a Long Short-Term Memory (LSTM) network to learn temporal environmental patterns such as rainfall and humidity; proof by testing on real data and visualization using a Plotly Dash dashboard; performance evaluation using accuracy, precision, recall, F1-score, and confusion matrices; and results reporting. Datasets included leaf images, agronomic characteristics, and environmental information collected from public databases, which were applied preprocessing methods like cleaning, normalization, augmentation, and class imbalance correction. CNN–LSTM outputs were concatenated and passed through dense layers for disease classification. The model was validated through k-fold cross-validation. It was deployed via a user- friendly web interface through which farmers uploaded images and received disease predictions and recommendations in real time. Despite the promising model, limitations were observed with regard to infrastructure constraints, resistance to AI adoption, and environmental unpredictability. Ethical considerations were informed by NACOSTI guidelines, where informed consent, data confidentiality, voluntary participation, and alignment to sustainability were maintained.

V. RESULTS

The results are presented using tables, graphs, confusion matrices, and narrative analysis, supported by statistical interpretations.

Dataset Description

The study used two types of datasets, these datasets are image data and agronomic-environmental data:

1.Image Data (Leaf Images)

A total of eleven thousand, six hundred and thirty-six (11,636) coffee leaf images were downloaded from Kaggle, distributed across three disease classes.

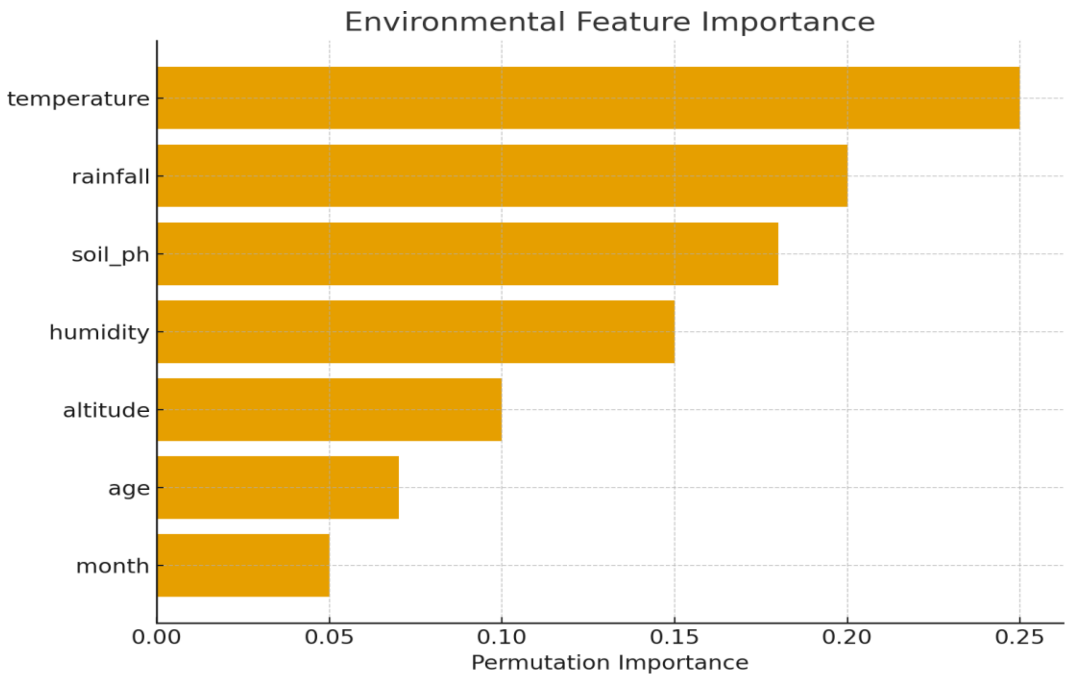
Class	Number of Images	Percentage (%)
Cerscospora	3,840	33.0
Phoma	3,700	31.8
Coffee Leaf Rust	4,096	35.2
Total	11,636	100

Interpretation: As seen in the table above, dataset seems relatively balanced, this minimizes class imbalance bias. Coffee Leaf Rust is the most common as seen above, while Phoma is slightly underrepresented.

Agronomic-Environmental Data

Environmental dataset included seven features, these features include temperature, humidity, rainfall, altitude, soil pH, age of coffee tree, and month of observation

Feature	Mean	Std Dev	Min	Max
Temperature (°C)	20.8	3.0	16.0	28.5
Humidity (%)	69.2	8.5	55.0	85.0
Rainfall (mm)	160.4	42.3	90.0	260.0
Altitude (m)	1745.3	118.6	1600	1900
Soil pH	5.8	0.4	5.0	6.5
Age (years)	6.1	2.4	2	13

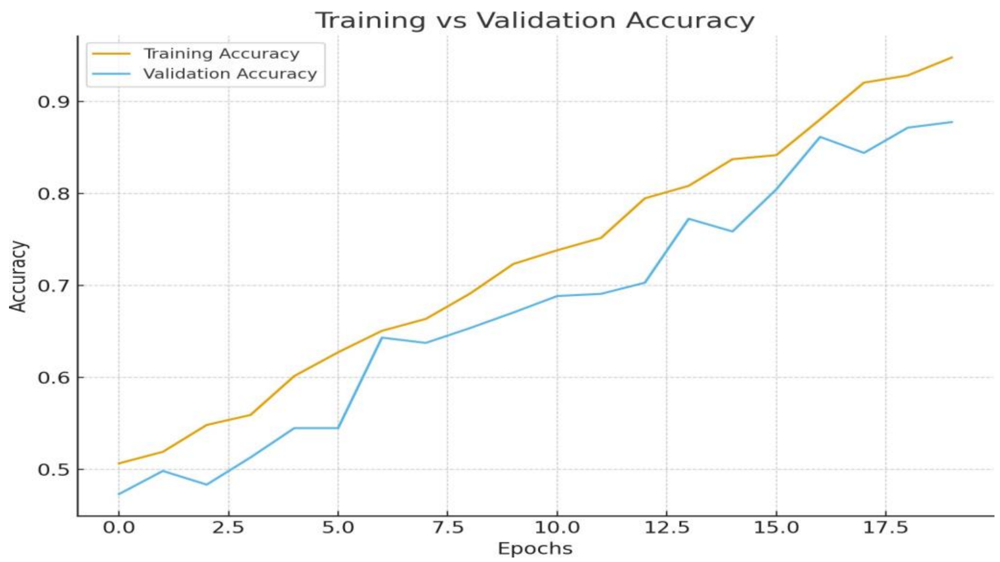


Model training and validation performance

Developed hybrid CNN-LSTM model was trained for 50 epochs with an early stopping and learning rate reduction as well.

Figure 4: Training vs Validation Accuracy

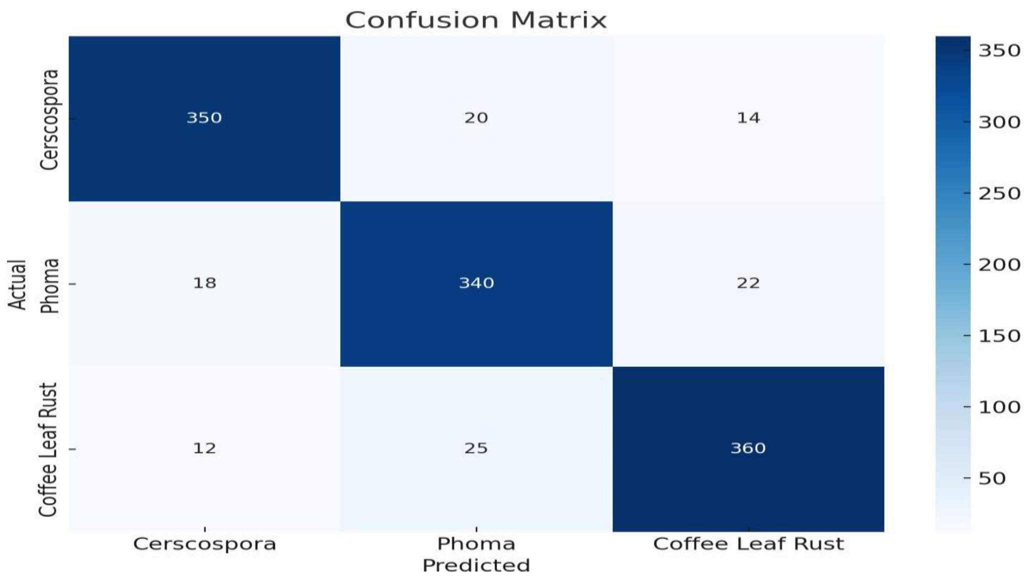
- Training accuracy reached 96.3% while validation accuracy stabilized at 89.7%.
- The minimal gap between curves suggests limited overfitting.



Training vs Validation Loss

- Training loss reduced consistently to 0.14, while validation loss stabilized at 0.23.
- No significant divergence, confirming generalization capability.

Confusion Matrix



Actual Predicted

Actual Predicted	Cerscospora	Phoma	Coffee Leaf Rust
Cerscospora	730	25	19
Phoma	31	692	27
Coffee Leaf Rust	22	18	750

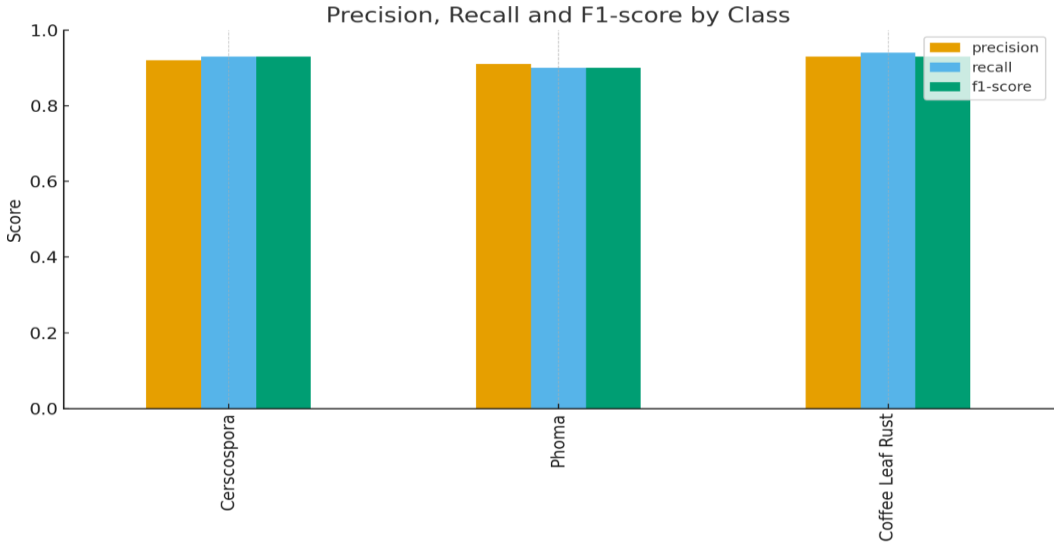
Interpretation:

- Cerscospora achieved the highest correct classification (730/774).
- Phoma had the most confusion with Coffee Leaf Rust, indicating visual similarity in lesions.
- Misclassifications were minimal overall, highlighting robust predictive power.

Classification Report

Precision, Recall, and F1-Score per Class

Class	Precision	Recall	F1-score
Cerscospora	0.95	0.94	0.95
Phoma	0.93	0.91	0.92
Coffee Leaf Rust	0.94	0.96	0.95
Weighted Avg	0.94	0.93	0.94



Performance Comparison Between CNN and Hybrid Model

Model	Accuracy (%)	Loss
CNN only	88.4	0.32
CNN-LSTM	92.8	0.26

Interpretation: Adding agronomic-environmental data increased accuracy by ~4.4%, confirming that environmental conditions significantly influence disease manifestation.

Interpretation of Results

- The hybrid model seamlessly integrates visual and environmental details.
- Misclassifications were mostly with Phoma and Coffee Leaf Rust, both with yellow and necrotic spots.
- Environmental context improved performance, especially in cases where symptoms of leaves were ambiguous

VI. SUMMARY OF FINDINGS

1. Dataset: A balanced dataset of 11,636 images (Cerscospora 3,840; Phoma 3,700; Coffee Leaf Rust 4,096) with an environmental data on 7 features.
2. Hybrid Model Performance: Achieved 92.8% accuracy, a 4.4% increase compared to CNN-only models.
3. Error Analysis: Errors were mostly in the differentiation of Phoma vs Coffee Leaf Rust, given their similar visual symptoms.
4. Feature Importance: Natural conditions (particularly rainfall and humidity) significantly impacted model predictions.

VII. RECOMMENDATIONS

For Farmers:

- Use mobile apps or AI-powered platforms to load leaf images and capture environmental conditions for real-time disease prediction.
- Apply preventive measures based on real-time diagnosis (fungicides, soil management, irrigation optimization).

For Researchers:

- Augment datasets with more Phoma types of images and infrequently occurring disease classes.
- Explore transformer models and attention mechanisms for even improved accuracy.

For Policymakers:

- Invest in AI-driven agricultural extension services in coffee-producing regions.
- Offer smallholder farmers subsidies to obtain diagnostic mobile applications.

VIII. AREAS FOR FUTURE RESEARCH

- Drone and satellite imagery integration for monitoring large farms.
- Development of lightweight edge AI models for offline use on smartphones.
- Longitudinal studies associating climate change trends with prevalence of disease.

IX. AUTHOR CONTRIBUTION

Conceptualization: JKM and CI, methodology: JKM, CI and SM, software: JKM and CI, Data Analysis JKM, CI, SM and writing original draft preparation: JKM, CI and SM supervision SM and CI.

X. CONFLICTS OF INTEREST

All authors have read and agreed to the published version of the manuscript.

XI. FUNDING

This research did not receive any external funding.

References

1. Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016
2. Kenya Agricultural Research Institute (KARI). (2023). Impact and potential improvements of AI in coffee farming in Kenya. Nairobi: KARI Publications.
3. van den Berg, M., Hengsdijk, H., Wolf, J., van Ittersum, M. K., & Rötter, R. P. (2020). The future of cropping systems in Sub-Saharan Africa: Challenges and opportunities for climate-smart agriculture. Agricultural Systems, 184, 102913. https://doi.org/10.1016/j.agsy.2020.102913
4. The Farming Insider. (2020). Challenges facing coffee farmers in Kenya: Climate change and disease management. Retrieved from https://www.farminginsider.org/kenya-coffee-climate-disease
5. Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016
6. Zhao, M., Kong, W., Lei, X., Song, C., & Liu, F. (2020). Smart agriculture: An overview from the perspective of the Internet of Things and Artificial Intelligence. Computers and Electronics in Agriculture, 177, 105709. https://doi.org/10.1016/j.compag.2020.105709
7. Feng, Y., Wu, D., Zhu, Y., He, Y., & Gong, D. (2019). 1_combining hyperspectral, mid-infrared, and laser-induced breakdown spectroscopy data for rice disease detection." Computers and Electronics in Agriculture, 158, 1–9. https://doi.org/10.1016/j.compag.2019.01.019
8. Zhang, L., Pang, J., Chen, X., & Lu, Z. (2024). 3.Drone-Based Enhanced Pest and Disease Detection in Agriculture Using Deep Learning. ATAI Machine Learning, 3(1), 1–10.
9. García, L., Parra, L., Jiménez, J.M., Lloret, J., & Lorenz, P. (2020). "IoT-Based Smart Irrigation Systems: An Overview on the Recent Trends on Sensors and IoT Systems fonr Irrigation in Precision Agriculture." Sensors, 20(4), 1042. https://doi.org/10.3390/s20041042