"ML-Based Disease Prediction in Tomato and Bean Crops in Sub-Saharan Africa: A Systematic Review"

Barbara Kunda[1], Jackson Phiri[2], Rebecca Lupyani[3]

ABSTRACT

This systematic review analyzes machine learning (ML) approaches for tomato and bean disease prediction in Sub-Saharan Africa, addressing gaps in real-world applicability and data integration. Findings show that deep learning (DL) models, especially CNNs and transfer learning, exceed 90% accuracy but demand high computational resources. Combining environmental and image data enhances reliability, yet infrastructure limitations hinder adoption. Most models rely on controlled datasets, limiting scalability, though mobile and edge computing show promise for low-resource settings. Key challenges include dataset biases, inadequate field validation, and minimal socio-economic integration. The study underscores the need for region-specific, efficient, and multimodal ML solutions to improve disease management. It advocates for tailored, sustainable AI strategies to bolster Sub-Saharan agriculture, calling for future research on deployable, context-aware technologies.

Key Words : Deep Learning, Sub-Saharan Africa, Transfer Learning, predictive analysis

INTRODUCTION

Machine learning (ML) and deep learning (DL) are transforming disease detection in Sub-Saharan Africa's tomato and bean crops, crucial for regional food security. While CNNs achieve high accuracy, challenges like computational costs, dataset biases, and limited real-world validation hinder scalability. Most studies rely on controlled datasets, neglecting field conditions and smallholder farmers' needs. This review evaluates ML methodologies, highlighting gaps in model generalizability, IoT integration, and socio-economic applicability. It advocates for context-aware, efficient solutions to enhance early disease prediction, reduce yield losses (20–40%), and support sustainable agriculture. Addressing these gaps can optimize precision farming and strengthen food security in resource-limited settings.

AIM OF THE STUDY

This study aims to synthesize existing research on machine learning approaches for predictive disease analysis in tomato and bean crops across Sub-Saharan Africa, by evaluating their effectiveness, identifying key technological gaps, and offering evidence-based recommendations to enhance modeling, ultimately supporting improved crop disease management and regional food security.

SYSTEMATIC REVIEW METHODOLOGY

This study employed a systematic literature review to analyze machine learning applications for tomato and bean disease prediction in Sub-Saharan Africa. The review involved:

1. **Database search**: Over 500 hundred papers were searched using targeted queries.
2. **Inclusion/exclusion criteria**: Peer-reviewed studies with empirical validation in Sub-Saharan contexts were included, while non-ML approaches and non-regional studies were excluded.
3. **Study selection**: 204 initial candidates were narrowed down using citation chaining and thematic analysis.
4. Thematic analysis: Findings were categorized into model performance, data integration, scalability, and practical challenges.

RESULTS AND DISCUSSION

:

I. SUMMARY OF FINDINGS

Theme	Key Observations
Model Accuracy	CNNs achieve >90% accuracy, but mostly on controlled datasets.
Data Use	Few models use multimodal data; most rely on images only.
Real-world Testing	Very limited; most studies are theoretical with no deployment.
Technological Constraints	High compute needs, limited mobile optimization, poor rural connectivity.
Promising Solutions	Edge computing, mobile-based apps, participatory design with farmers.

These results strongly suggest that for ML to be impactful in Sub-Saharan agriculture, future research must shift from theoretical accuracy to **deployable, efficient, and regionally-adapted solutions**. Prioritizing **context-aware models, lightweight deployment, field validation, and socio-economic alignment** is essential for transformative impact.

CONCLUSION

Machine learning, especially deep learning, shows great promise for early disease detection in tomato and bean crops in Sub-Saharan Africa, achieving over 90% accuracy. Key challenges include limited data, computational limits, and deployment difficulties. Progress requires collecting diverse field data, developing lightweight models for mobile use, integrating sensor data, and collaborating with local farmers to create practical, sustainable solutions.

REFERENCES

1. D. S. Abade, P. A. Ferreira, and F. D. B. Vidal, "Plant diseases recognition on images using CNN
2. J. Agarwal, S. Gupta, N. Sharma, and M. Manchanda, "A CNN method based predictive model for tomato leaf disease prediction