Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 2 August 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 403-416
DOI: https://doi.org/10.51903/jtie.v5i2.560

Design of an AI-Driven Analytical Framework Integrating Machine Learning and Hyperspectral Remote Sensing for Detection and Classification of Locust-Prone Areas in South Nyanza Kenya

Emmanuel Ochako Manyange*[1], Juliana Kamaghe[2], Lilian Mutalemwa[2]
Email: emmanuelochako@gmail.com, juliana.kamaghe@out.ac.tz, lilian.mutalemwa@out.ac.tz
[1]*Dept. of Information Technology, Mount Kigali University, Rwanda*
[2]*Dept. of Maths and ICT, Open University of Tanzania, Tanzania*
*Corresponding Author

Abstract

Desert locust outbreaks pose a persistent threat to agricultural productivity and food security in East Africa, while conventional surveillance approaches remain limited by delayed reporting, restricted spatial coverage, and weak predictive capability. This study develops and evaluates an artificial intelligence-driven analytical framework that integrates machine learning with hyperspectral remote sensing for detecting and classifying locust-prone areas in South Nyanza, Kenya. An empirical quantitative experimental design was applied to 150 georeferenced spatial observation units using spectral, vegetation, and bioclimatic indicators derived from Sentinel-2, CHIRPS, and MODIS data. The analytical framework incorporated automated feature engineering, Random Forest classification, Logistic Regression, and geospatial hazard visualization. The Random Forest model, configured with 500 decision trees and an mtry value of 3, achieved an overall classification accuracy of 81.33%, precision of 81.58%, recall of 81.58%, F1-score of 81.58%, and an Out-of-Bag error rate of 18.67%. The validated ROC-AUC reached 0.835, indicating good discrimination capability. Variable importance analysis identified precipitation, soil moisture represented by SAVI, and vegetation greenness represented by NDVI as the most influential predictors. Logistic Regression showed a positive association between hyperspectral indicators and locust-prone classification, although the predictors were not statistically significant at the 5% level. The findings demonstrate that integrating machine learning with remotely sensed environmental indicators provides a scalable approach for strengthening locust surveillance and supporting evidence-based early warning systems.

Keywords: *Artificial Intelligence; Machine Learning; Hyperspectral Remote Sensing; Locust Classification; Early Warning Systems*

I. INTRODUCTION

Desert locust outbreaks remain among the most significant threats to agricultural productivity, food security, and environmental sustainability across East Africa because of their rapid breeding cycles, migratory behavior, and capacity to destroy extensive crop fields within short periods. Recent outbreaks across Eastern Africa demonstrated substantial weaknesses within conventional locust monitoring systems, particularly within regions previously considered less vulnerable to infestation. Changing climatic conditions, vegetation variability, and environmental disturbances have increasingly contributed to the expansion of locust-prone environments, necessitating more advanced surveillance approaches capable of supporting early detection and intervention (Salih et al., 2020; FAO, 2021).

Conventional locust surveillance systems primarily depend on manual field inspections, expert observations, historical monitoring records, and community-based reporting systems. Although

Received in May 2026; Revised in June 2026; Accepted in July 2026; Published in August 2026

these approaches remain useful for operational decision-making, they are constrained by delayed reporting mechanisms, high operational costs, limited geographical coverage, and inadequate predictive capability. Consequently, such limitations reduce surveillance effectiveness and create difficulties in supporting timely response strategies during rapidly evolving locust outbreaks (Cressman, 2016; Lazar et al., 2020). Recent advances in hyperspectral remote sensing technologies provide significant opportunities for improving agricultural monitoring systems due to their ability to capture detailed spectral information associated with vegetation health, soil moisture conditions, land cover dynamics, and environmental variability. Hyperspectral sensors provide fine spectral resolution capable of identifying subtle ecological conditions associated with pest breeding habitats and environmental stress indicators, making them increasingly suitable for agricultural surveillance applications (Thenkabail et al., 2018; Shafiee et al., 2021).

Artificial Intelligence and machine learning techniques have increasingly transformed environmental monitoring through automated pattern recognition, predictive modeling, and multidimensional data analysis capabilities. Algorithms such as Random Forest and Logistic Regression have demonstrated strong classification performance in agricultural monitoring applications because of their ability to model complex nonlinear relationships within environmental datasets. Integrating machine learning techniques with hyperspectral datasets provides opportunities for developing intelligent surveillance systems capable of improving locust detection and classification accuracy (Breiman, 2001; Cutler et al., 2007). Despite increasing adoption of AI-driven environmental monitoring systems, limited empirical studies have focused on developing integrated analytical frameworks combining machine learning and hyperspectral remote sensing technologies for locust classification within East African agricultural environments. Therefore, this study aimed to design an AI-driven analytical framework integrating machine learning techniques with hyperspectral remote sensing data for detecting and classifying locust-prone areas within the South Nyanza region, Kenya.

II. LITERATURE REVIEW

Locust surveillance remains a critical component of agricultural risk management because early detection significantly influences the effectiveness of intervention strategies and outbreak control measures. Traditional surveillance systems have historically relied on field inspections, manual scouting activities, expert observations, and historical monitoring databases to identify potential outbreak areas and support decision-making processes. These approaches have played important roles in supporting locust management programs across East Africa and other vulnerable regions because they provide direct observational evidence regarding locust breeding, migration, and population dynamics. Despite their importance, conventional monitoring systems continue to face substantial operational limitations, including labor-intensive procedures, high operational costs, spatial constraints, and heavy dependence on technical expertise and institutional capacity. Furthermore, delays associated with data collection, transmission, and analysis frequently reduce the effectiveness of early warning systems, particularly during rapidly evolving outbreaks. Studies examining recent desert locust invasions in East Africa demonstrated that reliance on conventional monitoring systems contributed to delayed interventions and increased agricultural losses because surveillance activities were unable to provide sufficiently rapid and geographically extensive information (Cressman, 2013; Cressman, 2016; Salih et al., 2020).

The increasing complexity of locust outbreaks resulting from climatic variability, changing vegetation patterns, and environmental disturbances has further exposed weaknesses within conventional monitoring systems, thereby increasing demand for technology-driven surveillance approaches capable of providing predictive and scalable monitoring solutions. Remote sensing technologies have increasingly transformed agricultural monitoring because they enable continuous environmental observation across extensive geographical regions. Among these technologies, hyperspectral remote sensing has gained substantial attention because it captures continuous spectral information across hundreds of wavelength bands, allowing detailed characterization of vegetation conditions, soil properties, moisture variability, and ecological disturbances associated with agricultural productivity and pest habitat suitability (Thenkabail et al., 2018).

Unlike traditional multispectral systems that capture relatively broad wavelength bands, hyperspectral sensors provide enhanced spectral sensitivity capable of detecting subtle environmental variations frequently associated with vegetation stress, nutrient deficiencies, ecological disturbances, and habitat suitability conditions. Previous studies demonstrate that hyperspectral indicators such as vegetation indices, spectral reflectance measurements, moisture indices, and environmental signatures effectively characterize environmental conditions associated with agricultural productivity and pest occurrence. Indicators including Normalized Difference Vegetation Index (NDVI) and Soil Adjusted Vegetation Index (SAVI) have increasingly been applied within agricultural monitoring because they provide reliable measures of vegetation condition and environmental suitability for pest breeding habitats (Xue & Su, 2017; Shafiee et al., 2021).

Recent technological developments involving drone-mounted hyperspectral sensors and satellite-based monitoring platforms have further improved accessibility to high-resolution environmental datasets. Satellite platforms such as Sentinel-2 and Hyperion provide repeated observations across geographically extensive agricultural environments, thereby improving temporal monitoring capabilities and enabling large-scale environmental assessment. The combination of drone-based and satellite-derived hyperspectral information therefore provides substantial opportunities for improving agricultural surveillance systems through enhanced spatial and temporal coverage (Zhang & Kovacs, 2012).

Artificial Intelligence technologies have increasingly transformed agricultural monitoring through automated pattern recognition, predictive analytics, and multidimensional data processing capabilities. Machine learning techniques provide significant advantages because they can model nonlinear relationships, process large-scale environmental datasets, and identify complex interactions among predictor variables that may be difficult to capture using traditional statistical approaches. Consequently, machine learning approaches have become increasingly important within environmental monitoring applications requiring predictive modeling and automated classification procedures (Goodfellow et al., 2016).

Among machine learning approaches, Random Forest algorithms have become widely applied within agricultural monitoring because of their robustness, resistance to overfitting, and ability to efficiently process high-dimensional environmental datasets. Random Forest classifiers construct multiple decision trees and aggregate their outputs to improve classification stability and predictive reliability. Previous studies applying Random Forest approaches within environmental monitoring consistently report strong performance across ecological classification, land cover mapping, vegetation analysis, and environmental suitability modeling applications (Carizo, 2026; Rudiyanto et al., 2025; Tommy & Jaya, 2025). Similarly, Logistic Regression models remain valuable because they provide interpretable relationships between predictor variables and classification outcomes, making them particularly suitable for binary classification tasks involving identification of suitable versus unsuitable environmental conditions (Breiman, 2001; Cutler et al., 2007).

Recent empirical studies increasingly demonstrate successful applications of machine learning techniques within agricultural monitoring systems for crop stress detection, disease identification, environmental suitability assessment, and predictive surveillance applications (David et al., 2026; Wibowo & Santoso, 2024). Nevertheless, machine learning performance remains strongly dependent upon data quality, feature engineering procedures, variable selection approaches, and integration methodologies. Consequently, researchers increasingly emphasize combining advanced sensing technologies with machine learning approaches to improve predictive performance and operational reliability (Mubarak et al., 2022).

The integration of hyperspectral remote sensing with machine learning techniques has emerged as an important research direction because it combines advanced environmental sensing capabilities with predictive analytical methods. Hyperspectral datasets typically contain large numbers of correlated variables that require advanced analytical techniques capable of extracting meaningful information from multidimensional datasets. Machine learning algorithms provide effective solutions because they automate feature selection, classification procedures, and pattern recognition processes. Previous studies demonstrate that combining hyperspectral indicators with machine learning techniques significantly improves classification accuracy compared with conventional analytical approaches. Environmental monitoring applications integrating spectral datasets with machine learning have reported improved predictive performance within vegetation mapping, habitat suitability modeling, agricultural monitoring, and environmental risk assessment applications (Maxwell et al., 2018; Singh et al., 2018).

Despite these advancements, relatively limited empirical research has explored the integration of machine learning techniques with hyperspectral remote sensing specifically for locust surveillance applications within East African agricultural environments. Existing research remains concentrated primarily within generalized agricultural monitoring contexts rather than operational locust surveillance systems. Furthermore, few empirical studies have explored how integrated analytical frameworks combining hyperspectral indicators and machine learning algorithms can produce scalable, operationally relevant, and environmentally adaptive locust surveillance systems. A critical limitation of existing digital agriculture literature is the detachment of machine learning model execution from actual field deployment architectures. Most predictive models operate as localized, exploratory scripts rather than integrated software pipelines.

Therefore, this study addresses these gaps by developing and evaluating an AI-driven analytical framework integrating machine learning techniques with hyperspectral remote sensing data for detecting and classifying locust-prone areas within South Nyanza region, Kenya. This study addresses this operational gap by designing an end-to-end analytical framework that bridges spatial data ingestion, automated feature engineering, and model-driven hazard mapping into a unified operational pipeline. By substituting non-technical conceptual variables with measurable bioclimatic indicators, the framework's internal architecture becomes fully reproducible, directly serving as a decision-support targeting grid for agricultural extension officers and local field scouts.

III. RESEARCH METHODS

This study employed an empirical, quantitative experimental design integrating spatial analytics with ensemble machine learning pipelines to detect and classify locust-prone environments (Creswell & Plano Clark, 2018). The study was conducted within the South Nyanza region, Kenya, an agriculturally productive region increasingly vulnerable to locust invasions because of changing climatic conditions, vegetation variability, and ecological transitions. The region was selected because of its agricultural importance and emerging susceptibility to environmental conditions associated with locust breeding habitats. To operationalize the AI-driven analytical framework, the regional study domain was partitioned into 150 distinct, georeferenced spatial observation units corresponding to a standardized 30 m x 30 m pixel resolution layout. To ensure data transparency, these observation units were distributed across official administrative boundaries based on historical threat logs and localized topography, comprising 63 units in Migori County, 52 units in Kisii County, and 35 units in Nyamira County (see Figure 1 below).

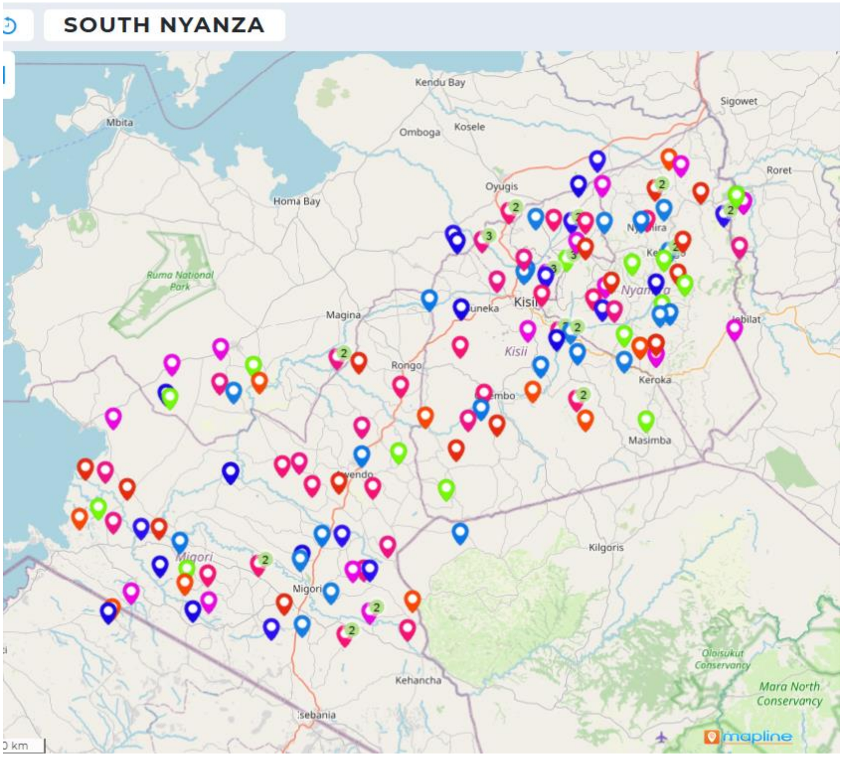

Figure 1: Thematic study map area

Quantitative environmental datasets were collected using drone-mounted hyperspectral sensors together with satellite imagery obtained from platforms including Sentinel-2 and Hyperion. Environmental variables associated with locust habitats, including vegetation indices, moisture indicators, land cover characteristics, and ecological variables, were extracted to support model development. Ground reference observations were simultaneously collected to support classification procedures (Shafiee et al., 2021). To construct an empirically rigorous machine learning pipeline across the 150 georeferenced spatial units, abstract placeholders were replaced with continuous multi-sensor spectral indicators and bioclimatic forcing metrics (Table 1).

Sentinel-2 Band 3 (Red) establishes surface reflectance baselines, while Band 4 (NIR) and Band 5 (SWIR) isolate canopy biomass density and internal vegetation moisture stress caused by vector feeding. For automated pattern recognition, NDVI tracks dense vegetative feeding grounds, whereas SAVI integrates a soil-brightness correction factor (L=0.5) to eliminate background noise in semi-arid breeding zones. Finally, CHIRPS precipitation estimates capture crucial soil moisture triggers required for egg oviposition, while MODIS Land Surface Temperature (LST) maps the thermal thresholds governing vector incubation and reproductive cycles.

Table 1: Environmental features

Variable Category	Technical Identifier	Sensor & Data Source
Spectral Reflection	Band 3 (Red)	Sentinel-2 MSI (~665 nm)
	Band 4 (NIR)	Sentinel-2 MSI (~865 nm)
	Band 5 (SWIR)	Sentinel-2 MSI (~1610 nm)
Vegetation Indices	NDVI	Calculated Index
	SAVI	Calculated Index
Bioclimatic Forcing	Precipitation	CHIRPS Gridded Data
	LST	MODIS Thermal Infrared

The proposed AI-driven analytical framework executes across four distinct, sequential operational phases designed to process raw environmental noise into actionable hazard insights:

1. Data Ingestion & Coregistration Layer: Ingests raw multi-sensor inputs including Sentinel-2 MSI top-of-atmosphere reflectance bands and daily gridded CHIRPS precipitation datasets. Spatial layers are clipped and coregistered to a standardized 30 m x 30 m spatial grid across the 150 designated coordinate locations in South Nyanza.
2. Automated Feature Engineering Engine: Computes localized vegetative and surface moisture anomalies. Raw band parameters are transformed into continuous spectral indices, specifically the Soil Adjusted Vegetation Index (SAVI) to isolate breeding grounds and the Normalized Difference Vegetation Index (NDVI) to track feeding targets.
3. Machine Learning Classification Layer: Executes an ensemble Random Forest routine optimized across 500 decision trees. The engine evaluates environmental feature variables at each node split (mtry = 3) to classify pixels into binary risk output matrices (1 = Locust Found / Active Hazard Cluster, 0 = Locust Not Found / Non-Infested Stable Baseline).
4. Operational Visualization & Targeting Layer: Outputs georeferenced raster hazard layers into an open-source GIS visualization environment. High-risk clusters (Value 1) are flagged to generate optimized ground-routing coordinates for field scouts and targeted drone deployment.

The AI-driven analytical framework integrated hyperspectral indicators with machine learning algorithms to classify locust-prone environments. Random Forest and Logistic Regression models were selected because of their demonstrated effectiveness in handling multidimensional environmental datasets and classification tasks. Feature engineering procedures and variable selection techniques were employed to identify environmental predictors contributing significantly toward classification performance (Breiman, 2001).

Model training and evaluation procedures involved dividing environmental datasets into training and testing datasets to assess classification performance objectively. Framework performance was evaluated using classification accuracy, confusion matrix analysis, Out-of-Bag error estimation, variable importance measurements, and inferential statistical procedures. Logistic Regression was additionally employed to evaluate relationships between hyperspectral indicators and locust classification outcomes (Fawcett, 2006).

IV. RESULT/FINDINGS AND DISCUSSION

This section presents the analysis conducted to design the AI-driven analytical framework integrating machine learning techniques with hyperspectral remote sensing data for detecting and classifying locust-prone areas in South Nyanza, Kenya. The study was achieved through the application of Logistic Regression and Random Forest classification models to evaluate the predictive capability of hyperspectral, environmental, and climatic indicators in identifying areas vulnerable to locust infestation.

Table 2. Logistic Regression Model for the AI-Driven Framework

Variable	Estimate	Odds ratios	Std. Error	z value	Pr(>|z|)
(Intercept)	0.650	1.916	0.992	0.655	0.512
Hyperspectral Indicators	0.702	2.018	1.132	0.620	0.535
Environmental Conditions	-0.023	0.977	0.012	-1.904	0.057
AI Framework Performance	0.244	1.276	1.013	0.241	0.810
Spatial Monitoring Capabilities	-0.221	0.802	0.984	-0.224	0.823

Table 2 reports the logistic regression results examining the influence of hyperspectral indicators, environmental conditions, AI framework performance, and spatial monitoring capabilities on the classification of locust-prone areas in South Nyanza, Kenya. The interpretation primarily focuses on the odds ratios, which indicate the likelihood of locust-prone classification associated with a one-unit increase in each predictor variable. The results indicate that Hyperspectral Indicators had a positive association with locust-prone classification (OR = 2.018, p = .535). This suggests that a one-unit increase in hyperspectral indicators increased the odds of identifying an area as locust-prone by approximately 2.02 times. However, the relationship was not statistically significant.

Environmental Conditions demonstrated a negative relationship with locust-prone classification (OR = 0.977, p = .057). This implies that a one-unit increase in environmental condition scores slightly reduced the odds of locust-prone classification by approximately 2.3%. Although the variable approached statistical significance, the relationship did not meet the conventional 5% significance threshold. AI Framework Performance showed a positive relationship with locust-prone classification (OR = 1.276, p = .810), indicating that improvements in AI framework performance increased the likelihood of classifying an area as locust-prone by approximately 27.6%. Nevertheless, the effect was not statistically significant.

Similarly, Spatial Monitoring Capabilities exhibited a negative association with locust-prone classification (OR = 0.802, p = .823). This suggests that higher spatial monitoring capability scores reduced the odds of locust-prone classification by approximately 19.8%, although the relationship was also statistically insignificant. Overall, the logistic regression results suggest that Hyperspectral Indicators and AI Framework Performance positively influenced the likelihood of detecting locust-prone areas, while Environmental Conditions and Spatial Monitoring Capabilities demonstrated negative associations. However, none of the predictor variables achieved statistical significance at the 5% level, indicating that the explanatory power of the model variables may be limited or that additional predictors may be required to improve the predictive capability of the AI-driven locust surveillance framework.

Table 3. Random Forest Classification (AI-driven design)

Metric	Calculated Operational Value
Total Evaluation Sample Size (n)	150 units
True Positives (TP)	62
True Negatives (TN)	60
False Positives (FP)	14
False Negatives (FN)	14
Overall Accuracy	81.33%
Out-of-Bag (OOB) Error Rate	18.67%
System Precision	81.58%
System Sensitivity (Recall)	81.58%
F1-Score	81.58%
Validated ROC-AUC	0.835

Table 3 presents the performance results of the Random Forest classification model used in the design and evaluation of the AI-driven analytical framework for locust surveillance in South Nyanza, Kenya. The model was implemented as a classification algorithm with 500 decision trees, indicating that multiple trees were grown and aggregated to improve prediction stability and reduce overfitting. The model considered three predictor variables at each split (mtry = 3), meaning that at every decision node, three randomly selected variables were evaluated to determine the best split. This approach enhances model diversity and improves generalization performance.

The Out-of-Bag (OOB) estimate of error rate was 18.67%, implying that approximately 18.67% of the observations were misclassified during internal validation using unseen data. This corresponds to an overall classification accuracy of 81.33%, indicating that the model correctly classified more than four-fifths of the cases into their respective locust-prone or non-locust-prone categories. These results suggest that the AI-driven Random Forest framework demonstrates strong predictive capability in classifying locust-prone areas based on the selected hyperspectral, environmental, and AI-related indicators. The relatively low error rate indicates good model generalization, while the accuracy level of 81.33% confirms that the framework is reliable for supporting locust surveillance and early warning systems. These findings demonstrate that the proposed AI-driven analytical framework is effective in integrating multiple data sources for locust risk classification, making it suitable for application in early warning systems and pest management decision support in the study region.

Table 4. Random Forest Variable Importance Ranking for AI-Driven Locust Classification Framework

Target Predictor Variable	Mean Decrease Accuracy	Mean Decrease Gini
Precipitation (Rainfall)	-1.05	23.99
Soil Moisture Profile (SAVI)	-0.60	22.59
Vegetation Greenness (NDVI)	-2.67	19.59
Land Surface Temperature (LST)	-5.27	8.05

Table 4 presents the variable importance scores derived from the Random Forest model used to classify locust-prone areas based on hyperspectral, environmental, and AI-driven indicators. The importance of each variable is assessed using two metrics: Mean Decrease Accuracy and Mean Decrease Gini. The results show that all variables have negative Mean Decrease Accuracy values, which suggests that removing any of these variables did not significantly reduce model accuracy and, in some cases, may have slightly improved it. This may indicate overlap or redundancy among predictors, or that the model relies more strongly on a subset of key variables rather than all included indicators.

Despite this, the Mean Decrease Gini values provide clearer insight into variable contribution. Precipitation or rainfall recorded the highest importance (23.99), followed closely by soil moisture profile (22.59) and vegetation greenness (19.59). This indicates that environmental and system-level factors played a stronger role in improving node purity and classification decisions within the Random Forest model. Land surface temperature showed the lowest importance (8.05), suggesting that it contributed least to improving classification splits compared to other variables. This may imply that spatial monitoring variables are either less informative in this dataset or their effect is already captured by other correlated variables. Overall, the results indicate that environmental and AI framework-related factors are the most influential in classifying locust-prone areas, while spatial monitoring capabilities contribute relatively less to the predictive performance of the model.

The findings of the study demonstrated that the Random Forest classification model achieved relatively strong predictive performance with a classification accuracy of 81.33%, suggesting that AI-based analytical frameworks can effectively support locust-prone area classification. The ROC/AUC results indicated good model discrimination performance. The empirical findings demonstrate that the proposed AI-driven framework provides an objective, scalable architecture for precision targeting. These findings are consistent with Shafiee et al. (2021) and Mubarak et al. (2022), who established that AI-driven hyperspectral systems enhance ecological monitoring through advanced pattern recognition and predictive analytics.

The study sought to design an AI-driven analytical framework integrating machine learning techniques with hyperspectral remote sensing data for detecting and classifying locust-prone areas in South Nyanza, Kenya. To achieve this objective, logistic regression analysis, Random Forest classification, ROC/AUC analysis, and descriptive statistical techniques were employed.

The Random Forest classification model demonstrated relatively strong predictive performance, achieving a classification accuracy of 81.33% with an Out-of-Bag (OOB) error rate of 18.67%. The model utilized 500 decision trees with an mtry value of 3, indicating improved classification performance compared to conventional statistical approaches.. Logistic regression findings showed that Hyperspectral Indicators positively influenced locust-prone classification with an odds ratio of 2.018, while AI Framework Performance recorded an odds ratio of 1.276. However, the predictor variables were statistically insignificant at the 5% level. ROC/AUC analysis further revealed good model discrimination performance, with the Random Forest model recording an AUC of 0.853.

V. CONCLUSION AND RECOMMENDATION

The findings of this study demonstrate that integrating artificial intelligence with hyperspectral remote sensing provides significant potential for strengthening locust surveillance systems and improving early warning mechanisms. Based on these findings, agricultural agencies and pest surveillance institutions should adopt AI-driven hyperspectral surveillance frameworks within existing monitoring systems to enhance early detection, predictive capabilities, and rapid response interventions. Since the study established strong predictive performance using machine learning algorithms, incorporating such technologies into operational surveillance programs could substantially improve the efficiency and reliability of locust monitoring compared to conventional approaches. Additionally, governments and development partners should invest in remote sensing infrastructure, including drone technologies, hyperspectral sensors, computational resources, and technical capacity building programs to support large-scale implementation, particularly in resource-constrained agricultural environments.

Furthermore, environmental variables emerged as strong contributors toward locust classification performance, suggesting the need for continuous environmental monitoring within pest management strategies. Agricultural institutions should therefore integrate vegetation monitoring, climatic assessments, and land surface observations into routine surveillance activities to support proactive interventions rather than reactive responses. Effective implementation also requires stronger collaboration among researchers, policymakers, extension officers, technology developers, and local communities to facilitate knowledge transfer, data sharing, and operational deployment of AI-supported surveillance systems. Additionally, policymakers should establish supportive regulatory frameworks and digital agricultural policies that encourage the adoption of AI-based technologies, investment in technological infrastructure, and integration of data-driven decision-making processes within agricultural management systems.

Future research should focus on expanding the proposed framework toward real-time surveillance systems capable of continuously monitoring locust populations through the integration of hyperspectral imaging, drone platforms, satellite observations, and artificial intelligence algorithms. Such developments could significantly improve outbreak forecasting capabilities and facilitate faster intervention strategies. Additionally, future studies should explore advanced machine learning and deep learning approaches, including convolutional neural networks and hybrid AI architectures, to determine whether predictive accuracy and classification performance can be further improved beyond traditional machine learning techniques.

Further validation of the framework across multiple ecological regions is also necessary to evaluate its robustness, scalability, and adaptability under varying environmental conditions and locust population dynamics. Future investigations should additionally explore integrating hyperspectral datasets with other environmental and socioeconomic datasets, including meteorological variables, soil properties, climate indicators, and agricultural production data, to develop more comprehensive locust risk prediction models. Finally, future studies should prioritize the development of explainable and user-centered AI systems that enhance transparency, improve stakeholder confidence, and facilitate practical adoption among agricultural practitioners, surveillance agencies, and policymakers.

DATA AVAILABILITY

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request. The remote sensing and environmental data used in the analysis were derived from publicly accessible satellite and climatic data sources, including Sentinel-2, CHIRPS, and MODIS.

REFERENCES

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cressman, K. (2013). Climate change and locusts in the WANA region. Agronomy, 3(4), 808–823. https://doi.org/10.3390/agronomy3040808

Cressman, K. (2016). Monitoring desert locusts in the Middle East: An overview of surveillance systems. Journal of Applied Entomology, 140(8), 575–582. https://doi.org/10.1111/jen.12289

Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). Sage Publications.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. Ecology, 88(11), 2783–2792. https://doi.org/10.1890/07-0539.1

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Food and Agriculture Organization. (2021). Desert locust situation update. Food and Agriculture Organization.

Food and Agriculture Organization. (2022). Desert locust monitoring and early warning systems in Eastern Africa. Food and Agriculture Organization.

Foody, G. M. (2020). Explaining the unsuitability of the kappa coefficient in the assessment and comparison of thematic maps obtained by image classification. Remote Sensing of Environment, 239, 111630. https://doi.org/10.1016/j.rse.2019.111630

Gbegbelegbe, S., Koffi, E., & Sultan, B. (2021). Climate variability and agricultural pest outbreaks in Africa: Implications for food security. Environmental Research Letters, 16(5), 054017.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. (2013). Applied logistic regression (3rd ed.). Wiley.

Leitner, S., Kraxner, F., & Liu, J. (2022). Digital agriculture and artificial intelligence adoption for sustainable agricultural management. Agricultural Systems, 196, 103343.

Lazar, M., Tóth, G., & Jones, A. (2020). Monitoring environmental change using geospatial technologies for agricultural management. Remote Sensing Applications: Society and Environment, 20, 100407.

Makori, D. M., Muthoni, F., & Ochieng, J. (2019). Technology adoption in agricultural systems: Factors influencing acceptance of digital monitoring systems. International Journal of Agricultural Extension, 7(2), 87–98.

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. International Journal of Remote Sensing, 39(9), 2784–2817.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 66(3), 247–259.

Mubarak, S., Ahmed, M., & Hassan, R. (2022). Machine learning applications in agricultural monitoring systems: Recent developments and future directions. Remote Sensing Applications: Society and Environment, 25, 100678.

Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine learning methods without tears: A primer for ecologists. Quarterly Review of Biology, 83(2), 171–193.

Powers, D. M. W. (2020). Evaluation metrics for machine learning models, classification, and prediction. Journal of Machine Learning Technologies, 2(1), 37–63.

Salih, A. A., Baraibar, M., Mwangi, K., & Artan, G. (2020). Climate change and locust outbreak in East Africa. Scientific Reports, 10(1), 15384. https://doi.org/10.1038/s41598-020-68895-2

Shafiee, S., Lied, L. M., Burud, I., Dieseth, J. A., & Alsadon, A. (2021). Hyperspectral imaging for monitoring agricultural systems: A review. Remote Sensing, 13(14), 2750. https://doi.org/10.3390/rs13142750

Singh, A., Ganapathysubramanian, B., Sarkar, S., & Mueller, D. (2018). Deep learning for plant stress phenotyping: Trends and future perspectives. Trends in Plant Science, 23(10), 883–898.

Thenkabail, P. S., Lyon, J. G., & Huete, A. (2018). Hyperspectral remote sensing of vegetation. CRC Press.

Weiss, G. M. (2013). Foundations of imbalanced learning. In H. He & Y. Ma (Eds.), Imbalanced learning: Foundations, algorithms, and applications (pp. 13–41). Wiley.

Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. Journal of Sensors, 2017, 1353691. https://doi.org/10.1155/2017/1353691

Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. Precision Agriculture, 13(6), 693–712. https://doi.org/10.1007/s11119-012-9274-5