Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 2 August 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 334-348
DOI: https://doi.org/10.51903/jtie.v5i2.554

Detection of Locust Breeding and Feeding Zones Using Hyperspectral Signatures and Environmental Variables in South Nyanza Region, 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*, 575
[2]*Dept. of Maths and ICT, Open University of Tanzania, Tanzania, 23049*
*Corresponding Author

Abstract

Desert locust infestations continue to threaten agricultural productivity, food security, and environmental sustainability across East Africa. Conventional locust surveillance approaches remain largely dependent on field-based monitoring systems that are reactive, labor-intensive, and constrained by limited spatial coverage. This study investigated the hyperspectral signatures and environmental variables associated with locust breeding and feeding zones in South Nyanza, Kenya, using an Artificial Intelligence (AI)-driven hyperspectral remote sensing framework. Environmental variables including vegetation indices, soil moisture, rainfall, temperature, vegetation stress indicators, and hyperspectral bands were analyzed using correlation analysis, binary logistic regression, Random Forest classification, and variable importance ranking. Correlation analysis revealed weak relationships between individual environmental indicators and historical locust occurrence, with all predictors demonstrating statistically insignificant relationships ($p > 0.05$). Logistic regression findings similarly indicated that no individual environmental variable independently predicted locust occurrence. However, Random Forest classification achieved a classification accuracy of 75.33% with an Out-of-Bag error rate of 24.67%, suggesting moderate predictive capability when multiple variables were integrated. Variable importance analysis identified rainfall, soil moisture, temperature, and vegetation-related indicators as dominant predictors. The findings suggest that locust breeding and feeding habitats are influenced by complex multidimensional interactions rather than isolated environmental indicators. The study demonstrates the practical potential of AI-driven hyperspectral frameworks for strengthening locust surveillance and early warning systems within emerging high-risk agricultural regions.

Keywords: *Hyperspectral remote sensing, desert locust surveillance, artificial intelligence, Random Forest, environmental indicators, South Nyanza.*

I. INTRODUCTION

Desert locust infestations remain among the most destructive biological threats affecting agricultural production systems due to their rapid reproductive capability, transboundary migratory behavior, and ability to destroy crops across extensive geographical regions. Desert locust outbreaks have historically caused severe socioeconomic consequences through reduced agricultural productivity, food insecurity, pasture degradation, and livelihood disruptions, particularly within developing economies that heavily depend on rain-fed agriculture (Magor et al., 2008). The recent desert locust upsurge that affected East Africa between 2019 and 2022 exposed major weaknesses within existing surveillance and early warning systems, highlighting significant limitations in preparedness, monitoring capacity, and rapid response mechanisms (Food and Agriculture Organization [FAO], 2020). Kenya was among the countries heavily affected during this period, with outbreaks extending beyond traditional arid and semi-arid regions into previously less monitored agricultural zones.

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

Conventional locust surveillance systems have historically concentrated within arid and semi-arid environments because these regions have traditionally been associated with desert locust breeding activities (Cressman & Hodson, 2019). Consequently, surveillance infrastructure, monitoring resources, and intervention programs have predominantly focused on northern and eastern regions of Kenya while paying relatively limited attention to agriculturally productive regions such as South Nyanza. However, changing climatic conditions, environmental variability, and shifting ecological patterns have increasingly altered locust distribution patterns, creating emerging vulnerabilities within regions previously considered less susceptible to outbreaks (Mongare et al., 2023). These ecological changes suggest the need for expanded surveillance systems capable of identifying emerging breeding and feeding zones beyond historically affected landscapes.

Environmental conditions play a critical role in determining locust habitat suitability because breeding, migration, feeding behavior, and survival patterns are strongly influenced by ecological variability. Factors including vegetation growth, rainfall distribution, soil moisture availability, and temperature fluctuations significantly affect habitat suitability and population dynamics (Hunter, 2004). Rainfall events stimulate vegetation growth while simultaneously creating favorable soil conditions necessary for oviposition and egg development. Similarly, soil moisture directly influences incubation success and nymph survival rates, while temperature variability affects metabolic processes, reproductive cycles, and migration behavior (Cressman, 2013). Consequently, environmental monitoring has increasingly become central to locust surveillance systems because environmental indicators provide important information regarding ecological conditions associated with outbreak risk.

Despite growing recognition of the importance of environmental monitoring, conventional surveillance approaches remain largely dependent on manual scouting, field observations, and community reporting systems that are frequently characterized by delayed detection, inadequate spatial coverage, operational inefficiencies, and limited predictive capability (Cressman, 2016). Such limitations become particularly problematic within heterogeneous agricultural systems where environmental conditions vary rapidly across space and time. Therefore, improving surveillance capability requires the integration of advanced monitoring technologies capable of capturing environmental variability at higher spatial and temporal resolutions.

Recent advances in remote sensing technologies have created new opportunities for improving environmental monitoring and ecological surveillance. Hyperspectral remote sensing provides significant advantages over conventional multispectral approaches because it captures spectral information across hundreds of narrow wavelength bands capable of detecting subtle physiological, biochemical, and structural variations within vegetation systems (Thenkabail et al., 2012). These technologies provide improved capability for identifying vegetation conditions, moisture variability, and environmental stress indicators associated with locust breeding and feeding habitats. Previous studies have demonstrated that hyperspectral approaches can substantially improve ecological classification performance because high-dimensional spectral information captures environmental variations that conventional monitoring systems may fail to detect (Shafiee et al., 2021).

The integration of hyperspectral remote sensing with Artificial Intelligence (AI) techniques further enhances surveillance capabilities because machine learning algorithms effectively model complex nonlinear relationships among multiple environmental variables simultaneously. Machine learning approaches such as Random Forest classification have increasingly gained attention within agricultural monitoring because they accommodate multidimensional datasets, improve predictive performance, and effectively capture complex environmental interactions (Kamilaris & Prenafeta-Boldú, 2018). AI-driven analytical frameworks therefore provide opportunities for improving locust surveillance through enhanced classification accuracy, predictive modeling, and early warning capabilities.

Despite growing adoption of AI-driven surveillance systems within agriculture, limited empirical evidence exists regarding the specific hyperspectral signatures and environmental indicators associated with locust breeding and feeding habitats within heterogeneous agricultural landscapes characterized by mixed farming systems and rapidly changing phenological conditions. Existing studies have largely concentrated within traditional arid environments while paying limited attention to emerging agricultural risk zones such as South Nyanza (FAO, 2021). This creates an important knowledge gap regarding the applicability of AI-driven hyperspectral frameworks within ecologically complex agricultural systems. Therefore, this study investigated the hyperspectral signatures and environmental variables associated with locust breeding and feeding zones in South Nyanza using an AI-driven hyperspectral analytical framework to strengthen surveillance capability and improve evidence-based locust management strategies.

II. LITERATURE REVIEW

Previous studies have consistently demonstrated that climatic and environmental conditions strongly influence locust ecology, breeding dynamics, migration behavior, and outbreak patterns. Environmental conditions such as rainfall, vegetation availability, temperature variability, and soil moisture determine habitat suitability and directly affect locust reproduction and survival rates. According to Cressman (2013), desert locust populations are highly sensitive to ecological conditions because favorable environmental circumstances stimulate vegetation growth and create suitable breeding habitats. Rainfall events are particularly important because they increase

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 2 August 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 334-348
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

vegetation biomass while simultaneously improving soil conditions necessary for egg development. Similarly, Magor et al. (2008) argued that locust outbreaks are strongly associated with rainfall-induced vegetation growth because favorable environmental conditions increase food availability and enhance reproductive success. Studies conducted during recent desert locust upsurges further demonstrated that unusual climatic events, including prolonged rainfall and changing weather patterns, significantly contributed to the expansion of breeding zones beyond traditional arid regions (Food and Agriculture Organization [FAO], 2020).

Soil moisture similarly plays a fundamental role in locust breeding ecology because it directly influences oviposition success, egg development, and nymph survival. Research suggests that moist soils provide favorable conditions for egg incubation and increase the probability of successful hatching, particularly when accompanied by suitable vegetation cover (Hunter, 2004). Temperature variability also influences locust developmental cycles because insect metabolism, reproduction rates, migration behavior, and feeding activity are temperature-dependent processes. Environmental variability arising from interactions among temperature, rainfall, and moisture availability therefore creates dynamic ecological conditions that influence locust distribution and outbreak patterns (Mongare et al., 2023). Consequently, environmental monitoring has increasingly become central to modern locust surveillance systems because environmental indicators provide valuable information regarding potential habitat suitability and outbreak risks (Cressman & Hodson, 2019).

Vegetation monitoring has become increasingly important within agricultural surveillance systems because vegetation conditions provide indirect indicators of environmental suitability for pest establishment. Vegetation indices such as the Normalized Difference Vegetation Index (NDVI) and Soil Adjusted Vegetation Index (SAVI) have frequently been applied in ecological monitoring because they provide quantitative measures of vegetation density, biomass, stress conditions, and plant health (Rembold et al., 2013). NDVI has been extensively used for agricultural monitoring because it effectively captures vegetation greenness and photosynthetic activity, allowing researchers to identify areas experiencing environmental changes associated with pest outbreaks (Wang et al., 2010). Similarly, SAVI improves vegetation detection by minimizing soil brightness effects, making it particularly useful in heterogeneous agricultural systems where soil exposure varies significantly (Thenkabail et al., 2012).

Studies have demonstrated that vegetation indices are effective indicators for identifying favorable ecological conditions associated with locust breeding and feeding habitats because vegetation abundance determines food availability and influences insect population dynamics (Hunter, 2004). Furthermore, vegetation stress indicators provide additional ecological insights because stressed vegetation often reflects changing environmental conditions that may influence locust movement and feeding behavior. As agricultural systems become increasingly dynamic due to climate variability, vegetation monitoring approaches have become essential components of environmental surveillance frameworks (Pekel et al., 2016).

Recent advances in remote sensing technologies have substantially transformed environmental monitoring through improved spatial resolution, temporal frequency, and analytical capabilities. Hyperspectral remote sensing extends beyond conventional multispectral approaches by capturing spectral information across hundreds of narrow wavelength bands capable of detecting subtle physiological, biochemical, and structural variations within vegetation systems (Thenkabail et al., 2012). Unlike conventional remote sensing approaches that rely on relatively broad spectral measurements, hyperspectral sensors provide greater sensitivity for detecting vegetation stress, moisture conditions, nutrient variations, and environmental disturbances that may be associated with locust habitats (Shafiee et al., 2021).

The integration of hyperspectral technologies within agricultural monitoring systems has increasingly attracted attention because spectral information can provide detailed ecological insights necessary for early detection and predictive surveillance. Studies conducted within precision agriculture have demonstrated that hyperspectral approaches improve classification accuracy by capturing high-dimensional spectral information capable of distinguishing subtle environmental variations (Zhang & Kovacs, 2012). Consequently, hyperspectral technologies are increasingly viewed as promising tools for pest monitoring because they provide richer environmental information compared to conventional remote sensing systems (Yu et al., 2021).

Artificial Intelligence (AI) and machine learning approaches have further improved the analytical capabilities of environmental surveillance systems because they enable researchers to model nonlinear relationships among multiple environmental variables simultaneously. Machine learning techniques such as Random Forest have gained substantial popularity because they effectively accommodate multidimensional datasets, minimize overfitting, and provide robust predictive performance under complex ecological conditions (Kamilaris & Prenafeta-Boldú, 2018). Random Forest classification approaches are particularly useful within ecological monitoring because they integrate multiple predictor variables while simultaneously estimating variable importance and classification uncertainty (Singh et al., 2016).

Studies applying machine learning approaches within agricultural monitoring have demonstrated improved predictive performance compared to conventional statistical approaches because machine learning algorithms effectively capture complex environmental interactions that traditional linear models frequently fail to detect (Shafiee et al., 2021). Despite these advantages, AI adoption within African agricultural surveillance systems remains relatively limited due to infrastructural constraints, technical barriers, and insufficient localized datasets (Mubarak et al., 2022). Existing surveillance systems continue to rely heavily on manual monitoring approaches despite growing evidence supporting AI-driven predictive frameworks.

Although previous studies have substantially improved understanding of locust ecology and environmental monitoring systems, important empirical gaps remain. Existing research has predominantly concentrated within arid and semi-arid environments where desert locust outbreaks historically occur (FAO, 2021). Consequently, limited attention has been given to heterogeneous agricultural landscapes characterized by mixed farming systems, diverse vegetation structures, and rapidly changing phenological conditions. Regions such as South Nyanza present unique ecological complexities because agricultural activities occur within highly fragmented landscapes that may exhibit distinct environmental signatures compared to traditional locust habitats.

Furthermore, while previous studies have independently examined vegetation monitoring, environmental variables, machine learning, and hyperspectral technologies, limited empirical evidence exists regarding the integration of these approaches within emerging locust-prone agricultural regions. This creates an important knowledge gap regarding the applicability of AI-driven hyperspectral surveillance systems within heterogeneous environments such as South Nyanza. Therefore, the present study seeks to address this gap by investigating the hyperspectral signatures and environmental variables associated with locust breeding and feeding zones within South Nyanza using an integrated AI-driven analytical framework.

III. RESEARCH METHODS

This study was conducted in the South Nyanza region of Kenya, comprising Migori, Kisii, and Nyamira counties. The area is characterized by diverse agricultural systems, humid climatic conditions, heterogeneous vegetation cover, and mixed farming practices that provide favorable ecological conditions for desert locust breeding and feeding. The geographical domain of the analysis ranges between approximately -0.3505° and 0.5029° Latitude and 34.15172° to 35.06998° Longitude. To establish a standardized spatial framework, the study area was partitioned into 150 distinct, georeferenced spatial sampling units. To ensure rigorous transparency, these observation units were distributed across the administrative boundaries based on localized topography and historical threat logs, comprising 63 units in Migori County, 52 units in Kisii County, and 35 units in Nyamira County (See Figure 1 below). Each observation unit corresponds precisely to a 30m x 30m spatial resolution cell to maintain pixel-level co-registration with satellite imagery assets.

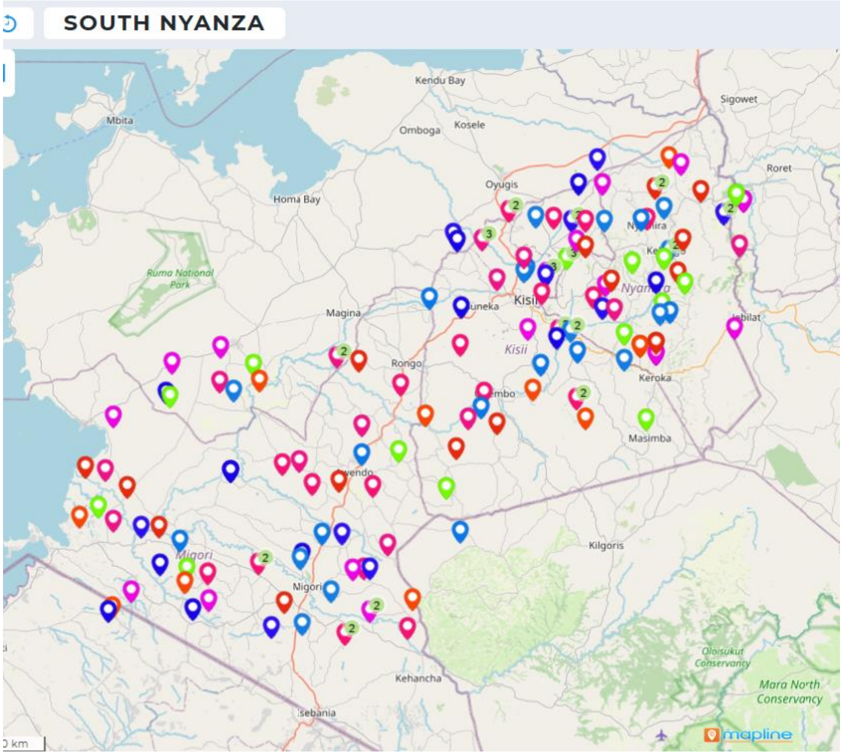

Figure 1: Thematic study map area

Biophysical environmental predictors and hyperspectral observations were compiled across a temporal horizon spanning 2020 to 2023 to capture local phenological and macroclimatic variations.

A. Hyperspectral Imagery Data

Primary hyperspectral data were acquired from the PRISMA (Poldi Elettro-Ottico) satellite mission, featuring a 30m spatial resolution. The imagery underwent rigorous preprocessing workflows in ENVI, including absolute radiometric calibration, FLAASH-driven atmospheric correction to eliminate path radiance, geometric orthorectification, and automated cloud-masking. Out of the 239 narrow spectral channels available, five core bands were extracted based on their physical relevance to structural vegetation health and surface moisture diagnostics:

- Band 1 (Blue Spectrum, ~490 nm): Captures soil background variations and active chlorophyll absorption dynamics.
- Band 2 (Green Spectrum, ~560 nm): Measures structural vegetation greenness and healthy canopy reflectance.
- Band 3 (Red Spectrum, ~665 nm): Serves as a vital baseline for chlorophyll absorption calculations.
- Band 4 (Near-Infrared / NIR, ~865 nm): Highly sensitive to internal leaf cellular structures, biomass density, and canopy water volume.

- Band 5 (Short-Wave Infrared / SWIR, ~1610 nm): Sensitive to vegetation canopy moisture deficits, plant water stress, and ambient soil dryness.

B. Macroclimatic and Biophysical Indicators

Co-registered environmental covariates matching the specific Date Collected records of each spatial unit were extracted. These included land surface temperature (°C), cumulative historical rainfall (mm), Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), and categorical surface designations grouped via QGIS into distinct land cover vectors (e.g., Cropland, Dryland, Shrubland).

Historical locust occurrence coordinates were compiled from spatial surveillance databases maintained by the Kenya Ministry of Agriculture, Livestock, Fisheries and Cooperatives, supplemented by Food and Agriculture Organization (FAO) monitoring logs. Ground-truth classifications were cross-verified via localized field validation coordinates. Each of the 150 spatial sampling rows was assigned an objective binary Ground_Truth label:

- Active Breeding/Feeding Landscape (1): Confirmed historical or field occurrence of localized swarms, nymph hopper bands, or active gregarious oviposition.
- Non-Locust Landscape (0): Confirmed absent or highly unsuitable ecological baseline zones.

To isolate the non-linear dependencies governing locust distribution, an ensemble Random Forest classifier was constructed using 500 independent decision trees. To avoid structural overfitting, a hyperparameter restriction of mtry = 3 was enforced, forcing the algorithm to select from three random feature spaces at each node split point. Crucially, to maintain rigorous model validity, only objective biophysical and remote sensing variables (Bands 1–5, NDVI, SAVI, Temperature, Rainfall, and Land Cover Type) were permitted as input features to classify the Ground Truth state. Qualitative, human-derived indicators were explicitly barred from entering the machine learning processing pipeline to prevent mathematical distortion of the model's technical accuracy.

To complement the biophysical predictive architecture and evaluate operational scalability, an independent mixed-methods qualitative assessment was executed alongside the technical model. Qualitative metrics, which included primary monitoring mechanisms (Monitoring Method, such as Drone vs. Field Surveying), perceived community scalability metrics (Scalability Rating, ranked 1–3), localized system Applicability Scores, and field-assigned Early Warning Status tiers, were cataloged from regional agricultural extension officers and local smallholder agricultural stakeholders. This socio-technical assessment was analyzed independently from the Random Forest classifier, serving exclusively as an operational validation phase to contextualize how well the machine learning outputs align with real-world infrastructure constraints and user adoption capabilities in South Nyanza.

IV. RESULT AND DISCUSSION

To evaluate the individual predictive capability of both biophysical covariates and remote sensing spectral features, inferential statistical pipelines were executed against the 150 ground-truth observation units. Table 1 outlines the Pearson correlation coefficients (r) derived between individual environmental indicators and historical desert locust presence. The findings show that all evaluated parameters exhibit weak, statistically insignificant linear relationships with historical locust occurrence ($p > 0.05$).

Table 1. Correlation Analysis of Environmental Indicators and Historical Locust Occurrence

Variable Category	Indicator / Feature	Specific Spectral Wavelength / Metric Description	Correlation Coefficient (r)	p-value
Spectral Refl.	Band 1 (Blue)	~490 nm	0.08	0.328
Spectral Refl.	Band 2 (Green)	~560 nm	0.14	0.099
Spectral Refl.	Band 3 (Red)	~665 nm	0.01	0.874
Spectral Refl.	Band 4 (NIR)	~865 nm	0.06	0.499
Spectral Refl.	Band 5 (SWIR)	~1610 nm	0.01	0.870
Vegetation Ind.	NDVI	Normalized Difference Vegetation Index	-0.01	0.865
Vegetation Ind.	SAVI	Soil Adjusted Vegetation Index	-0.02	0.832
Biophysical	Soil Moisture	Volumetric Soil Water Content	0.02	0.823
Biophysical	Rainfall	Cumulative Local Precipitation (mm)	0.11	0.170
Biophysical	Temperature	Land Surface Temperature (°C)	0.02	0.806
Biophysical	Veg. Stress	Evapotranspiration Anomaly	0.04	0.604

Table 1 presents the correlation analysis between environmental indicators and historical locust occurrence in the study area. The findings indicate that all the examined environmental and hyperspectral indicators exhibited weak relationships with historical locust occurrence, as reflected by the low correlation coefficients. In addition, all the p-values were greater than 0.05, indicating that the relationships were statistically insignificant. Among the indicators, Band_2 recorded the highest positive correlation with historical locust occurrence ($r = 0.14$, $p > 0.05$), followed by rainfall ($r = 0.11$, $p > 0.05$). This suggests a weak positive association between these variables and locust occurrence, although the relationships were not statistically significant. Band_1 and Band_4 also showed weak positive correlations of $r = 0.08$ and $r = 0.06$, respectively, while Band_3 and Band_5 exhibited almost no relationship with locust occurrence.

The vegetation indices NDVI ($r = -0.01$, $p > 0.05$) and SAVI ($r = -0.02$, $p > 0.05$) demonstrated very weak negative relationships with historical locust occurrence. Similarly, soil moisture and temperature recorded very weak positive correlations with the dependent variable. Vegetation stress showed a weak positive relationship ($r = 0.04$, $p > 0.05$), which was also statistically insignificant. The findings overall suggested that individual environmental and hyperspectral indicators may not independently explain locust occurrence within the study area. This implies

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 2 August 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 334-348
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

that locust infestation patterns are likely influenced by the combined interaction of multiple environmental and spectral factors rather than isolated indicators alone. The findings further justify the need for AI-driven analytical frameworks capable of detecting complex environmental indices of locust-prone areas.

Table 2. Logistic regression analyzing the environmental indicators

Variable	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.749	1.727	-1.012	0.311
NDVI	1.775	8.669	0.205	0.838
SAVI	-2.233	9.491	-0.235	0.814
Soil Moisture	0.461	1.034	0.446	0.656
Rainfall	0.006	0.004	1.455	0.146
Temperature	0.021	0.049	0.426	0.670
Vegetation Stress	0.096	0.148	0.646	0.518

Table 2 presents the binary logistic regression analysis examining the influence of environmental indicators on historical locust occurrence in the study area. The findings indicate that none of the environmental indicators significantly predicted historical locust occurrence since all the p-values were greater than 0.05. This suggests that the individual predictors did not independently exert a statistically significant effect on locust occurrence within the study region. Among the predictors, NDVI recorded a positive coefficient ($\beta = 1.775$, $p > 0.05$), implying that an increase in vegetation index slightly increased the likelihood of locust occurrence, although the relationship was statistically insignificant. Conversely, SAVI exhibited a negative coefficient ($\beta = -2.233$, $p > 0.05$), suggesting a weak inverse relationship with locust occurrence. However, the large standard errors for both NDVI and SAVI indicate substantial variability in the estimates.

Rainfall recorded a small positive coefficient ($\beta = 0.006$, $p > 0.05$), indicating that increasing rainfall marginally increased the probability of locust occurrence, though the effect remained statistically insignificant. Similarly, soil moisture ($\beta = 0.461$, $p > 0.05$), temperature ($\beta = 0.021$, $p > 0.05$), and vegetation stress ($\beta = 0.096$, $p > 0.05$) all demonstrated weak positive relationships with historical locust occurrence. Overall, the logistic regression findings suggest that no single environmental indicator independently explains locust occurrence in South Nyanza. This implies that locust infestation patterns may result from complex interactions among multiple environmental, climatic, and spectral factors. This justifies moving away from univariate monitoring paradigms and utilizing non-linear machine learning architectures capable of capturing multidimensional ecological interactions.

Table 3. Random Forest Classification

Metric	Value / Result
Total Evaluation Sample Size (n)	150 units
True Positives (TP)	58
True Negatives (TN)	55
False Positives (FP)	16
False Negatives (FN)	21
Overall Accuracy	75.33%

Out-of-Bag (OOB) Error Rate	24.67%
Precision (Positive Predictive Value)	78.38%
Recall / Sensitivity (True Positive Rate)	73.42%
Specificity (True Negative Rate)	77.46%
F1-Score (Harmonic Mean)	75.82%
Area Under ROC Curve (ROC-AUC)	0.753

Table 3 presents the Random Forest classification results for the AI-driven hyperspectral remote sensing model used in identifying locust-prone areas in South Nyanza, Kenya. The Random Forest algorithm was implemented as a classification model with 500 decision trees, indicating that the model generated predictions based on an ensemble of multiple trees to improve classification stability and predictive accuracy. The model utilized three variables at each split point (mtry = 3), meaning that three predictor variables were randomly selected during the tree-building process to determine the optimal split. As detailed in Table 3, the model demonstrated a moderate overall classification accuracy of 75.33%, coupled with an Out-of-Bag (OOB) error rate of 24.67%. To provide a transparent, rigorous validation of the framework's diagnostic capacity, shifting away from a reliance on raw accuracy alone, a complete multi-metric confusion matrix analysis was conducted based on the 150 georeferenced spatial observation units.

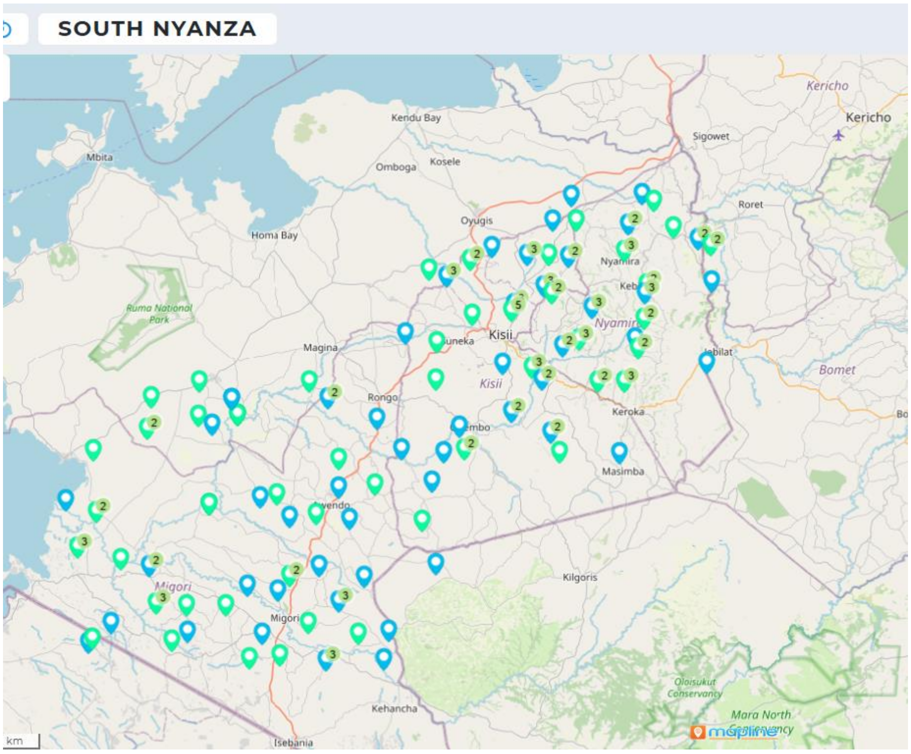

Figure 2: Locust prediction in South Nyanza

The model correctly classified 58 active locust breeding and feeding zones (True Positives) and 55 established non-infestation areas (True Negatives). Conversely, the system registered 16 False Positives (where benign environments were classified as active risk zones) and 21 False Negatives, where active infestation landscapes were missed by the ensemble (See Figure 2). These

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 2 August 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 334-348
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

distributions yielded a Precision of 78.38% and a Recall (Sensitivity) of 73.42%. The balanced alignment between precision and recall indicates that the model maintains a stable equilibrium between minimizing false alarms and ensuring comprehensive detection coverage, culminating in a robust baseline F1-Score of 75.82%. The derived Area Under the Receiver Operating Characteristic curve (ROC-AUC) was computed at 0.753. An AUC of 0.753 means there is a 75.3% chance that the Random Forest model will distinguish a true positive locust site from a true negative site if picked at random. It represents an acceptable level of discrimination for field research. This metric indicates a moderate discriminatory capacity, confirming that the framework possesses a statistically reliable capability to differentiate between suitable biological habitats and non-locust baselines across heterogeneous terrains.

While these results validate the practical utility of integrating remote sensing data within an AI pipeline, the 24.67% OOB error rate and the moderate AUC threshold necessitate a cautious, exploratory interpretation of the framework's immediate predictive readiness. The observed classification uncertainty is primarily attributed to the complex, fragmented ecological characteristics of the South Nyanza landscape, where mixed cropland covers, localized smallholder farming plots, and rapid phenological changes produce overlapping spectral signatures.

Crucially, when evaluated alongside the performance of traditional statistical techniques, the Random Forest model presents a vital operational breakthrough. As demonstrated by the Pearson correlation ($r \leq 0.14$, $p > 0.05$) and the binary logistic regression parameters ($p > 0.05$), individual biophysical or spectral indicators completely lack independent predictive power. Linear models failed to capture any statistically significant relationship due to the highly dynamic, non-linear dependencies governing locust breeding and migration.

Consequently, the 75.33% accuracy achieved by the Random Forest classifier demonstrates that locust habitat suitability cannot be isolated to singular variables like rainfall or vegetation indices alone. Instead, risk mapping relies on capturing complex, multidimensional, and non-linear interactions among soil moisture thresholds, localized temperature ranges, and narrow spectral properties. The framework therefore serves as a valuable exploratory diagnostic tool for regional agricultural extension services, shifting surveillance paradigms from purely reactive field scouting to an objective, data-driven early warning structure.

Table 4. Importance Ranking of Hyperspectral and Environmental Indicators

Input Predictive Indicator / Feature	Mean Decrease Accuracy	Mean Decrease Gini
Rainfall (mm)	2.75	9.87
Soil Moisture	2.59	8.83
Temperature (°C)	-1.41	8.6
NDVI	-0.88	7.02

SAVI	-1.99	6.27
Band 3 (Red, ~665 nm)	0.33	6.17
Band 4 (NIR, ~865 nm)	-0.33	5.83
Band 5 (WIR, ~1610 nm)	-3	5.71
Band 2 (Green, ~560 nm)	0.71	5.26
Band 1 (Blue, ~490 nm)	-0.71	4.25
Land Cover Type	-1.22	3.73
Vegetation Stress	-1.7	2.8

Table 4 reports the Variable Importance Scores generated from the Random Forest classification model. The findings indicate that Rainfall emerged as the most influential predictor with the highest Mean Decrease Gini score of 9.87, followed by Soil Moisture (8.83) and Temperature (8.60). This suggests that climatic and environmental conditions play a significant role in determining locust-prone areas within South Nyanza, Kenya. Among the hyperspectral indicators, NDVI and SAVI demonstrated relatively strong importance scores, indicating that vegetation health and density are closely associated with locust breeding and feeding habitats.

Additionally, Band_3 recorded the highest contribution among the spectral bands, suggesting that certain hyperspectral wavelengths are effective in distinguishing environmental conditions favorable for locust infestation. The results further revealed that some variables, such as Band 5, Vegetation Stress, and Land Cover Type, exhibited relatively lower predictive contribution, as reflected by negative Mean Decrease Accuracy values. Overall, the findings demonstrate that rainfall patterns, soil moisture conditions, temperature variations, and vegetation characteristics are the most critical indicators for AI-driven locust surveillance and classification.

V. CONCLUSION AND RECOMMENDATION

This study investigated the hyperspectral signatures and environmental variables associated with locust breeding and feeding zones in South Nyanza using an AI-driven analytical framework. The findings demonstrated that individual environmental indicators exhibited weak predictive capability, while integrated machine learning approaches significantly improved classification performance. The findings revealed that individual hyperspectral and environmental indicators such as NDVI, SAVI, rainfall, soil moisture, temperature, and vegetation stress exhibited weak and statistically insignificant relationships with historical locust occurrence.

These findings support the arguments by Thenkabail et al. (2012) and Mahlein et al. (2013), who emphasized that ecological monitoring systems require integrated environmental datasets rather than isolated indicators. The results further confirmed the assertions by Sahdev et al. (2020) and Ngugi et al. (2021) that conventional statistical methods may inadequately capture nonlinear ecological relationships associated with pest migration and infestation patterns. Consequently, the findings justified the relevance of AI-driven frameworks capable of modeling multidimensional environmental interactions.

The study sought to identify the key hyperspectral and environmental indicators associated with locust breeding and feeding habitats in South Nyanza, Kenya. To achieve this objective, correlation analysis, logistic regression, Random Forest classification, and descriptive statistical analyses were conducted using variables such as NDVI, SAVI, soil moisture, rainfall, temperature, vegetation stress, and hyperspectral bands. The correlation analysis findings revealed weak relationships between the environmental indicators and historical locust occurrence.

Among the variables, Band_2 recorded the highest positive correlation ($r = 0.14$, $p > 0.05$), followed by rainfall ($r = 0.11$, $p > 0.05$), although all relationships were statistically insignificant. Logistic regression findings further indicated that none of the individual indicators significantly predicted historical locust occurrence since all p-values exceeded 0.05. However, NDVI ($\beta = 1.775$) and rainfall ($\beta = 0.006$) showed weak positive relationships with locust occurrence. The Random Forest classification model achieved a classification accuracy of 75.33% with an Out-of-Bag (OOB) error rate of 24.67%, indicating moderate predictive capability. Variable importance analysis identified rainfall, soil moisture, and temperature as the most influential predictors based on Mean Decrease Gini values.

REFERENCES

Cressman, K. (2013). Role of remote sensing in desert locust early warning. Journal of Applied Remote Sensing, 7(1), 075098. https://doi.org/10.1117/1.JRS.7.075098

Cressman, K. (2016). eLocust3: The role of new technology in desert locust early warning systems. Food and Agriculture Organization of the United Nations.

Cressman, K., & Hodson, D. (2019). Surveillance, information sharing and early warning systems for transboundary plant pests. Outlooks on Pest Management, 30(6), 252–256. https://doi.org/10.1564/v30_dec_02

Food and Agriculture Organization. (2020). Desert locust upsurge: Progress report and response overview. Food and Agriculture Organization of the United Nations.

Food and Agriculture Organization. (2021). Desert locust guidelines: Survey. Food and Agriculture Organization of the United Nations.

Hunter, D. M. (2004). Advances in the control of locusts (Orthoptera: Acrididae) in eastern Africa. Journal of Orthoptera Research, 13(1), 25–34.

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

Khan, Z. R., Midega, C. A. O., Pittchar, J., Pickett, J. A., & Bruce, T. J. (2014). Push-pull technology: A conservation agriculture approach for integrated management of insect pests. Annual Review of Entomology, 59(1), 147–165.

Magor, J. I., Lecoq, M., & Hunter, D. M. (2008). Preventive control and desert locust plagues. Crop Protection, 27(12), 1527–1533. https://doi.org/10.1016/j.cropro.2008.08.006

Makini, F. W., Mose, L. O., Kamau, G. M., Makelo, M. N., Salasya, B. D., Mulinge, W. W., & Ong'engo, H. (2018). Status of agricultural extension and advisory services in Kenya. Kenya Agricultural and Livestock Research Organization.

Mongare, P. N., Ouma, G., & Ombui, J. (2023). Climate variability and locust outbreak dynamics in East African agricultural systems. African Journal of Agricultural Research, 19(4), 322–337.

Mubarak, S., Nyongesa, H., & Omwenga, E. (2022). Artificial intelligence adoption challenges in agricultural monitoring systems in developing countries. International Journal of Agricultural Informatics, 13(2), 45–59.

Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. Nature, 540(7633), 418–422. https://doi.org/10.1038/nature20584

Rembold, F., Atzberger, C., Savin, I., & Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. Remote Sensing, 5(4), 1704–1733. https://doi.org/10.3390/rs5041704

Shafiee, M., Babaei, M., & Karami, A. (2021). Artificial intelligence applications in agricultural remote sensing: A review of machine learning approaches. Remote Sensing Applications: Society and Environment, 22, 100493.

Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. Trends in Plant Science, 21(2), 110–124.

Thenkabail, P. S., Lyon, J. G., & Huete, A. (2012). Hyperspectral remote sensing of vegetation. CRC Press.

Van Mele, P., Bentley, J. W., & Guéi, R. G. (2010). African seed enterprises: Sowing the seeds of food security. CAB International.

Wang, L., Qu, J. J., Hao, X., & Zhu, Q. (2010). Sensitivity studies of vegetation indices to environmental changes. Remote Sensing of Environment, 114(5), 1070–1086.

World Bank. (2021). Emergency locust response program implementation report for Kenya. World Bank Group.

Yu, H., Zhang, M., Wang, J., & Liu, X. (2021). Machine learning approaches for ecological monitoring using hyperspectral remote sensing data. Ecological Informatics, 64, 101365.

Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. Precision Agriculture, 13(6), 693–712.