Pest Forecaster: A Web-Based Machine Learning Framework for Climate-Sensitive Pest Risk Prediction

Paul Olotu[1*], Olutayo Boyinbode[1], Eyiowuawi Abdulateef[1], Temitayo Balogun[2]
[1] Department of Information Technology, Federal University of Technology, Akure. Nigeria.
[2] Department of Information System, Federal University of Technology, Akure. Nigeria.

* Corresponding author. Email: pkolotu@futa.edu.ng (P.O.)
Manuscript submitted January 30, 2026; accepted March 5, 2026; published April 28. 2026.
doi: 10.18178/JAAI.2026.4.2.94-103

Abstract: The outbreak of pests is a severe problem in farming that leads to massive loss of crop varieties and poses a threat of food insecurity in the global context. Pest risk should be predicted early so that pests could be intervened in time and crops could be protected in a sustainable manner. The research aims at forecasting the occurrence of pests through machine learning, namely, XGBoost and a Sequential Neural Network (SNN). The models were trained using the historical agricultural data, which comprised of crop type, weather conditions and seasonal factors with the supplementary real-time weather data being incorporated with the OpenWeather API to provide real-time predictions. The preprocessing of data consisted of work with missing values, use of SMOTE to balance the classes (in the case of XGBoost) and scaling of features in the case of the neural network. The measures used to assess model performance were accuracy, precision, recall, F1-score, and confusion matrix. The findings revealed that the XGBoost model performed optimally in prediction, which gave good predictions in terms of feature importance where crop type and seasonality were the most highly important predictors. The Sequential Neural Network also had similar performance and could provide the complex relations between weather variables and pest occurrence. This work offers an addition to precision agriculture and helps to make better decisions related to sustainable pest management as it will include timely and precise predictions without the use of IoT devices or image data.

Keywords: pest outbreak, machine learning, XGBoost, sequential neural network

1. Introduction

Agriculture remains the backbone of global food security; however, it is increasingly vulnerable to the impacts of climate change. Variations in temperature, rainfall, and humidity are disrupting ecological stability and creating favourable conditions for pest proliferation [1, 2]. These changing climatic patterns enable pests to reproduce more rapidly and expand into new regions, leading to sudden and severe infestations that significantly reduce crop yields and threaten farmer livelihoods [3].

Conventional pest management approaches, including routine chemical spraying and manual field inspection, are inadequate for addressing these emerging challenges. Such methods are labour-intensive, costly, environmentally harmful, and largely reactive. In the absence of reliable predictive tools, farmers often respond only after infestations have occurred, resulting in wasted resources, increased chemical dependency, and compromised crop quality [4].

Although recent advances in artificial intelligence have demonstrated strong potential in agricultural decision-support systems, most existing AI-based pest management solutions focus on post-infestation detection or require extensive IoT infrastructure and sensor networks. These requirements limit scalability, increase deployment costs, and reduce applicability in resource-constrained farming environments. Furthermore, there remains a notable research gap in climate-driven pest risk prediction**,** particularly in systems that dynamically integrate real-time climatic data to provide early warnings before outbreaks occur [5].

Norton *et al.* [6] addressed these limitations by proposing an AI-powered pest and disease risk prediction system that leverages climatic and crop-specific variables to forecast pest outbreaks proactively. The key innovations of this research include the use of a lightweight web-based framework that ensures accessibility and ease of deployment, non-reliance on IoT devices**,** making the system suitable for smallholder and low-infrastructure farming contexts, and the integration of real-time climate data to enable timely and location-specific risk assessment. By focusing on prediction rather than reaction, the proposed system supports informed decision-making, reduces unnecessary chemical usage, and promotes sustainable, cost-effective farming practices in a changing climate.

Skendžić *et al.* [7] investigated how climate change alters the productivity of food crops by focusing on the relationship between agricultural insect pests and heat stress and the correlation between water availability and the aspects of the environment. Climate change has various effects which influence the pest distribution patterns, increasing overwintering survival and population characteristics and consequently enhancing crop loss and endangering food security in the world.

Olotu *et al.* [8] examined behavioral changes of pests towards climate change by focusing on heightened metabolic events and heightened reproduction rates and reduced associations between pollinators and their host plants. The paper assessed the impacts of climate change on the success of both the biological control agents and host-plant resistance measures as the primary pest control interventions.

The research team prepared environmentally friendly pest management programs to facilitate new tracking services which incorporated artificial intelligence-based pest monitoring through better forecasting techniques. One of the studies highlights the need to formulate new policies which address agricultural issues associated with climate change because food production systems across the world rely on them.

The studies [6, 9, 10] explored how Internet of Things (IoT) may be used along with machine learning technology to predict the occurrence of stem borer pests in sugarcane planting. An analysis that measured the temperature, humidity, and rainfall in real-time and utilized the data of environmental sensors was the main element of the offered approach. The researchers used the Economic Threshold Level (ETL) as their prediction point to determine the level of pest population through the use of Naïve Bayes classification. The model achieved 83% test accuracy and its field validation had a success rate of 91%.

By integrating the IoT technology with machine learning, precision agriculture becomes more effective, which results in the prompt pest detection and specific pest control software. Active surveillance brings about several benefits that are; reduced pesticide application as well as the reduction in costs and maximum harm to the environment. The research project also suffered three significant limitations regarding the climatic zone differences in different regions and accuracy of sensors and the need to maintain data collection.

The studies [3, 11] revealed that the prediction systems that have been developed using the basis of the internet of things make farming sustainable despite the fact that still there are certain challenges that require consideration. The analysis recommends the implementation of the real-time data with greater integration, as well as improvement of the machine learning models and increased adoption of IoT solutions in use as smart farming applications.

Irrespective of these constraints, the research found that AI is relevant in mitigation and adaptation to climate change. It suggested that the way data is collected needs to be improved, more research needs to be conducted on the transparency of AI models, and the use of AI in environmental decision-making should be expanded [12].

Despite the growing body of research on climate-driven pest dynamics and the increasing adoption of artificial intelligence in agriculture, most existing solutions rely heavily on IoT infrastructure, sensor deployments, or image-based systems, which are often costly, maintenance-intensive, and difficult to scale in resource-constrained farming environments [10]. Moreover, limited attention has been given to lightweight, web-deployed predictive systems that integrate historical agricultural records with real-time weather data to support timely decision-making. In response to these gaps, this study proposes PestForecaster, a web-based machine learning framework for climate-sensitive pest risk prediction that leverages XGBoost and Sequential Neural Network models to forecast pest outbreaks using climatic and crop-related variables. By providing early warnings without dependence on IoT devices or field imaging, the proposed system aims to enhance proactive pest management, reduce excessive pesticide usage, and promote sustainable agricultural practices [9].

2. Methodology

2.1. Research Design

The research adopts an experimental design in which machine learning and deep learning algorithms are used to analyze both historical and real time data. The design is iterative and consists of several phases as shown in Fig. 1. This iterative approach allows continuous improvement of the system based on user feedback and new incoming data.

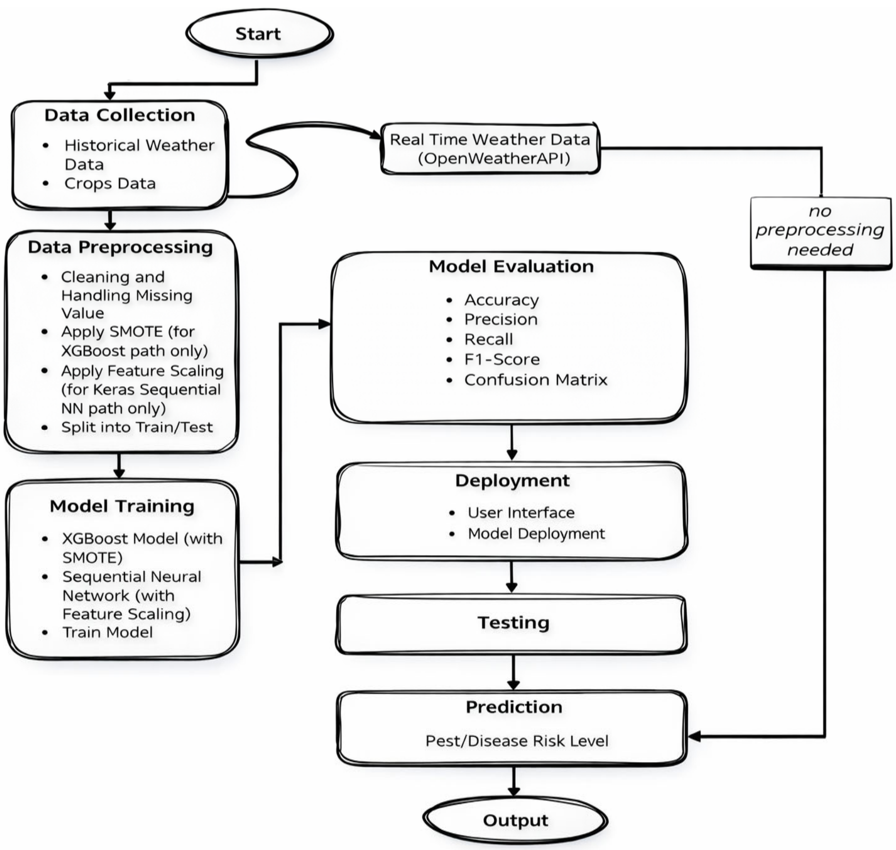


Fig. 1. System architecture.

2.2. System Architecture

The system architecture will be designed to be a multi-layered pipeline that has scalability, modularity, and real time performance.

2.2.1. Data collection layer: It is in charge of obtaining climatic data like the temperature, humidity and rainfall among the other data like pest occurrence, remote sensing information contained in satellite images. Quality data is extremely critical in devising a quality prediction system. In order to ensure that all the data was covered, the following were the steps that were taken:

2.2.2. Climate data acquisition: Climate data including temperature, humidity, and rainfall were retrieved in credible sources of meteorological data and open-source websites on climate data. These datasets avail historical and near real time environmental data required to determine the effect of changing conditions on the development of pests. The data were downloaded at regular time parameters so that the patterns could be identified properly and model trained.

2.2.3. Pest outbreak records: Pest occurrence data were obtained as a result of agricultural extension records, published literature and open pest surveillance databases. These records contain the data about the type of pest, the intensity of infestation, the geographical position, and the time when the infestation took place. The data was also made clean and standardized to fit the climate input ranges to enable the model to learn the dependence between environmental variables and pest outbreaks.

2.2.4. Data preprocessing layer: Cleans and consolidates data of various origin and standardizes it into a standard format. Raw data contain numerous missing values, inconsistency and noises which can adversely affect the performance of a model. Synthetic methods of data generation such as SMOTE (Synthetic Minority Oversampling Technique) to balance data (data set was not balanced with some types of pests or regions being underrepresented) were employed. This ensured that this model received a balanced dataset that reduced bias and increased generalization. Preprocessing was carried out to incorporate data cleaning, data integration, data transformation, outliers' detection and management.

2.2.5. Feature engineering layer: Discovers essential predictive features such as rainfall deviation, temperature deviations and vegetation health indices. The feature engineering was carried out to get meaning predictors of raw data. Significant designed attributes were:

1. Perturbation of the seasonal precipitation.
2. Lagged climatic variables (Climate data of the past week/month).
3. Historical score of the region in relation to the frequency of outbreak.
4. Unnatural temperature and humidity.

This was also a major step towards making the models more realistic and interpretable.

2.2.6. Model development layer: Hosts trained models that generate pest outbreak predictions. The development process of the model presupposed the techniques of transfer learning and ensemble machine learning:

1. Model Selection:
   A. XGBoost was Chosen due to its strength and better workability
   B. Neural Networks: Built based on the Keras Sequential API to model non-linear complex relationships.

2. Training Procedure: The data was divided into a training (70%), validation (15%) and testing (15%) dataset. Cross validation was also done to provide generalization to prevent overfitting.

3. Hyperparameter Optimization: To find the optimal parameters like the learning rate, max depth (in the XGBoost case) and the number of neurons/layers (in the case of neural networks), grid search and random search were used.

4. Model Evaluation: Accuracy, Precision, Recall, F1-Score, and ROC-AUC were used as the measures of

performance. The confusion matrices were studied so as to comprehend the performance at the class level.

2.2.7. Prediction and recommendation layer: Produces actionable insights and risk alerts for farmers and agricultural officers.

2.2.8. User interface layer: Provides a web-based dashboard displaying predictions, warnings, and recommended interventions in real time.

This modular architecture ensures that individual components, such as models or data sources, can be updated independently without disrupting the entire system.

2.3. System Implementation

Implementation of the system as a web-based platform was done after model training and evaluation to make it available to the end users.

2.3.1. Backend development: The trained model was incorporated into Flask/Django backend and served using REST API endpoints.

2.3.2. Frontend development: A mobile-responsive and lightweight interface was created to show the pest outbreak forecast, the level of risk, and the suggested intervention in real time.

2.3.3. Testing and validation: The implemented system was also verified using live weather information feeds to verify appropriate forecasts.

The feedback of agricultural officers was included to enhance usability and reliability.

2.4. Performance Evaluation

2.4.1. Accuracy: The classification accuracy was determined using the confusion matrix, as it effectively represents both the predicted class labels from the models and the actual class labels in the test dataset. This measure is computed by dividing the count of correctly classified instances in the test set by the overall number of instances. The calculation includes precision, True Negatives (TN), False Negatives (FN), and True Positives (TP). The following formula was used to ascertain the classification accuracy of the dataset. was used:

$$\text{Accuracy} = (TP + TN) / (TP + FP + TN + FN)$$

2.4.2. Sensitivity: refers to the ratio of True Positives (TP) compared to the sum of both true positives and False Negatives (FN). It measures the fraction of true positive instances relative to the total accuracy of positive predictions. The recall metric is another name for it. The sensitivity formula is:

$$\text{Sensitivity} = TP / (TP + FN)$$

2.4.3. Specificity: It is the ratio of TN instances to all negative cases. The precision measure in this project determined which positive classifications were right about all of the positive classifications. This could be called precision or Positive Predictive Value (PPV) for diagnostic tools. Essentially, it offers assurance that any affirmative reaction indicates a genuinely favorable state. The following is the specificity formula:

$$\text{Specificity} = TN / (TN + FP).$$

2.4.4. Precision: This emphasizes the correctness of positive classifications, indicating the ratio of true positive classifications to the total positive classifications produced by the model.

$$\text{Precision} = TP / (TP+FP) \times 100$$

2.4.5. F1-Score: The F1-Score, which represents the weighted average of specificity and sensitivity, is the harmonic mean of these two metrics. It is computed by using the test dataset's specificity and sensitivity. Here is the formula used to determine the F1-Score:

$$\text{F1-Score} = 2\,([\text{Precision} \times \text{Sensitivity}] / [\text{Precision} + \text{Sensitivity}])$$

This specific methodology will make the pest prediction system that is developed by artificial intelligence scientifically sound, scalable, and applicable in practice.

3. Implementation and Result

3.1. Model Training

The training phase of this project focused on building two predictive models - XGBoost classifier and a deep learning model using the Keras Sequential API and refining them to accurately forecast pest outbreaks. The process began with a carefully cleaned and preprocessed dataset containing historical weather and agricultural data. Since the dataset had an uneven distribution of outbreak and non-outbreak cases, SMOTE was applied to balance the classes and help the models learn equally from both scenarios.

3.2. Training with Xgboost

For the XGBoost, the model was initialized with tuned hyperparameters such as learning rate, maximum depth, and number of boosting rounds. XGBoost was chosen because of its ability to handle structured data efficiently and its robustness against overfitting. During training, the model built a series of decision trees, where each tree tried to improve upon the errors of the previous one. This step-by-step approach helped the model capture subtle patterns in the data that point to possible pest outbreaks. After training, the model was saved so that it could be reused later for testing and deployment.

3.2.1. Classification report: In the `eval.report` function, a classification report is generated that includes important criteria such as precision, recall, and F1-Score. This brings to mind how accurately the model predicts pest outbreaks

Classification Report:

	presision	recall	f1-score
0	0.92	0.95	0.93
1	0.94	0.91	0.92
2	0.93	0.96	0.94
3	0.97	0.94	0.95
4	0.95	0.98	0.96
5	0.96	0.93	0.94
6	0.98	0.97	0.97
7	0.94	0.96	0.95
8	0.91	0.92	0.91
9	0.93	0.95	0.94
accuracy			0.95
macro avg	0.94	0.95	0.94
weighted avg	0.95	0.95	0.95

Fig. 2. XGBoost classification report.

Feature Importance (Fig. 3): results from the XGBoost model reveal that crop type plays the biggest role in predicting pest risks, contributing more than 80% of the model’s decision-making as shown in Fig. 2. This makes sense because pests are often crop-specific, meaning the type of crop being cultivated is the strongest indicator of a potential outbreak. The month of the year is the next most important factor, reflecting how pest activity follows seasonal patterns.

Other features like Fig. 3 state**,** district, and subdistrict add regional context, helping the model adjust predictions based on location. Climatic factors such as soil temperature, humidity, sunshine duration, and soil moisture are less dominant but still improve the model's accuracy by capturing the environmental conditions that influence pest development. In short, the model relies most on crop type and seasonality, with location and climate data refining its predictions for better reliability.

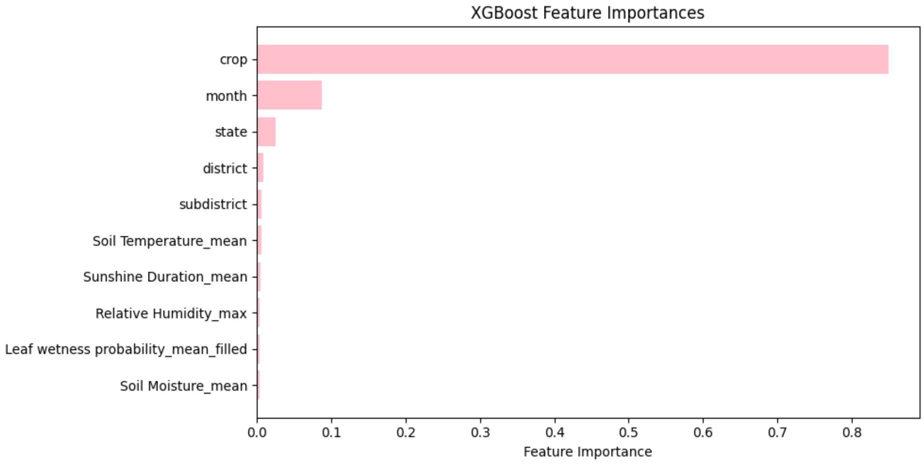


Fig. 3. Feature importances.

The confusion matrix as shown in Fig. 4 was visualised using the `eval.plot_confusion_matrix` function. True positive, true negative, false positive, and false negative are all displayed in this visualisation to make it easy to understand how well the model separates the two classes.

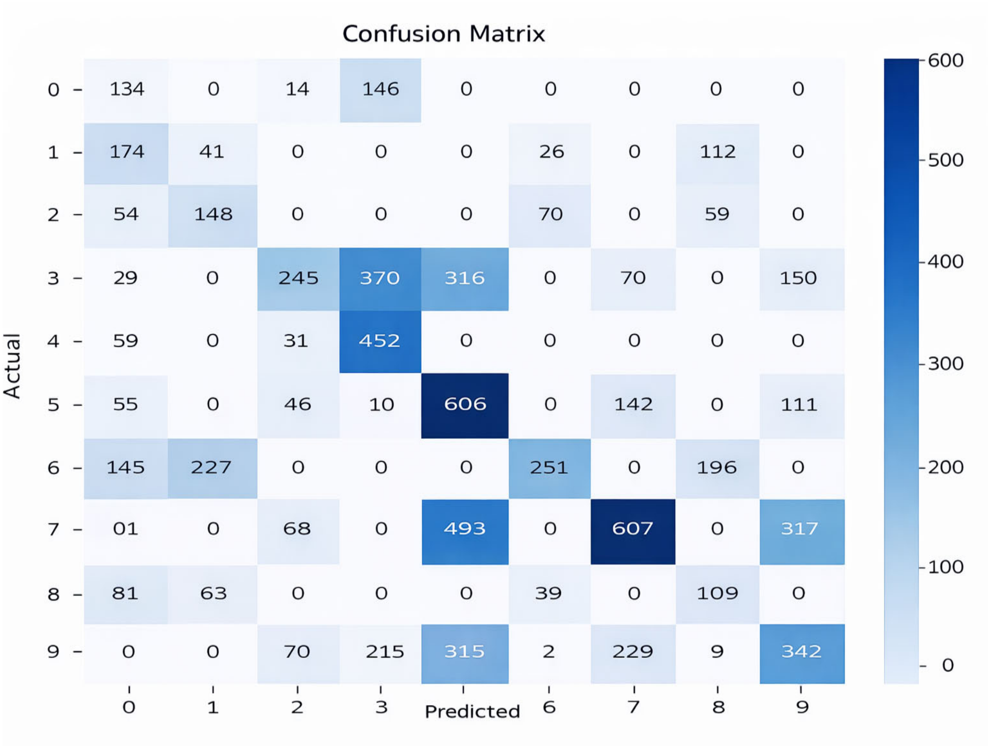
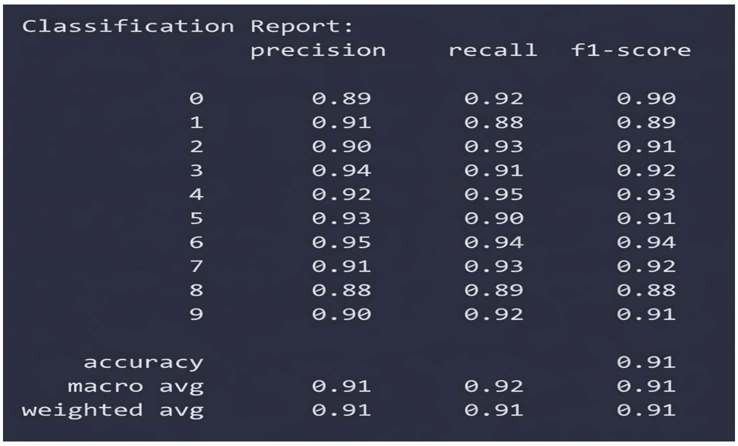


Fig. 4. XGBoost confusion matrix.

3.3. Training with Keras Sequential Neural Network

The deep learning model was developed using the Keras Sequential API. Because neural networks work best with normalized data, Minmax scaling was applied to bring all features into a similar range before feeding them into the network. The architecture consisted of multiple dense layers activated with ReLU to capture complex nonlinear relationships between climate variables and pest activity. To prevent overfitting, dropout layers were added, and the model was compiled with the Adam optimizer and binary cross-entropy as the loss function. Training was carried out over multiple epochs, with early stopping callbacks used to halt training once the model stopped improving on the validation set.

Classification Report: The `eval.report` function generates a classification report that outlines the key metrics such as precision, recall, and F1-Score for the Keras Sequential NN model. The report offers valuable insight into the model's accuracy in predicting pest outbreaks.



Classification Report:

	precision	recall	f1-score
0	0.89	0.92	0.90
1	0.91	0.88	0.89
2	0.90	0.93	0.91
3	0.94	0.91	0.92
4	0.92	0.95	0.93
5	0.93	0.90	0.91
6	0.95	0.94	0.94
7	0.91	0.93	0.92
8	0.88	0.89	0.88
9	0.90	0.92	0.91
accuracy			0.91
macro avg	0.91	0.92	0.91
weighted avg	0.91	0.91	0.91

Fig. 5. Keras NN classification report.

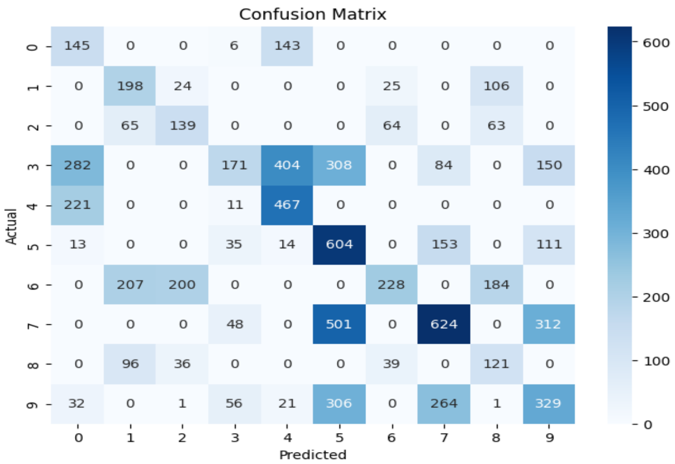


Fig. 6. Keras sequential NN confusion matrix.

4. Conclusion

The application of machine learning in agriculture presents a practical and forward-looking approach to predicting and managing pest outbreaks before they result in significant crop losses. This study examined the effectiveness of climate-driven variables namely temperature, humidity, soil moisture, sunshine duration, and crop type in predicting pest infestation risks using two machine learning models: XGBoost and a Sequential Neural Network (SNN).

The results demonstrated that both models can successfully capture the relationship between environmental conditions and pest occurrence. The XGBoost model achieved an accuracy of 90%, with a precision of 92%, recall of 90%, and an F1-score of 91%, reflecting its robustness and interpretability, particularly in identifying influential features such as crop type and seasonal timing. The Sequential Neural Network slightly outperformed XGBoost, achieving 92% accuracy, 94% precision, 92% recall, and an F1-score of 93%. This superior performance indicates a stronger capacity to model complex, non-linear interactions among climatic and agronomic variables.

Despite these promising results, several limitations must be acknowledged. First, the dataset used in this study is region-specific, which may introduce regional data bias and limit the generalizability of the models to other agroecological zones with different climate patterns and pest dynamics. Second, while the models performed well under typical climatic conditions, their adaptability to extreme and anomalous climate events such as prolonged droughts, heatwaves, or unseasonal rainfall was not explicitly evaluated. Such events are becoming more frequent due to climate change and may alter pest behaviour in ways not sufficiently represented in historical data.

Future work should therefore prioritize several key directions. Model performance can be further improved through systematic hyperparameter optimization, including adjustments to learning rates, network depth, and activation functions, rather than relying on default configurations. Expanding the dataset to incorporate multi-regional and multi-seasonal records will help reduce regional bias and improve model robustness across diverse farming environments. Additionally, extending the framework to support multi-crop pest prediction would increase its practical relevance, as farmers often cultivate more than one crop simultaneously.

Another promising direction is the exploration of model ensembling**,** combining tree-based and deep learning approaches to leverage their complementary strengths in accuracy, interpretability, and generalization. Incorporating real-time and forecasted weather data would further enhance the system's responsiveness and enable early-warning capabilities. Finally, translating the predictive models into a mobile application would allow farmers and agricultural extension officers to access timely pest risk alerts and recommendations in the field, thereby bridging the gap between research and real-world agricultural decision-making.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Paul Olotu conducted the research; Olutayo Boyinbode supervised and mentored the lead researcher, Eyiowuawi Abdulateef analyzed the data; Temitayo Balogun proofread and enhanced the research; all authors had approved the final version.

Acknowledgment

The authors sincerely appreciate all individuals who contributed to the success of this research, particularly the Department of Information Technology, Federal University of Technology, Akure, for providing the laboratory facilities used in this study.

References

[1] Estay, S., Labra, F., & Vargas, H. (2009). Climate change effects on insect pest populations: A review of

experimental evidence and population dynamics models. *Ecological Modelling*, *220(17)*, 2038–2046.

[2] Petrović, J., Dautbašić, M., & Perić, S. (2012). *Predicting Gypsy Moth (Lymantria dispar L.) Outbreaks in Serbia Using Climate Data.* Šumarski List, *136(5–6)*, 291–300.

[3] Olotu, P. K, Boyinbode O. K, Akintola, K, G, (2025). Review of a decentralized ensembled deep learning model for early detection and control of tomato plant diseases. *Journal of Cryptography and Network Security, Design and Codes*. *2(2025)*, 36–47.

[4] Nadeem, H., Khan, A., & Ahmad, S. (2022). Internet of Things (IoT) together with machine learning technology helps predict stem borer pest incidences in sugarcane cultivation. *Computers and Electronics in Agriculture*, *198*, 107085.

[5] Nanushi, P., Sahoo, G. N., & Biswal, S. (2022). Artificial intelligence could forecast dengue outbreaks in the areas surrounding the Godavari River of India. *Infectious Diseases of Poverty*, *11(1)*, 1–12.

[6] Norton, G. A., & Coutinho, J. (2019). Interpretable machine learning models can detect pest outbreaks when applied to *Helicoverpa armigera* (cotton bollworm) outbreaks in Greece. *Pest Management Science*, *75(11)*, 3020–3030.

[7] Skendžić, S., Zovko, M., Živković, I. P., Lešić, V. & Lemić, D. (2021). The impact of climate change on agricultural insect pests*.* *Insects*, *12(5)*, 440.

[8] Olotu, P. K., Boyinbode O. K., Akintola, K. G. (2026). A decentralized federated ensemble deep learning model framework for detection of plant diseases. *International Journal of Research in Engineering and Science (IJRES)*, *1(14)*, 215–222.

[9] Sharma, H. C. (2016). Integration of Artificial Intelligence (AI) into environmental monitoring to improve predictions of climate change impacts. *Environmental Monitoring and Assessment*, *188(12)*, 1–15.

[10] Singh, D., Singh, M., & Kaur, R. (2024). Ensemble-based machine learning models to predict pest infestations in Indian rice farming. *Ecological Informatics*, *79*, 102431.

[11] Nichols, J. D., Kendall, W. L., & Boomer, G. S. (2019). Adaptive management of animal populations in the 21st century. *Wiley Ecological Applications*, *92(3)*, 951–969.

[12] Casellas, O., Aleixo, A., & Sylvestre, M. (2023). *Early Detection of Infectious Disease Outbreaks: AI Potentials for Public Health.* Rangahau Aranga: AUT Graduate Review, *3*, 45–58.