

Explainable Climate-Aware Framework for Multi-Horizon Rice Pest Infestation Forecasting and Pest Risk Assessment Using Dual Regression–Classification

Aripirala Venkata Sai Pavan¹, B.Thimma Reddy², J.Swami Naik³

¹ Student, *Department of Computer Science and Engineering G Pulla Reddy Engineering College(Autonomous) Kurnool-518002, Andhra Pradesh, India*

^{2,3} Associate Professor, *Department of Computer Science and Engineering G.Pulla Reddy Engineering College(Autonomous) Kurnool-518002, Andhra Pradesh, India*

Abstract—Accurate and early prediction of rice pest infestation is vital to proactively protect the rice crops and to achieve good pest management. Current pest forecasting methods, however, typically use only a small amount of climatic data and make predictions at a single time into the future, and the results often are difficult to interpret for effective field decision making. The present study introduces an Explainable Multi-Horizon Early Warning Framework using a dual framework of regression and classification, which combines features derived from climate information, temporal pest history, multi-horizon forecasting, risk classification and Explainable Artificial Intelligence (XAI). The framework was tested with the Rice Pest Dataset with the evaluation based on infestation data for Yellowstemborer (YSB) and temporal and climatic data. The features were engineered using derived climate variables (Mean temperature, Temperature range, Mean relative humidity, Relative humidity difference in seasons), location encoding, temporal lag features and rolling features of pest intensity. A Random Forest regression model was trained to forecast the pest intensity at forecasting horizons of T+1, T+2 and T+3, and a Random Forest classifier was developed to classify the pests infestation state as Low, Medium or High risk. The regression model achieved an MAE of 0.8301, RMSE of 1.2020, and R² of 0.7801 for T+1. The models were found to be useful for predicting the values across the different forecasting horizons as their R² values were 0.7346 for T+2 and 0.6987 for T+3. The risk classification model performed well with a macro F1 score of 0.84 and overall accuracy of 84%, which shows a good discrimination between the 3 risk levels. The dual framework offers numerical pest-intensity forecasts that can be beneficial for research and agricultural professionals and translate infestation conditions and into easily interpretable risk categories

that are valuable for farm and decision-oriented pest management. Moreover, the model transparency was enhanced via SHAP analysis, and the most important factors affecting predictions were identified. The results indicated that recent pest history and trends of infestation, such as Pest_Log_lag_1 and pest_roll_4, were among the most influential features and other historical pest and climatic features had lesser influences. Overall, the proposed framework is based on multi-horizon prediction, risk-oriented classification, and explainable machine learning (X-ML), to develop an interpretable early warning mechanism that can facilitate both technical analysis and practical decisions for pest management in rice production.

Index Terms—Random Forest, Xtreme Gradient Boosting, Explainable Boosting Machine, Recurrent Neural Network, K-Nearest Neighbors, Artificial Neural Networks, Support Vector Regression, Absolute Log Error, Yellow Stem Borer, Root Relative Square Error, pest early warning system, farmer-oriented decision support.

I. INTRODUCTION

Rice is one of the important food crops and is a significant contributor to global food security. Rice production, however, is seriously impacted due to insect pest infestation, which can lead to loss of yield and quality of crop. Yellow Stem Borer is one of the important rice pests which can inflict a significant damage to rice plants. Hence, the need for timeous prediction of pest infestation for early monitoring and effective pest-management practices.

Pest monitoring is primarily conducted using field observation, traps, historic records and experience.

These methods give useful information, but can be time consuming and may not give enough warning of pest outbreaks. Machine learning (ML) has potential to make use of the climatic and historical data to detect patterns and facilitate automated pest prediction. Climate, including temperature, humidity, and rainfall, can affect insect development and population dynamics, and past observations of pests can be useful in determining past changes in pest populations.

Machine learning and explainable AI have been shown to be effective in rice pest predictions in recent studies. The study 'A generalizable and interpretable model for early warning of pest-induced crop diseases using environmental data' by Wadhwa et al., and Malik et al., 2024 [1] used environmental data from various parts of India to predict three classes: Green Leafhopper, Yellow Stem Borer and No Pest, employing a machine learning based model that is interpretable. They analyzed climatic factors like temperature, humidity and precipitation, and created more environmental features to enhance the prediction power. The models, evaluated using machine learning techniques, were tuned to a CatBoost model, which was able to perform classification with good success. SHAP was also used to determine the relative importance of the environmental factors for pest occurrence. The results show that machine learning and explainable AI can be applied to create agricultural early-warning systems.

The currently used method in [1] however is mostly about classification of the presence of pests based on environmental data. It doesn't cover the prediction of pest intensity for future periods of time. Past pest activity and/or lagged environmental effects can also affect pest populations. So, the historical information on pest observation, climatic lag characteristics, seasonal trends and recent infestations can be added to help forecast future pest conditions. Furthermore, predicting pest levels at several time horizons can be used to give more time to warn before the pest appears.

A second important issue is the ability to interpret numerical predictions of pests. A regression model could give a numerical value of the pest-intensity that is useful to agricultural researchers and experts. But farmers might not be able to easily understand the numerical counts of pests and the pest condition

might not be directly related to the numerical count of a pest species. This means that a representation of pest infestation that is more straightforward with respect to risk and that can facilitate practical decision making is needed.

To tackle these challenges, in this study, a dual regression–classification based Explainable Multi-Horizon Early Warning Framework (EWF) for Pest Infestation Prediction (PIP) is proposed. The regression component forecasts the intensity of the pest at the three forecasting horizons, T+1, T+2 and T+3, which correspond to one-, two-, and three-week-ahead forecasts. In the classification component, the proposed Pest Risk Index (PRI) is introduced, which will be a tool for converting the conditions of pest infestation to the categories of Low, Median and High risk. The aim of the PRI is to give a farmer a readable and meaningful assessment of the pest problem, without having to comprehend the numerical pest counts or the sophisticated output from machine learning.

The proposed framework also includes features derived from the climate data, seasonal data, historical lag data, and rolling window data to account for environmental conditions and for historical pest behaviour. Moreover, using SHAP-based explainability, important factors that contributed to the pest-intensity prediction are identified. Therefore, the framework delivers complementary outputs: the future intensity of the pest, the understandable risk of the pest, and explanations of prediction factors with SHAP.

The significant contribution of this study is:

- The multi-horizon forecasting method for forecasting the intensity of pests at T+1, T+2, and T+3.
- A dual regression–classification system of quantitative pest-intensity prediction and risk-level classification.
- A proposed Pest Risk Index (PRI) in the simplified farmer-friendly interpretation of Low, Medium and High risk.
- Climatic, seasonal, historical lag and rolling features integrated to include environmental and temporal pest patterns.
- Use of SHAP-based explainability in integration to find the most important factors affecting pest predictions.

The proposed framework overall extends the existing pest prediction based on machine learning, integrating multi-horizon forecasting, risk-oriented interpretation and explainable AI into a single early-warning system. It seeks to move the needle from technical, pest-intensity forecasting to more practical, farmer-pest-risk based decision making.

II. LITERATURE REVIEW

A machine-learning model for classification of three classes of Green Leafhopper, Yellow Stem Borer and No Pest based on environmental factors like temperature, humidity and rainfall was proposed in the study [1] which was easy to interpret. To arrive at the best performance, feature engineering was used before comparing various ML models, where CatBoost performed the best. The environmental factors that affected the occurrence of the pests were also identified with SHAP. The study showed that ML and XAI are helpful methods for early detection of pests.

The study in [6] studied the use of machine learning approaches to implement automatic pest detection and forecasting from insect traps outfitted with sensors. The study in this paper examined the feasibility of combining camera based pest observation systems with the environmental parameters (air temperature and air humidity) to predict the occurrence of *Helicoverpa armigera*. The idea of the time windowed data set was introduced and gave a substantial increase in prediction accuracy due to considering the temporal patterns in pest development. The results showed that the system was able to effectively decrease the amount of manual monitoring needed to increase the accuracy of early pest detection systems using sensor technologies, climatic data, and machine learning algorithms.

In this context, to solve the different difficulties faced in predicting the target values with high accuracy, the Multiple Machine Learning approach, namely the Model Ensemble Learning was recommended in [12] and it was made up by four machine learning models: Artificial Neural Networks (ANN), Support Vector Regression (SVR), K-Nearest Neighbors (KNN), and Random Forest (RF). The whole framework of this approach is different in modeling a contribution as using adaptive weights based on the prediction error, which is contrary to the conventional ensemble

method. The approach developed was able to predict the infestation of rice crop by YSB with an acceptable accuracy by incorporating climatic variables such as temperature, rainfall and relative air humidity with the approach. The study showed the effectiveness of good quality of DYN-ENS to increase the prediction power and prediction robustness of the agricultural applications.

In [4] the authors have developed an Explainable Deep Learning (EDL) model for forecasting pest outbreaks, called Attention-based Long Short-Term Memory Interaction Convolutional Neural Network (ALIC). It brought in attention mechanisms, Long Short-Term Memory (LSTM) networks, and interaction convolutional layers were incorporated, achieving effective modeling of temporal dynamics and feature interactions in the meteorological and historical data on pests. Moreover, feature importance analysis was also inbuilt, which helped in understanding the model and identified the topmost factors in the environmental factors that led to the emergence of unwanted. The findings of the experiments suggested the possibility to enhance the forecasting accuracy when compared with other existing benchmark models, indicating the potential of explainable deep learning processes in precision agriculture.

The drawbacks of the traditional iterative forecasting, so a novel multi-step forecasting model named Varied Length Mixture (VLM) model was introduced in [11] was introduced. A few different length autoregressive predictors were used as components of the model and they were combined with a probabilistic mixture model. This had been successful in keeping the temporal dependence and preventing it from building up over longer forecast periods. The VLM framework was proven to be the most successful method when applied to long-term forecasting scenarios on synthetic and actual data sets and it can also be applied in agro-forecasting prediction systems that rely on climatic scenarios, which require an early warning were crucial.

To provide extensive analysis of various applications of machine learning techniques towards the detection, classification and prediction of the crop disease and pest, [2] performed a review study. Various methodologies such as Artificial Neural Networks, Support Vector Machines, Convolutional Neural Networks and ensemble learning were discussed and

studied. The review noted the trend of moving in the direction of technology and automation in agriculture, replacing hairy manual monitoring methods. It also discovered that AI and machine learning tools have vast potential in the field of pest and disease control as they can greatly enhance the accuracy of predictions, resource utilization, reduce dependence on pesticides, and contribute to sustainable agriculture practices.

In [13] an explainable machine learning model for the prediction of olive fly and red spider mite infestations has been developed in olive groves and vineyards. To enhance transparency in model predictions, the study used a combination of classification and regression models, and featured explainable artificial intelligence (XAI) methods such as SHapley Additive exPlanations (SHAP) values and Individual Conditional Expectation (ICE) plots. The proposed methodology not only yielded high level of predictive performance, but also helped to identify the environmental variables that had the highest association with pest outbreaks. The findings showed the potential of using explainable AI techniques to improve the interpretability and usability of machine learning models for pest forecasting.

The study conducted in [15] suggested a machine learning model, suitable for human interpretation that predicted the incidence of *Helicoverpa armigera* on cotton cultivation. The model incorporated data on vegetation indices along with meteorological and insect trap data to produce forecasts of pest occurrences. One of the more important contributions of the study was the use of a transparent and accessible “glass-box” approach to modelling that was able to explain not only at the global level, but also at the local level. Results from the experiments showed the framework had acceptable predictive validity which also preserved high interpretability, thereby fostering trust and better-informed decision-making for farmers.

III. LIMITATIONS OF PREVIOUS WORKS

While there have been tremendous steps forward in modelling pests with machine and deep learning, advancements in sensing technologies and explainable AI, there are still some issues. The current approach to pest monitoring is largely an

activity of the field, carried out in a total "reactive" fashion, and hence both labor intensive and time consuming, and may only be undertaken after noticeable crop losses have occurred.

It has been studied recently to use climatic conditions together with sensors and historical pest records to perform automatic predictions. Most of the systems that exist, however, only use limited data sources, just a few features and lack a true representation of the complex relationship between environmental conditions and pest dynamics. However certain treatments do not provide good representation of historical pest behaviour, seasonal patterns, lagged climatic effects, and recent infestation trends. Existing models tend to be limited to a single prediction or a short-term future prediction, which do not have the potential of being used for multi-horizon prediction required for prior pest control.

The performance prediction of ensemble and deep-learning models is found to be promising, however, the accuracy of these models varies according to the data set, climate elements and geographical regions. Data balance, few observations and fluctuations in climate conditions may also influence models and their reliability and generalisability. Moreover, the explainable AI techniques (e.g., SHAP) may be able to determine important features but reporting just the feature importance does not necessarily convey sufficient understanding of the role of climatic/temporal factors on pest development to predict future infestation.

The other limitation is the absence of practical steps to support decision making for farmers. There are numerous systems in existence that produce pest predictions in terms of numbers that farmers may not be able to easily interpret. A pest count doesn't really signal whether it is a low, medium or high risk situation. Thus, it is required to transform the prediction results to understandable and valuable risk categories that could be used for timely decision making.

Effectiveness of these restrictions points towards a need for an interpretable and climate-aware early-warning system that incorporates: climatic and historical pest information, temporal dependencies, multi-horizon, and explainable AI with risk-based decision support. The proposed work introduces a dual regression-classification framework is developed, with regression predicting the coming

intensity of pests and the proposed Pest Risk Index (PRI) turning infestation situations into low, medium, and high risk. This combination is designed to help technical users with comprehensive forecasting and generate an easier-to-understand and more manageable risk picture for farmers.

IV. METHODOLOGY

The suggested system is an Explainable Machine Learning methodology for Multi-horizon pest forecasting and pest risk classification. The system combines Climatic Variables, Temporal Pattern and Pests History to forecast the pest intensity at several future steps (T+1, T+2, T+3) and define the risk level of pests infestation. The framework also integrates techniques from explainable AI (XAI) to explain model predictions.

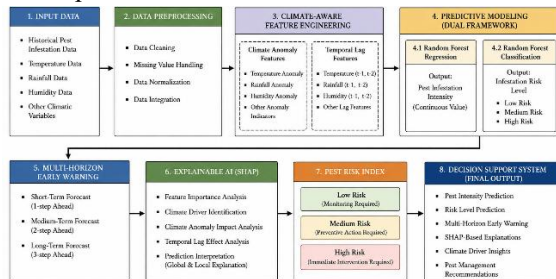


Fig1: System Architecture Diagram

A. Dataset Description

This study is based on the Rice Pest Dataset which is a field-collected agricultural time-series dataset developed for monitoring and forecasting rice pest infestation. The records are weekly observations of pests and their meteorological conditions.

- Temporal features: Observation Year, Standard Week
- Climatic features: Max Temperature (MaxT), Min Temperature (MinT), Relative Humidity (RH1, RH2), Rainfall (RF), Wind Speed (WS), Sunshine Hours (SSH), Evapotranspiration (EVP)
- Categorical features: Location, Collection Type, Pest Name
- Target variable: Pest Value (infestation intensity)

B. Data Preprocessing

Only the records for the pest Yellowstemborer are retained from the dataset. This helps to ensure

uniform pest behaviour modelling. The data is then sorted by Observation Year, Standard Week.

The intensity distribution of a pest was highly variable and skewed, so a logarithmic transformation is made in this method:

$$\bullet \text{Pest_Log} = \log(1 + \text{Pest Value})$$

This transformation helps to dampen variation and minimise the influence of extreme values of outbreak on the variance.

C. Feature Engineering

Feature engineering involves creating new variables that are informative from existing data in order to improve the predictive power.

1. Climate-Derived Features:

- $\text{MeanTemp} = (\text{MaxT} + \text{MinT}) / 2$ (Represents average thermal condition affecting pest development.)
- $\text{Temp_Range} = \text{MaxT} - \text{MinT}$ (Captures daily temperature variability influencing pest stress adaptation.)
- $\text{MeanRH} = (\text{RH1}(\%) + \text{RH2}(\%)) / 2$ (Average atmospheric moisture level)
- $\text{RH_Diff} = \text{RH1}(\%) - \text{RH2}(\%)$ (Represents diurnal humidity variation.)

2. Seasonal Encoding:

To determine cyclical pest behaviour throughout the agricultural seasons, sinusoidal transformation is done:

$$\bullet \text{week_sin} = \sin(2\pi \times \text{Standard Week} / 52)$$

$$\bullet \text{week_cos} = \cos(2\pi \times \text{Standard Week} / 52)$$

This encoding preserves periodic continuity between week 52 and week 1.

3. Location-Based Encoding:

The data type of the categorical variable Location is changed to another, numerical feature (Location_encoded). This is done using their respective historical observations to calculate their respective mean Pest_Log. This encoding provides spatial information of the extent of the occurrence of pests in different areas.

4. Collection Type Encoding:

One-hot encoding is used in the derive of a variable "Collection Type_Number", which accounts for the

different methodologies used in collecting the pest's data.

5. Lag Features (Temporal Dependency Modeling):

Lags are built to represent the history of the system when creating a model of the effects of the historical event on the system.

- Meteorological lag features: RF(mm)_lag_1, RF(mm)_lag_2, RF(mm)_lag_3, RF(mm)_lag_4, RF(mm)_lag_5, RF(mm)_lag_8, MeanTemp_lag_1 to MeanTemp_lag_8, MeanRH_lag_1 to MeanRH_lag_8
- Pest history lag features: Pest_Log_lag_1 to Pest_Log_lag_8

These features represent ecosystem feedbacks of pest populations to environmental changes that take time to occur.

6. Rolling Window Features:

These are made to represent the trends of a short period, and the momentum of outbreaks:

- pest_roll_4 → 4-week moving average of Pest_Log
 - pest_roll_8 → 8-week moving average of Pest_Log
- All of these features help to control the fluctuations and determine if build-up of infestations is occurring over time.

D. Target Variable Construction (Multi-Horizon Forecasting)

To enable early warning prediction, three forecasting horizons are defined:

- t1 = shift(-1) → Pest intensity in next week
 - t2 = shift(-2) → Pest intensity in two weeks ahead
 - t3 = shift(-3) → Pest intensity in three weeks ahead
- All targets based off Pest_Log, so regression based forecasting is possible.

E. Risk Classification Formulation (Pest Risk Index)

A multi-class risk labeling approach is used to categorize pest infestation severity. Risk levels are determined by setting thresholds on the quantiles of Pest Value:

- Low Risk (0) → $\leq$ 33rd percentile
- Medium Risk (1) → 33rd–66th percentile
- High Risk (2) → $\geq$ 66th percentile

Then, this turns the regression problem into a supervised classification problem to offer a support for decision making.

F. Final Feature Set Construction

- Raw climate variables: MaxT, MinT, RH1(%), RH2(%), RF(mm), WS(kmph), SSH(hrs), EVP(mm)
- Engineered climate features: MeanTemp, Temp_Range
- MeanRH, RH_Diff.
- Temporal features: week_sin, week_cos
- Spatial feature: Location_encoded
- Categorical feature: Collection Type_Number
- Temporal dependency features: Pest_Log_lag_, RF_lag_, pest_roll_4, pest_roll_8

G. Data Splitting Strategy

Training data and testing data are split randomly, 80% training, 20% testing and this is done to ensure an unbiased evaluation of the model under all intensities of pest classes.

H. Machine Learning Model

An ensemble learning approach known as the random forest is employed. Builds a number of decision trees and uses the averaged predictions as its result.

I. Risk Classification Models

Random Forest Classifier employ the same engineered features, and are evaluated for multi-class pest risk prediction.

J. Evaluation Metrics

- Regression Evaluation: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Coefficient of Determination (R^2). These are evaluated for T+1, T+2, T+3 forecasting horizons.
- Classification Evaluation: Precision, Recall, F1-score, Confusion Matrix

K. Explainable AI (SHAP Analysis)

The trained models are used to explain predictions with the aid of SHAP (SHapley Additive exPlanations). It charts the contribution of each feature for example:

- Pest_Log_lag_1 (historical pest influence)
- RF(mm) (rainfall impact)
- MeanTemp (temperature effect)
- Location_encoded (spatial influence)
- pest_roll_8 (trend effect)

SHAP provides global feature importance ranking, local prediction explanations, model transparency for agricultural decision-making.

V. RESULTS AND DISCUSSION

The performance of the proposed Explainable Multi-Horizon Early Warning System for pest infestation prediction was evaluated using both regression and classification approaches. The experimental results demonstrate the effectiveness of climate-aware feature engineering, temporal lag modeling, and explainable AI in improving prediction performance.

A. Regression Results (Pest Intensity Prediction)

The intensity of pest infestation was predicted using random forest regression model with climatic and historical data of pests. The models were compared against each other based on the performance measures MAE, RMSE, R^2 score.

The obtained results are:

Table1:Regression Performance Results

MAE	0.8301044207378162
RMSE	1.2020145954319865
R^2 Score	0.7800872082743631

These error values are low which means that the model can closely approximate the levels of pest. The R^2 value is high indicating that there is a large model trait explaining the pest intensity. This supports the point of climate anomaly and temporal lag and attributes them as a major reason to increase predictive accuracy.

B. Classification Results (Risk Level Prediction)

Pest infestation was classified as a Low risk, medium risk and high risk. The Accuracy, Precision, Recall and F1 score were evaluated for the model.

The results obtained are:

Table2:Classification Performance Results

	Precision	Recall	F1-Score	Support
Low Risk	0.95	0.87	0.91	280
Medium Risk	0.73	0.83	0.85	274

High Risk	0.87	0.83	0.85	311
Accuracy			0.84	865
Macro Avg	0.85	0.84	0.84	865
Weighted Avg	0.85	0.84	0.84	865

The results show the reliability of the model in discriminating the levels of the pest risk. The balanced F1-score implies that the model holds onto all classes without any problems, even if the dataset is imbalanced.

C. Confusion Matrix Analysis

The performance of the Random Forest classification model was evaluated by using a confusion matrix to determine the classification performance to predict the risk level of pest infestations. As shown in the Fig2, most of the samples were correctly classified in their respective tags, with 244, 227, and 257 samples being classified correctly for the classes 'Low-risk', 'Medium-risk', and 'High-risk' respectively.

There were a few similarities recorded, primarily between nearby risk rates (such as Low and Medium, or Medium and High). This is typical since pest levels will not fluctuate dramatically from one host to another. In summary, the confusion matrix showed that the proposed model can distinguish the different categories of pests well and has a high classification performance for pest outbreaks prediction.

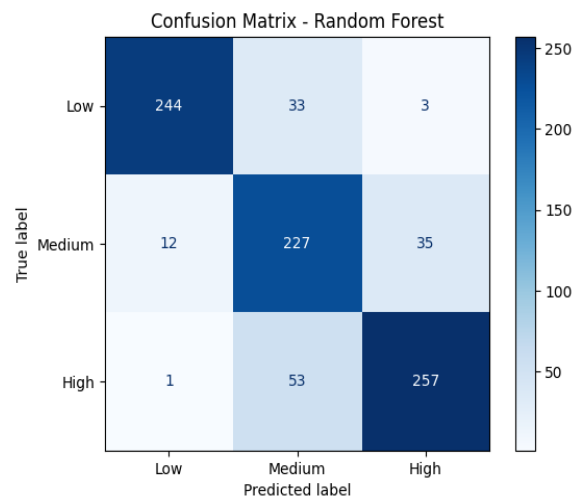


Fig2:Confusion Matrix

D. Multi-Horizon Forecasting Results

The results were a multi horizon forecasting module which made predictions for various time horizons (T+1, T+2 and T+3). Based on the results, the model's performance is relatively consistent in all the horizons of forecasting with a small increase in the error, at longer forecasting horizons.

Table3:Multi-Horizon Regression Performance Results

T+1	
MAE	0.8301044207378162
RMSE	1.2020145954319865
R2	0.7800872082743631
T+2	
MAE	0.9607008615221444
RMSE	1.3280656781776738
R2	0.7345665805291433
T+3	
MAE	1.0397372080380434
RMSE	1.431132876858923
R2	0.6987055518634664

E. SHAP Explainability Results

The features most influencing the prediction of pest infestation were identified using SHAP approach. Fig3 and Fig4 illustrates the contribution of each of the features according to the average contribution provided to the output of the model.

The results suggest that the most important features are Pest_Log_lag_1 and pest_roll_4, reflecting that pest occurrence data in recent years has a large impact on future pest occurrence. Other lag-based features including Pest_Log_lag_2, Pest_Log_lag_3 and Pest_Log_lag_8 also played an important role in the predictions.

Further, the model was sensitive to climatic factors like Mean Temperature, Relative Humidity and Temperature Range, which meant that climate state had a significant impact on model performance reflecting the impact of climatic factors on pest development. The seasonal characteristics of Standard Week, week_sin and week_cos were also added to help the model detect periodic patterns in pest infestation.

Overall, the SHAP analysis was helpful in elucidating the main factors contributing to pest infestations and to improve the interpretability of the

proposed model. The results indicate that having some historical pest data and climatic data is important in predicting pest infestation correctly.

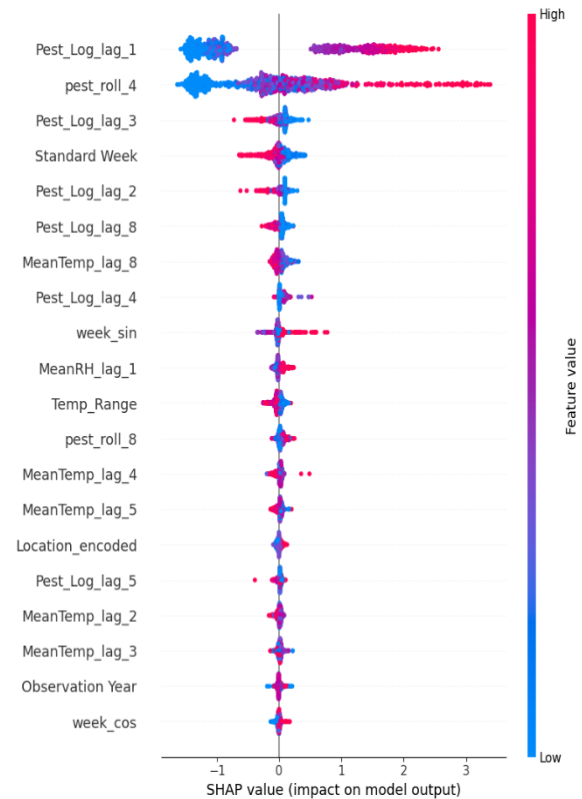


Fig3: SHAP feature importance summary plot

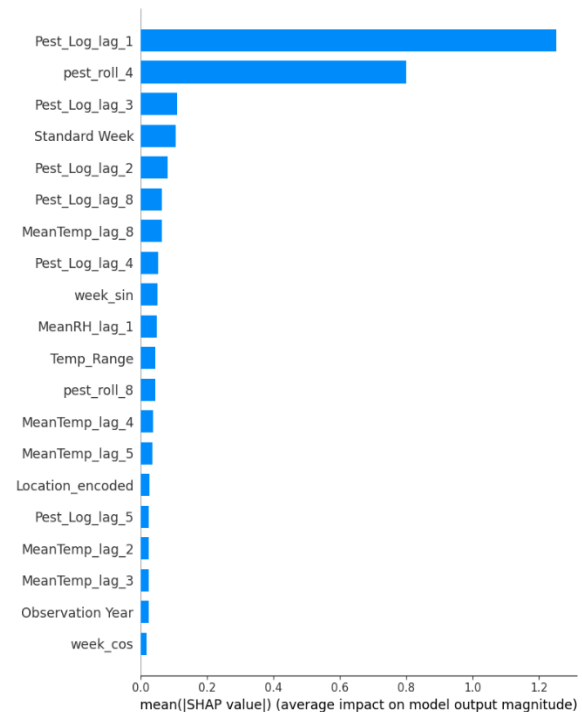


Fig4: SHAP feature importance Bar plot

F. Overall Discussion

Combining climate intelligent feature engineering, machine learning models, multi-horizon forecasting, and explainable AI boosts the prediction accuracy and explains the results. The proposed framework could offer more realistic and decision support solutions for pest management as opposed to traditional methods that seek to base estimates entirely on raw climatic data and to use only one-step prediction.

VI. CONCLUSION

The purpose of this study was to develop an Explainable Multi-Horizon Early Warning Framework for predicting pest infestation based on climatic data and the past history of pest records. The framework incorporates climate-driven feature engineering, machine learning-driven prediction, transparency of AI, and risk assessment to enhance forecasting of pest outbreaks. To model environmental influences on pest population dynamics climate anomaly features and temporal lag features were used, and Random Forest models were used predict the intensity of occurrence and for risk classification of the pests.

The SHAP analysis improved the interpretability of the model that was used by identifying the most influential factors to explain the presence of pest outbreaks, while the multi-horizon forecasting capacity gave ahead warnings of pest outbreaks for proactive pest management strategies. Moreover, the outputs of the Pest Risk Index could be converted to actionable risk levels, enabling informed pest risk decision making process within the agricultural sector.

REFERENCES

- [1]. Daisy Wadhwa, Kamal Malik. A generalizable and interpretable model for early warning of pest-induced crop diseases using environmental data. *Computers and Electronics in Agriculture*. 2024;227:1. <https://doi.org/10.1016/j.compag.2024.109472>.
- [2]. Tiago Domingues, Tomás Brandão, João C. Ferreira. Machine Learning for Detection and Prediction of Crop Diseases and Pests: A Comprehensive Survey. *Agriculture*.2022;12(9):1350.<https://doi.org/10.3390/agriculture12091350>
- [3]. Lee, S., Yun, C.M. A deep learning model for predicting risks of crop pests and diseases from sequential environmental data. *Plant Methods*. 2023;19:145.<https://doi.org/10.1186/s13007-023-01122-x>
- [4]. Jiale Wang, Dong Zhang. Intelligent pest forecasting with meteorological data: An explainable deep learning approach. *Expert Systems with Applications*.2024;252:124137.<https://doi.org/10.1016/j.eswa.2024.124137>
- [5]. Johannes Kopton, Sytze de Bruin, Dario Schulz, Eike Luedeling. Combining spatio-temporal pest risk prediction and decision theory to improve pest management in smallholder agriculture. *Computers and Electronics in Agriculture*.2025;236:110426. <https://doi.org/10.1016/j.compag.2025.110426>
- [6]. Dušan Marković, Dejan Vujičić, Snežana Tanasković, Borislav Đorđević, Siniša Randić, Zoran Stamenković. Prediction of Pest Insect Appearance Using Sensors and Machine Learning. *Sensors*. 2021; 21(14): 4846. <https://doi.org/10.3390/s21144846>
- [7]. Amit Bijlwan, Rajeev Ranjan, Manendra Singh, Rahul Purohit, Arun Jyoti Nath, Sumit Chakravarty. Predicting crop disease severity using real time weather variability through machine learning algorithms. *Scientific Reports* .2025;15:34767. <https://doi.org/10.1038/s41598-025-18613-7>
- [8]. Sita Rani, Amit Kumar Mishra, Aman Kataria, Saurav Mallik, Hong Qin. Machine learning-based optimal crop selection system in smart agriculture. *Scientific Reports*.2023;13:15997.<https://doi.org/10.1038/s41598-023-42356-y>
- [9]. Ranjit Kumar Paul, Sengottaiyan Vennila, Md Yeasin, Satish Kumar Yadav, Shabistana Nisar, Amrit Kumar Paul, Ajit Gupta, Seetalam Malathi, Mudigulam Karanam Jyosthna, Zadda Kavitha, Srinivasa Rao Mathukumalli, Mathyam Prabhakar. Wavelet Decomposition and Machine Learning Technique for Predicting Occurrence of

- Spiders in Pigeon Pea. *Agronomy*.2022;12(6):1429.
<https://doi.org/10.3390/agronomy12061429>
- [10]. Eric A. Ibrahim, Daisy Salifu, Samuel Mwalili, Thomas Dubois, Richard Collins, Henri E.Z. Tonnang. An expert system for insect pest population dynamics prediction. *Computers and Electronics in Agriculture*.2022;198:107124.
<https://doi.org/10.1016/j.compag.2022.107124>
- [11]. Yicun Ouyang, Hujun Yin. Multi-Step Time Series Forecasting with an Ensemble of Varied Length Mixture Models. *International Journal of Neural Systems*.2018;28(04):1750053.<https://doi.org/10.1142/S0129065717500538>
- [12]. Ankit Kumar Singh, Md Yeasin, Ranjit Kumar Paul, A. K. Paul, Anita Sarkar. Dynamic ensemble-based machine learning models for predicting pest populations. *Frontiers in Applied Mathematics and Statistics Journal*. 2024; 10: 1435517.
<https://doi.org/10.3389/fams.2024.1435517>
- [13]. F. Rodríguez-Díaz, A.M. Chacón-Maldonado, A.R. Troncoso-García, G. Asencio-Cortés. Explainable olive grove and grapevine pest forecasting through machine learning-based classification and regression. *Results in Engineering*. 2024; 24 : 103058.
<https://doi.org/10.1016/j.rineng.2024.103058>
- [14]. Wasswa Shafik, Ali Tufail, Liyanage Chandratilak De Silva, Rosyzie Anna Awg Haji Mohd Apong, Ki-Hyung Kim. Deep learning technique for plant disease classification and pest detection and model explainability elevating agricultural sustainability. *BMC Plant Biol*.2025;25:1491.
<https://doi.org/10.1186/s12870-025-07377-x>
- [15]. O. Nanushi, V. Sitokoustantinou, I. Tsoumas and C. Kontoes. Pest Presence Prediction Using Interpretable Machine Learning. 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), Nafplio, Greece, 2022, pp. 1-5, doi: 10.1109/IVMSP54334.2022.9816284.
<https://doi.org/10.1109/IVMSP54334.2022.9816284>
- [16]. S. Ajith, S. Vijayakumar, N. Elakkiya. Yield prediction, pest and disease diagnosis, soil fertility mapping, precision irrigation scheduling, and food quality assessment using machine learning and deep learning algorithms. *Discover Food*.2025;5:67.
<https://doi.org/10.1007/s44187-025-00338-1>
- [17]. Ritu Raj Kumar, Satendra Kumar, Arun Mittal, Ajesh Chauhan, Shivam Rajput, Rishabha Malviya. Application of Machine Learning and Deep Learning in Precision Agriculture for Crop Disease Detection and Pest Management. *Recent Advances in Computer Science and Communications*. 2025;19(8):151675.
<http://dx.doi.org/10.2174/0126662558393789251013123155>
- [18]. Imran Zualkernan, Daa Addeen Abuhani, Maya Haj Hussain, Jowaria Khan, Mohamed ElMohandes. Machine Learning for Precision Agriculture Using Imagery from Unmanned Aerial Vehicles (UAVs): A Survey. *Drones*.2023;7(6):382.<https://doi.org/10.3390/drones7060382>
- [19]. Muhammad Hassam Ejaz, Muhammad Bilal, Usman Habib, Muhammad Attique, Tae-Sun Chung. Crop Pest Classification Using Deep Learning Techniques: Areview. *arXiv:2507.01494v3*.2025.<https://doi.org/10.48550/arXiv.2507.01494>
- [20]. Balaban Ismail , Acun Fatih , Arpali Onur ,Murat Furkan ,Babaroğlu Numan, Akci Emre, Çulcu Mehmet, Özkan Mümtaz, Temizer Selim. (2019). Development of a Forecasting and Warning System on the Ecological Life-Cycle of Sunn Pest. *arXiv.1905.01640*.<https://doi.org/10.48550/arXiv.1905.01640>
- [21]. A.M. Chacón-Maldonado, G. Asencio-Cortés, A. Troncoso. A multimodal hybrid deep learning approach for pest forecasting using time series and satellite images. *Information Fusion*. 2025;124:103350.<https://doi.org/10.1016/j.inffus.2025.103350>