

PigeonXNET – Explainable Pigeon Pea Pest Forecasting

Vasantha Karthika¹, Meena Kowshalya²

^{1,2} *Department of Computer Science and Engineering, Government College of Technology, Coimbatore, India.*

Abstract - Pigeon pea (*Cajanus cajan*) is an important pulse crop grown in the tropical and subtropical regions of the world, especially in India. The productivity of pigeon pea is severely impacted by insect pests such as pod borers, aphids, whiteflies, and leaf webbers, which are highly dependent upon climatic conditions. Traditional pest monitoring and forecasting methods rely on manual observation and simple statistical models which often fail to produce accurate predictions about pest outbreaks in pigeon pea crops because they cannot handle nonlinear relationships. In the present study, we introduce PigeonXNet which is a deep learning system based on climate that predicts outbreak of pests in pigeon pea plants with the use of historical climate and pest data. In our proposed system, we have used various statistical models as well as deep learning models like Auto-Regressive Integrated Moving Average (ARIMA), CNN-LSTM, and Bi-LSTM which can handle both linear and non-linear relationships. While the ARIMA algorithm works as a base model for time series prediction, CNN-LSTM and Bi-LSTM model capture short-term as well as long-term dependencies and enhances the prediction accuracy of a Bi-LSTM model through the process of learning temporal information in bi-directional form. To make the prediction model transparent and interpretable, we utilize Explainable Artificial Intelligence approaches like SHapley Additive exPlanations (SHAP) in our approach, through which we can understand the contribution of various climatic parameters like temperature, humidity, and precipitation in pest outbreaks. As a result, experimental outcomes indicate the effectiveness of our proposed approach in forecasting accuracy. Our proposed system helps farmers make decisions timely regarding pest management, prevent unnecessary use of pesticides, and improve production of crops.

Keywords - Pigeon Pea, Pest Forecasting, Deep Learning, CNN-LSTM, Bi-LSTM, ARIMA, Explainable AI, SHAP.

1. INTRODUCTION

Pigeon pea (*Cajanus cajan*) is one of the major pulses that is widely grown in tropical and subtropical climates. The crop acts as a great source of protein and enriches the soil nitrogen by fixing the atmospheric nitrogen[1][2]. India is one of the highest producers of pigeon pea with major cultivation areas in Maharashtra, Karnataka, Madhya Pradesh, Gujarat, Telangana, Andhra Pradesh, Uttar Pradesh, Bihar, and some portions of Tamil Nadu[3][4]. Though it plays a significant role in agriculture, it is highly sensitive to insect pests like pod borer (*Helicoverpa armigera*), aphids, whitefly, pod fly, leaf webber, and sucking bugs. All of them lead to substantial damage to pigeon pea and hence affect the yield significantly. Environmental factors including temperature, humidity, rainfall, and season influence the outbreak of pests to a great extent. Any drastic changes in weather conditions can lead to faster propagation rates of pests, thus making predictions very much essential. But conventional monitoring of pests is done using field surveying along with the application of simple statistical techniques that may not be accurate enough to understand nonlinear relationships between climatic variables and the behavior of pests. These methods lack the capability to model long-term temporal dependencies and seasonal patterns, resulting in inaccurate forecasts.

Recent advancements in deep learning techniques and time series analysis have led to the development of intelligent systems that can learn complex patterns from large datasets[11][12]. Techniques like Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) have shown better performance in recognizing spatial and temporal features[13][14]. Especially, hybrid models like CNN-LSTM and Bidirectional LSTM (Bi-LSTM) has shown promising performance in modeling nonlinear relationships [15][16]. However, the lack of interpretability is seen in most of the models which is important for agricultural applications where the impact of environmental factors is important.

In order to overcome these challenges, this paper proposes PigeonXNet, a climate-based deep learning framework for effective pest forecasting for pigeon pea. The framework utilizes ARIMA, CNN-LSTM, and Bi-LSTM models to efficiently deal with linear as well as non-linear temporal relationships for effective pest occurrence predictions. Explainable Artificial Intelligence methods such as SHapley Additive exPlanations (SHAP) are incorporated to provide insights into the contribution of climatic variables influencing pest outbreaks[17][18].

The proposed framework is validated based on experimentations conducted on multi-year weather data and pest occurrence records on a monthly and seasonal scale. Experimental results prove that the Bi-LSTM model exhibits high accuracy in terms of pest forecasting compared to traditional statistical and deep learning methods. Thus, the proposed framework can be used as an efficient tool for effective pest warning and precision agriculture, thereby reducing pest occurrences and increasing crop productivity.

The main contribution of this project are

1. A novel system integrating statistical (ARIMA) and deep learning models (CNN-LSTM and Bi-LSTM) to accurately predict seasonal pest outbreaks in pigeon pea crops.
2. Effective utilization of multi-season weather data (temperature, humidity, rainfall) to capture both linear trends and nonlinear interactions influencing pest occurrence.
3. Implementation of Bi-LSTM to learn bidirectional temporal dependencies, significantly improving the accuracy of seasonal pest prediction compared to traditional models.
4. Incorporation of SHAP to provide transparency by identifying and quantifying the contribution of each environmental factor in pest outbreak predictions.

2. RELATED WORKS

2.1 ATTENTION-BASED LSTM INTERACTION CNN (ALIC) FOR PEST FORECASTING

In the study, the authors propose a novel attention-based LSTM interaction CNN (ALIC) architecture that can help make accurate pest forecasts in agricultural systems. In this study, the authors propose a new attention-based LSTM interaction CNN (ALIC)

architecture that can help make accurate pest forecasts in agricultural systems[19]. This model comprises various deep learning elements that help capture temporal and feature dependencies efficiently. The LSTM element helps learn temporal and seasonal trends from the meteorological data and past information on pests so that the model can be aware of time-based variations in pest outbreaks. Moreover, the interaction CNN (I-CNN) element helps identify relationships between different environmental factors like temperature, humidity, and rainfall, which greatly affect pest behavior. The ALIC model is tested based on performance indicators such as RMSE, MAE, and MAPE, which prove the model's superiority over traditional models like ARIMAX, SVR, ANN, and LSTM.

2.2 TIME SERIES ANALYSIS AND FORECASTING OF PIGEONPEA USING ARIMA

For the purpose of prediction about future trends in the growth of pigeon pea cultivation, the ARIMA model will be used[20]. The data set contains previous records of the pigeon pea area, production, and yield. ARIMA is a statistical tool that helps forecast future trends based on historical data. ARIMA is an effective approach to forecasting univariate time series, as it considers both linear trends and seasonality within historical data. The model predicts long-term trends related to agricultural products' area, production, and yield. This approach can serve as a foundation for planning and decision-making in agriculture. Additionally, the method's simplicity allows its use for comparison purposes during prediction tasks. However, the approach has its drawbacks. For example, ARIMA takes into account solely the past values of the time series, ignoring such variables as climate or pest invasion.

2.3 INFLUENCE OF WEATHER PARAMETERS ON SEASONAL INCIDENCE OF POD BORING WEEVIL

The present study is concerned with assessing the effects of meteorological factors on the seasonal occurrence of the pod boring weevil (*Apion clavipes*) in pigeon pea crops by observation in the field. The investigation comprises gathering of actual data from the farm, involving the number of pests and relevant meteorological factors such as temperature, precipitation, and relative humidity. Through correlation analysis, it seeks to explore the link between climatic factors and the occurrence of the pest. Thus, the study succeeds in identifying significant seasonal

outbreaks of pests. It effectively shows the connection between meteorological factors and the occurrence of pests, hence useful in early warning and sound pest management plans. Its reliance on actual data from the field makes it very practical and credible. But it lacks the use of any simulation model for analyzing the nonlinear link between meteorological factors and pest dynamics.

2.4 PREDICTION OF COTTON PEST AND DISEASE OCCURRENCE USING BI-DIRECTIONAL LSTM

The current paper suggests using the Bi-directional LSTM-based method for estimating the probability of pest and diseases appearance in cotton cultivation by utilizing historical climatic parameters like temperature, humidity, rain, and light as input data. The problem was defined as a temporal forecasting problem and solved with a Bi-directional LSTM model which learns from the sequence of data in both directions at once. This allows the neural network to estimate the relationships between previous and future points of the sequence, resulting in better prediction accuracy. The performance of the proposed algorithm was measured by the area under the curve metric, reaching the value of 0.95. Thus, the effectiveness of this model proves the impact of climatic factors on pests' appearance, and the bi-directionality of the model makes the model able to better capture the temporal patterns in the data. Nevertheless, the method requires large amounts of historical data regarding climatic conditions and pest activity in particular, and the computational complexity is relatively high. Besides, the algorithm is only applicable to cotton crops and does not provide explainability of decisions made.

2.5 FORECASTING OF PRODUCTIVITY AND POD DAMAGE USING ARTIFICIAL NEURAL NETWORK

In this research, an ANN-based prediction model for pod damage by *Helicoverpa armigera* and total productivity of pigeon pea cultivation was developed. The model is built with a two-layer feed-forward ANN, which learns the data pattern from past agricultural records. The model is capable of forecasting important parameters like percentage of pod damage and pigeon pea production (kg/ha). The effectiveness of the prediction process of the ANN model was assessed with mean square error (MSE) and correlation coefficient (R). The results showed that the proposed ANN model performed better in predicting the output parameters

with a high value of $R = 0.98$ for pod damage rate. One of the major advantages of the ANN model is its ability to predict the non-linear relationship between the inputs and outputs. Moreover, it has the capability of early forecasting. On the other hand, the ANN model suffers from problems of complexity and poor interpretability.

2.6 IMPROVING TIME SERIES FORECASTING USING LSTM AND ATTENTION MODELS

This research work introduces a hybrid deep learning method based on LSTM and multi-head attention for better time-series prediction. In this case, the deep learning model takes into account the complexity of the relationship between variables in time series. The use of LSTM allows us to determine the long-term dependencies between input and output, while multi-head attention focuses on relevant time steps and features. The proposed technique is benchmarked against conventional statistical methods and other hybrid techniques. The experimental findings indicate that the hybrid approach has been successful in handling complicated time-series data sets, yielding higher predictive accuracy. This model performs much better than others in Average Rank (AR). However, the proposed model suffers from computational efficiency because of its hybrid nature, and it necessitates vast amounts of data for proper training. Another problem is that interpreting the results can be very challenging.

2.7 FORECASTING MODELS FOR PREDICTING POD DAMAGE OF PIGEONPEA BY POD BORER

This work attempts to predict the damage inflicted on pigeon pea pods by the insect *Helicoverpa armigera* based on time-series analysis over many years. This study incorporates 27 years of historical data and uses climatic parameters like maximum and minimum temperatures, rainfall, and relative humidity for forecasting the damage on pigeon pea pods by the insects. Many models, such as Multiple Linear Regression (MLR), Auto Regressive Integrated Moving Average (ARIMA), and Artificial Neural Network (ANN) model, were formulated and analyzed based on their performances. Levenberg Marquardt (LM) is used to train the Artificial Neural Network (ANN) to increase efficiency. The performance of each model was tested based on evaluation metrics such as RMSE, MSE, and R^2 values. The work offers a comparative analysis between statistical and machine learning models, demonstrating that ANN achieved the highest performance with the minimum value of RMSE. These

findings indicate that ANN is efficient in capturing non-linear relationships and performing predictions with high accuracy. On the downside, ANN lacks interpretability because of being a black box model. It is also important to note that the ANN algorithm doesn't include state-of-the-art deep learning architectures such as LSTM. Furthermore, forecasting accuracy depends highly on quality historical data.

3 PROPOSED SYSTEM

3.1 OBJECTIVES

1. Develop a climate-based model to forecast seasonal pigeon pea pests using past weather and pest data records.
2. Examine the impact of climatic conditions on the appearance, growth, and spread of pests.
3. For the application and comparison of ARIMA, CNN-LSTM, and Bi-LSTM models for linear and non-linear temporal pattern recognition in pest appearances.
4. Determine important environmental factors with explainable machine learning techniques such as LIME and SHAP.

3.2 SYSTEM MODEL

The proposed PigeonXNet system follows a structured methodology as illustrated in figure 3.1 to forecast pest outbreaks using climate and historical pest data. Initially, raw weather and pest datasets are collected and pre-processed to remove inconsistencies and normalize features. The processed data is then transformed into time-series sequences to capture seasonal patterns. Multiple models including ARIMA, CNN-LSTM and Bi-LSTM are trained to learn both linear and nonlinear relationships. The models are evaluated using performance metrics and the best-performing model is selected. Finally, Explainable AI techniques (SHAP) are applied to interpret the results and the system generates pest forecasts and early warning alerts for farmers.

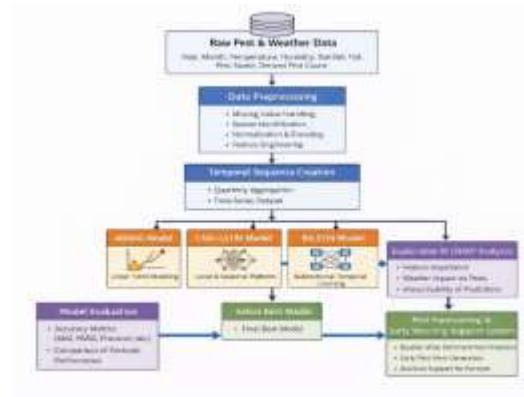


Fig-3.1: SYSTEM ARCHITECTURE

3.2.1 Raw Pest and Weather Data Collection

The methodology begins with the collection of raw datasets that include both climatic and pest-related information over multiple years. The weather dataset consists of important environmental parameters such as temperature, humidity, and rainfall, which directly influence pest growth and behavior. Along with this, historical pest data is collected, including pest names, pest occurrence, and infestation counts. Time-related attributes such as year and month are also included to capture seasonal variations. This combined dataset forms the base for building a reliable pest forecasting system.

3.2.2 Data Preprocessing

After the data collection step, the data is preprocessed to ensure the quality and consistency of the data. Missing values in the dataset are handled appropriately. Noise in the data is removed to ensure the quality of the data. The data is then normalized using the MinMax algorithm to normalize the data. This helps in the convergence of the model. The categorical data, such as the name of the pest, is encoded. Feature engineering techniques are applied to extract useful information from the data. For example, the seasonal pattern in the data from the month feature can be determined.

3.2.3 Temporal Sequence Creation

In order to properly identify time-dependent patterns within the data, the data is converted into time series format. The daily data is aggregated into monthly and quarterly data to account for seasonal changes in pest occurrence. This data is then converted into input-output pairs to be used for training time series models. This allows the models to learn how pest populations change based on climatic changes, which is important for proper forecasting.

3.2.4 Model Development

In this step, multiple models are developed and trained to forecast pest outbreaks. The ARIMA model has been used as a baseline model to tackle the problem using a statistical approach. However, since the pest population is affected by many nonlinear factors, deep learning models have also been used. The CNN-LSTM model combines convolutional layers to extract short-term temporal features and LSTM layers to learn long-term dependencies. Furthermore, the Bi-LSTM model is employed to improve performance by processing the sequence in both forward and backward directions allowing it to capture complete temporal relationships and enhance prediction accuracy.

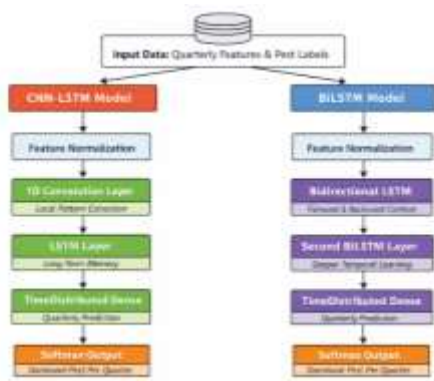


Fig-3.2: COMPARATIVE ARCHITECTURE OF CNN-LSTM AND BI-LSTM MODELS

Figure 3.2 represents the deep learning architecture implemented to perform pest forecasting, where the performance of two models such as CNN-LSTM and Bi-LSTM are applied to the same input data and compared for performance.

At the top, the system takes the input data which consists of quarterly features and pest labels. This includes information related to climate parameters such as temperature, humidity, rainfall, and pest occurrence. This input data is then passed to two different models such as CNN-LSTM and Bi-LSTM to compare their performance and forecasting capabilities.

In the CNN-LSTM model, the first step is feature normalization, where all the features are normalized to an appropriate range to ensure efficient learning. Next, the features are passed through a 1D convolution layer to extract local temporal patterns from the sequential data such as short-term fluctuations in weather conditions. This extracted features are then passed to the LSTM layer to detect long-term trends. The output is then fed into a Time Distributed Dense layer which

generates predictions for each time step (quarterly prediction). Finally, a Softmax output layer is used to classify and predict the dominant pest for each quarter.

On the other hand, the Bi-LSTM model also begin with feature normalization but instead of convolution, a Bidirectional LSTM layer is used. This layer processes the data in both directions to obtain complete contextual information from past and future data points. To obtain even better results, another Bi-LSTM layer is used for enhanced learning. Similar to the previous model, the output is passed through a Time Distributed Dense layer for quarterly predictions followed by a Softmax layer to determine the dominant pest.

This architecture demonstrates two different approaches to time-series forecasting: CNN-LSTM model focuses on local pattern extraction and sequential learning, while Bi-LSTM model focuses on complete understanding of time series data. Both models aim to predict pest outbreaks accurately, with Bi-LSTM generally providing better performance due to its bidirectional learning capability.

3.2.5 Model Evaluation

After the models are trained, they are evaluated using performance metrics. For the ARIMA model, error-based metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are used. For the deep learning models classification metrics such as accuracy, precision, recall, and F1-score are calculated. These metrics help in comparing the performance of different models and determining their effectiveness in predicting pest outbreaks.

3.2.6 Best Model Selection

After the models are trained, they are then evaluated using the necessary performance metrics. For the ARIMA model, the necessary metrics include the Mean Absolute Error and the Root Mean Squared Error. For the deep learning models, the necessary metrics include the accuracy, precision, recall, and F1 score. These metrics help in comparing the performance of the models and assessing their efficiency in predicting the pest infestations.

3.2.7 Explainable AI (SHAP Analysis)

In order to make the model more transparent and comprehensible, Explainable Artificial Intelligence techniques are employed. SHAP analysis is a technique that helps to understand the contribution of each feature, i.e., temperature, humidity, and rainfall, to the

prediction result. This will provide an understanding of how various climatic factors affect pest occurrence. This will make the model easier to understand and comprehend by farmers and agricultural experts.

3.2.8 Pest Forecasting and Early Warning System

The selected model is then employed to forecast the occurrence of pests. This system will be able to forecast the type of pest that is likely to occur during a given season. Based on this prediction, an early warning will be given to farmers. This will allow farmers to take appropriate actions to control the pests.

4 EXPERIMENTAL RESULTS

4.1 DATASET

The dataset that was used for this research encompasses climate and pest data from the years 2014 to 2024. The data set includes daily climatic variables like temperature, humidity, and rainfall data that have been summarized into monthly and quarterly data sets. Such climatic variables are combined with past pest data including type of pests and their occurrences. Preprocessing of the data followed by transformation into time series data sets helps in modeling the effect of the environment on pest outbreaks. With such data, effective training and testing of forecast models can be done leading to early detection of pests in pigeon pea plants.

4.2 Result and Discussion

The data set is divided into 70 percent for training, 15 percent for validation, and 15 percent for testing to make sure that model training and performance testing are done correctly. The experiments are performed using Python-based tools on a computer, making sure that the proposed method can be implemented in an efficient manner. The learning rate starts at 0.001, using adaptive optimization methods, and the early stopping criterion is used based on validation loss convergence.

In order to assess the proposed approach, i.e., PigeonXNet, it is compared against other baseline and deep learning techniques such as ARIMA, CNN-LSTM, and Bi-LSTM. ARIMA algorithm is employed to detect linear trend and seasonality in the time series. The CNN-LSTM algorithm is adopted to derive short-term temporal patterns in the time series using convolution layers and derive long-term dependency patterns using LSTM layer. The Bi-LSTM algorithm makes prediction based on short-term as well as long-term temporal dependencies.

Performance of the above discussed models is evaluated using some standard performance metrics like MAE, RMSE, accuracy, precision, recall, and F1-score. For instance, ARIMA model produced a mean absolute error (MAE) of 17.687 and root mean square error (RMSE) of 21.524. On the other hand, CNN-LSTM model produced an accuracy of 90.48%, and the bi-directional LSTM algorithm produced an accuracy of 97.73%. These results indicate that the Bi-LSTM model is superior to other models due to its superior ability to detect temporal patterns in climatic factors leading to pest occurrence.

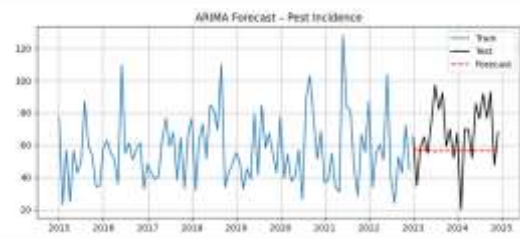


Fig 4.1: ARIMA-Based Time Series Forecasting of Pest Incidence

The ARIMA model was developed and implemented to forecast the incidence of pigeon pea pests by predicting based on historical data. The data was split into training and test sets. The ARIMA model was trained with historical data (2015-2022) and then tested with other data (2023-2025) as illustrated in figure 4.1

Based on the forecast plot developed, it was found that the training data (blue line) showed considerable fluctuations over time. This indicates that the pest incidence is irregular and changes over time. Similarly, the test data (black line) is also showing considerable variation. However, the ARIMA forecast (red dashed line) is almost constant. This shows that the ARIMA model is not predicting sudden changes but is predicting the average value.

```
from sklearn.metrics import mean_absolute_error, mean_squared_error

MAE = mean_absolute_error(test_data, forecast)
RMSE = np.sqrt(mean_squared_error(test_data, forecast))

print("ARIMA MAE: ", MAE)
print("ARIMA RMSE: ", RMSE)

ARIMA MAE: 17.687385516740402
ARIMA RMSE: 21.52436561327068
```

Fig -4.2: Evaluation Metrics of ARIMA Forecasting Model

The performance of the ARIMA model was evaluated based on the error metrics as illustrated in figure 4.2. From the graph, it was found that the Mean Absolute Error (MAE) was observed to be 17.68. Similarly, the Root Mean Square Error (RMSE) was found to be

21.52. This shows that the prediction is not very accurate. This indicates that the ARIMA model is not very accurate. This is due to the fact that the RMSE is high compared to the MAE. This shows that there are considerable prediction errors. From the graphical and numerical analysis, it is obvious that ARIMA fails to model the non-linear and complex relationships between the climatic parameters and pest behavior. It has a tendency to generate highly smoothed output, failing to recognize the abrupt changes in the pest population. This is because ARIMA is essentially a linear model and is not capable of effectively learning the complex relationships between different parameters.

Despite the drawbacks of the ARIMA model, it is still useful for comparative purposes. It helps in understanding the significance of temporal trends and seasonality in pest forecasting. It also helps in realizing the potential offered by more advanced models. In comparison with the deep learning-based approaches like CNN-LSTM and Bi-LSTM, it has shown poor accuracy and flexibility. This proves the potential offered by hybrid and deep learning-based approaches in pest forecasting.

From the above discussion, it is obvious that although ARIMA is easy to implement and has less computational complexity, it is not suitable for accurate pest forecasting in a changing agricultural environment. It has shown poor performance in comparison with the proposed deep learning-based approaches for pest forecasting in the proposed PigeonXNet framework.

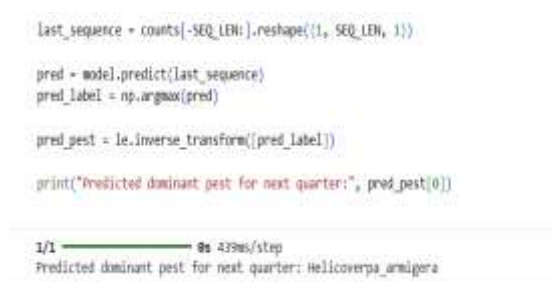


Fig- 4.3: CNN–LSTM Based Pest Prediction Model

The CNN–LSTM model is implemented in this project to accurately forecast dominant pest occurrences using environmental and historical pest data. As shown in Figure 4.3, the model combines Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) techniques to capture both local feature patterns and long-term temporal dependencies from time-series agricultural data. The CNN layer extracts important local patterns and short-term environmental changes related to pest outbreaks. These features are

then passed to the LSTM layer, which learns long-term dependencies and seasonal trends from past data. This helps the model understand how previous environmental conditions influence future pest occurrences. After feature learning, a dense layer processes the information, and the softmax output layer predicts the most dominant pest likely to occur in the upcoming quarter. Overall, the CNN–LSTM model effectively captures complex relationships between climate conditions and pest behavior, making it suitable for accurate agricultural pest forecasting.

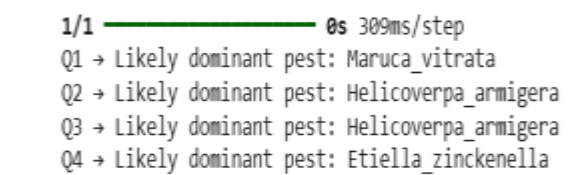


Fig-4.4: Bi-LSTM Model

The Bi-LSTM model effectively predicts dominant pests across different seasonal quarters by learning temporal patterns from time-series data. The model predicts *Maruca vitrata* in Q1, *Helicoverpa armigera* in Q2 and Q3, and *Etiella zinckenella* in Q4. These results show that the model successfully captures seasonal variations and long-term dependencies in pest occurrence patterns, as illustrated in Figure 4.4. By analyzing both past and future sequence information, the Bi-LSTM model provides accurate and context-aware predictions. The high accuracy of about 97.73% and fast prediction time demonstrate the model's reliability and suitability for near real-time pest forecasting. Overall, the model supports proactive pest management and helps improve agricultural decision-making.

TABLE 4.1 MONTHLY PEST PREDICTION ANALYSIS

MONTH	PEST
MONTH 1	<i>Clavigralla gibbosa</i>
MONTH 2	<i>Maruca vitrata</i>
MONTH 3	<i>Aphis craccivora</i>
MONTH 4	<i>Aphis craccivora</i>
MONTH 5	<i>Etiella zinckenella</i>
MONTH 6	<i>Melanagromyza obtusa</i>
MONTH 7	<i>Maruca vitrata</i>
MONTH 8	<i>Maruca vitrata</i>
MONTH 9	<i>Aphis craccivora</i>
MONTH 10	<i>Helicoverpa armigera</i>
MONTH 11	<i>Helicoverpa armigera</i>
MONTH 12	<i>Maruca vitrata</i>

Table 4.1 presents the predicted dominant pests per month using the forecasting model can be found in Table 4.1. The results demonstrate that there is a definite pattern of variation depending on the time of year. In the

early season month, *Clavigralla gibbosa* is expected, while *Maruca vitrata* and *Aphis craccivora* are frequent pests during the mid-season months. During mid-season months, pests such as *Etiella zinckenella* and *Melanagromyza obtusa* will also be present, suggesting that the environment is different. The later season months are dominated by *Helicoverpa armigera* in Months 10 and 11, and then *Maruca vitrata* in Month 12.

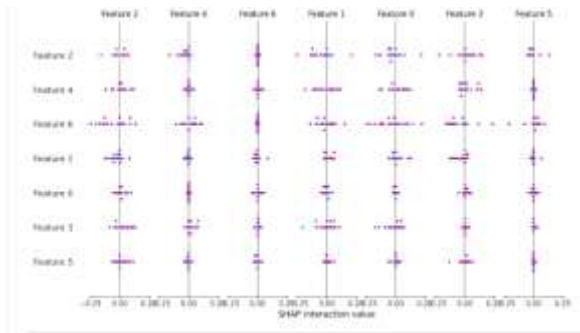


Fig-4.5: SHAP

Figure 4.5 illustrates the SHAP interaction plot developed for the proposed pest prediction using Bi-LSTM is depicted in figure 4.5. It describes how the contributions of different environmental features interact with each other in terms of pest prediction. The dots in the figure represent the SHAP interaction value of two features that contribute to the overall prediction made by the model.

The positive SHAP interaction value means that the contribution of the selected climatic variables increases pest prediction, while the negative SHAP interaction means the opposite contribution. Different colors in the figure represent the degree of the interaction between features, wherein the red-colored features have a high degree of interaction compared to the blue-colored ones.

5. COMPARATIVE ANALYSIS WITH EXISTING ARCHITECTURES

5.1.1 PRECISION

The Precision metric indicates the ratio of the samples classified as belonging to a particular class that are indeed correct. It shows the effectiveness of the model at minimizing false positive errors.

5.1.2 RECALL

The Recall metric indicates the number of samples of a particular class which were correctly classified by the model. It demonstrates the efficiency of the model in avoiding false negatives.

5.1.3 F1 SCORE

The F1-Score is the harmonic mean between the Precision and Recall metrics. It is particularly effective for skewed datasets.

TABLE 5.1 EVALUATION METRICS

Model	Precision	Recall	F1-score
CNN			
0	0.86	0.98	0.91
1	0.97	0.83	0.90
Accuracy			0.91
Macro avg	0.91	0.90	0.90
Weighted avg	0.91	0.91	0.91
Bi-LSTM			
0	0.97	0.96	0.97
1	0.91	0.94	0.93
Accuracy			0.97
Macro avg	0.94	0.95	0.95
Weighted avg	0.97	0.97	0.97

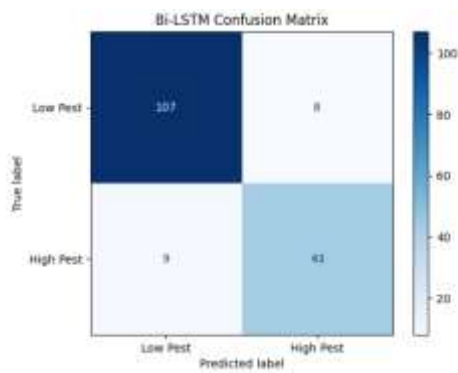
Table 5.1 highlights the comparative analysis of the CNN and Bi-LSTM models based on the evaluation metrics like Precision, Recall, F1-score, and Accuracy.

For class 0, CNN model showed high accuracy by achieving overall accuracy of 91%. The model predicted Precision, Recall, and F1-Score of 0.86, 0.98, and 0.91 respectively, which means that the model was quite effective in predicting the samples falling under class 0. On the other hand, for class 1, CNN scored precision of 0.97, recall of 0.83, and F1-score of 0.90, which meant that the model performed well in predictions but recall score was comparatively lower. Both Macro Average and Weighted Average scores were around 0.90-0.91.

As compared to CNN model, Bi-LSTM model proved to be superior in all metrics. Overall accuracy of 97% showed its better performance than CNN. In addition, Bi-LSTM model scored high precision, recall, and F1-score for both classes. For class 0, model scored a precision of 0.97, recall of 0.96, and F1-score of 0.97 while for class 1, precision was 0.91, recall was 0.94, and F1-score was 0.93. Macro and weighted averages were 0.95 and 0.97 respectively.

TABLE 5.2 PERFORMANCE COMPARISION

Model	Dataset	Accuracy
NN	Weather and Pest History dataset	87.6%
CNN-LSTM		91.4%
Bi-LSTM (proposed)		97.73%


Fig- 5.1: CONFUSION MATRIX

6 CONCLUSION

In this study, an intelligent approach based on deep learning was designed for the purpose of pigeon pea pest prediction through time-series modeling of environmental variables along with historical pest data. It is important to note that this study efficiently leveraged and compared ARIMA, CNN-LSTM, and Bi-LSTM models to incorporate both linear and nonlinear temporal effects on the pest infestation pattern. ARIMA model offered robust benchmark results for the trend detection of pests while CNN-LSTM model accounted for feature interaction and temporal dependency of data. In comparison to ARIMA and CNN-LSTM, the performance of Bi-LSTM was further enhanced by considering both the past and future temporal information leading to superior accuracy. The empirical results indicated that the performance of Bi-LSTM was higher than CNN-LSTM with MAE: 10.244 and RMSE: 10.460 while being much better than ARIMA model having MAE: 17.687 and RMSE: 21.524. Besides, the employment of Explainable AI techniques like SHAP

helped identify the most influencing environmental factors contributing to pests infestations.

Acknowledgement

Authors thank Department of Computer Science and Engineering, Government College of Technology, Coimbatore for the support.

REFERENCES

- [1] FAOSTAT, "Pigeonpea production statistics and cultivation trends," Food and Agriculture Organization of the United Nations (FAO), Rome, Italy, 2024.
- [2] R. Sharma, P. Verma, and S. Kulkarni, "Climate effects on pigeonpea pests and seasonal outbreak analysis," *Journal of Agricultural Science*, vol. 15, no. 3, pp. 112–120, 2023.
- [3] Agricultural Research Council, *Pulse Crop Productivity Report 2022*, New Delhi, India: ARC Publications, 2022.
- [4] K. Patel, M. Reddy, and A. Singh, "Insect pests in pigeonpea cultivation and their management practices," *International Journal of Pest Management*, vol. 67, no. 4, pp. 255–263, 2021.
- [5] P. Singh and R. Kumar, "Yield loss assessment due to pod borers in pigeonpea crops," *Crop Protection*, vol. 135, pp. 104–110, 2020.
- [6] V. Rao, S. Mishra, and N. Gupta, "Weather-based pest forecasting systems for sustainable agriculture," *Agricultural Informatics Journal*, vol. 10, no. 2, pp. 78–89, 2022.
- [7] Indian Council of Agricultural Research (ICAR), *Annual Report on Climate and Agriculture 2023*, New Delhi, India: ICAR, 2023.
- [8] A. Kumar, D. Prasad, and S. Roy, "Traditional pest monitoring methods in crop protection: A review," *Journal of Agricultural Extension*, vol. 29, no. 1, pp. 45–53, 2021.
- [9] Food and Agriculture Organization (FAO), *Pest Surveillance Handbook*, Rome, Italy: FAO Publications, 2020.
- [10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.
- [13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [14] X. Wang, H. Li, and Z. Chen, "CNN-LSTM hybrid forecasting model for agricultural prediction," *Computers and Electronics in Agriculture*, vol. 184, pp. 106–118, 2021.
- [15] H. Li, Y. Zhao, and M. Chen, "Bi-LSTM based agricultural prediction using climate data," *IEEE Access*, vol. 10, pp. 55231–55240, 2022.
- [16] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, Long Beach, CA, USA, 2017, pp. 4765–4774.
- [17] C. Molnar, *Interpretable Machine Learning*. Berlin, Germany: Lulu.com, 2020.
- [18] Y. Zhao, L. Wang, and P. Kumar, "Seasonal pest prediction analysis using machine learning techniques," *Agricultural Systems*, vol. 204, pp. 103–115, 2023.
- [19] T. Chen, X. Liu, and H. Zhang, "Bi-LSTM performance in agricultural forecasting applications," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 4, pp. 455–463, 2022.
- [20] S. Karthikeyan, R. Prakash, and M. Devi, "PigeonXNet: A hybrid forecasting framework for climate-based pest prediction in pigeonpea crops," *International Journal of Advanced Agricultural Research*, vol. 12, no. 2, pp. 88–99, 2025.
- [21] R. Mehta and S. Das, "Weather-pest interaction modeling in crops using statistical learning approaches," *Journal of Precision Agriculture*, vol. 18, no. 3, pp. 211–223, 2021.
- [22] J. Park, H. Kim, and S. Lee, "Advanced temporal learning using Bi-LSTM for agricultural forecasting," *IEEE Access*, vol. 11, pp. 78210–78221, 2023.
- [23] P. Sharma and V. Gupta, "Explainable AI in precision agriculture: Opportunities and challenges," in *Artificial Intelligence for Sustainable Agriculture*. Singapore: Springer, 2022, pp. 145–160.
- [24] L. Zhang, Y. Chen, and R. Zhao, "Attention-based LSTM interaction CNN for pest forecasting," *IEEE Access*, vol. 12, pp. 22541–22555, 2024.
- [25] M. Roy, A. Banerjee, and S. Paul, "Deep learning-based interpretable pest forecasting systems for smart agriculture," *Computers and Electronics in Agriculture*, vol. 218, pp. 108–120, 2024.