Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 04 Issue 01
2022

Research Article
https://doi.org/10.56979/401/2022/67

Forecasting Cotton Whitefly Population Using Deep Learning

Abdul Majid Soomro[1*], Awad Bin Naeem[1], Khuram Shahzad[1], Abdul Mustaan Madni[1], Muhammad Sajid[1], Armado Dacalap Del Mundo[2], and Muhammad Ashad Baloch[1]

[1]Department of Computer Science, National College of Business Administration & Economics, Multan, Pakistan.
[2]Department of Computer Science, University of Technology and Applied Sciences Salala, Oman

[*]Corresponding Author: Abdul Majid Soomro. Email: gi180004@siswa.uthm.edu.my

Received: July 12, 2022 **Accepted:** September 15, 2022 **Published:** December 29, 2022.

Abstract: Agricultural is the primary source on which a country's economy depends. Pakistan is the world's fourth-largest cotton producer, making it one of its top cash crops. They can't figure out what sort of cotton is best for their climate because of government regulations, crop risks, and a low literacy rate among the general public and farmers. Using a model that considers temperature, this system attempts to describe the different types of plants. More than anything else, the pest's population is concerned with its diversity, how the weather affects its population, and when it is high or low. The primary purpose of this study was to develop a framework that can handle the complex process of the population of cotton whitefly. Our main aim was to know the population of insects with their eggs and with their parents. And another purpose is to find out information about the variety with a low population of insects. To get early knowledge about the people against the temperature. So, cotton yield could be increased, and fewer chemicals should be used. Therefore, farmers' income could be improved. So, we will make the best model for predicting whiteflies on cotton ARIMAX. This model's accuracy was nearly on par with the statistical forecasting models. Because ARIMAX is a statistical model, it may also be used for forecasting.

Keywords: Forecasting Cotton, Deep Learning, Whitefly Population, Feature Extraction

1. Introduction

Agriculture's contribution to Pakistan's economy is expected to be substantial. Agriculture is the country's main economic activity and source of wealth. Gossypium hirsutum L., the botanical name for cotton, is the world's most significant crop for manufacturing textile fibers. Soft and breathable textiles can be made from wool by spinning it into yarn or thread. Today's apparel's most common natural fiber is made from this particular substance. Cotton accounts for more than 40% of the world's natural fiber production, making it the most important source of fiber. Cotton's long history and output are an inseparable link to the earliest stages of human civilization." White Gold," known in Pakistan, is a significant agricultural commodity. As a result, it significantly impacts Pakistan's agricultural industry, regardless of how developed or underdeveloped it is. Oil, lint, and hulls are only a few products made from them, all of which are generated in warmer locations. Regarding cotton output, Pakistan is ranked fourth in the world. Eighty percent of the nation's oil supply comes from cotton farming.

Ginning and turning operations employ cotton as their primary raw material. Various uses make cotton essential to the economy, from food and fuel to fibers and even foreign exchange. Cotton's impact on our society and culture stretches far beyond economics. Due to various reasons, including pests, climate change, and illnesses, the quality and quantity of cotton crops have decreased significantly. They can't figure out what sort of cotton is best for their climate because of government regulations, crop risks, and a low literacy rate among the general public and farmers. Cotton-eating insects are thought to number as many as 1,326 distinct species worldwide. However, the Indian subcontinent only has 162 bug species that feed on cotton. According to a study, Pakistani cotton is plagued by 93 pests and creepy crawlies. Experiments show that modern technologies and new methodologies should be used to implement novel approaches. Crop production, protection, picking, transportation, and storage all play a role in regulating Pakistan's cotton industry.

Pakistan is the world's fourth-largest cotton producer, making it one of its top cash crops. They can't figure out what sort of cotton is best for their climate because of government regulations, crop risks, and a low literacy rate among the general public and farmers. Using a model that considers temperature, this system attempts to describe the different types of plants. More than anything else, the pest's population is concerned with its diversity, how the weather affects its population, and when it is high or low. So major problem is to develop a framework that can handle the complex process of the people of cotton whitefly. Our main aim was to know the people of insects with their eggs and with their parents. And another purpose is to find out information about the variety with a low population of insects. To get early knowledge about the people against the temperature. So, cotton yield could be increased, and fewer chemicals should be used. Therefore, farmers' income could be improved. The main goals of doing this research are:

1. To completely understand the insect population, including their eggs and their parents.
2. To obtain more information regarding the plant type with a low bug population, go here.
3. To obtain preliminary information regarding the population concerning the temperature.

This work is organized as follows: Section 2 contains the literature review. Section 3 presents the methodology of the study. Section 4 discusses the results. In last, the conclusion and future work is described in section 5.

2. Literature Review

This Bulletin describes using a fungal pathogen with a native strain, V.lecanii, to control aphids and whiteflies in Korea. The highest cotton aphid mortality was produced by V. lecanii CS-625 among the six isolates gathered in Korea [5]. Based on bioassay and field research, the strain CS-626 was selected as the test subject, and a biological control agent for whiteflies was created. Insects are killed by pesticides. Dimethomorph and procymidone did not affect the germination of V. spores or the growth of mycelium [17]. The ebb and flow of insect populations are significantly influenced by the weather. One of the finest methods for tracking and controlling harmful insects is pest forecasting, which is especially helpful in areas where pest control is expensive [4]. The Pests Warning and Quality Control of Pesticides Department of Agriculture, Government of Punjab, Multan, collected data on pests over five years in several places around the Multan district to create a pest forecasting model [27]. The district as a whole has several of these locations. The association between weather and infestations of sucking insects was summarised using multivariate regression and correlation analysis approach. In this study, the Jassid Amrasca large tulle big tulle population averages, as well as those of other insects such as the Whitefly, Thrips, Cotton Mealybug, Dusky Cotton Bug, and Cotton Leaf Curl Virus, were examined CLCuV [3]. The results showed that all studied insects positively related to relative humidity. Between 2006 and 2008, cotton mealybug populations grew quickly, but after that, they started to decline as a result of changes in environmental factors [26]. Investigations were conducted on the Dusky cotton insect's resurgence as a new issue in the cotton crop. A regression study revealed a negative linear link between the highest temperature and the whitefly population, contributing 5.9 to 21.6%. Rainfall had a negative linear regression effect on the Jassid population from 1.3 to 3.4% [25]. The project's study was done in 2008 and 2009 during the Kharif seasons. The feeding harm caused by sucking insect pests was investigated in 19 genotypes, including 17 Bt hybrids, one conventional hybrid, and one variety, in the absence of pesticides. Sucking insect observations were meticulously recorded weekly [24]. Most leafhoppers were seen between the 28th and 32nd weeks of the year. The highest mean incidence rate ever recorded was in population H-1226. RCH 134 BG-II didn't reach the economic threshold level ETL in 2008, 6.00–6.23 adults per leaf, but it did in 2009. In 2009, the population exceeded it. The IT-905 Bt and RCH-134 BG-II strains had the highest population densities. While rainfall had a significant and opposite impact, temperature strongly and positively correlated with the frequency of leafhoppers and whiteflies. The genotypes were utterly identical to one another [29]. The evolutionary history and ecological ramifications of the Bemisia tabaci complex have been studied in Ecuador.

Samples of whiteflies were collected in Ecuador from nine different areas ranging in latitude from 2 degrees to 5 degrees south and in longitude from 78 degrees to 81 degrees west). On the 3′-mt COI-tRNAleu regions, genetic and adjusted pairwise distance studies were carried out. Thus, the mitotypes may be identified (832 bp) [22]. MaxEnt was used to simulate mitotype distributions and forecast mitotype niches based on environmental gradients. There is an imported B mitotype of B. tabaci in addition to the three native ECU mitotypes. Mitotypes ECU1 44%, ECU2 0.74%, and ECU3 1.47% were discovered in the American Tropics AMTROP species. These mitotypes diverged from one another by as much as 10%, surpassing prior AMTROP forecasts [23].

Between 7.7 and 8.6% of Ecuador's population has never been exposed to the ECU3 haplotype, which is unique from ECU1 and ECU2 in the American tropics. This discovery suggests that ECU3 is restricted to the region where it was found. The three ECU-endemic mitotypes were discovered in high-altitude pockets on the western side of the Andes [2]. These microclimates, located on the west side of the Andes, were characterized by sweeping changes in temperature and humidity. The non-endemic B mitotype was only present in irrigated agriculture systems in hot or dry tropical ecological niches, 47% [30]. ECU1 fared better than other native mitotypes in the most important places to the ecology. The amount and frequency of rainfall primarily affected where the ECU1 and B mitotype niche ranges were located. The previously assumed steady mt COI-tRNAleu sequence of B. tobacco was shown to be more variable than expected [31].

In the late phases of the new coronavirus, several factors are impacting people's food consumption, and the present forecast technique is hampered by duplicate data, resulting in low prediction accuracy [21]. Some elements are as follows: A prediction model was chosen in the context of big data, and its forecast accuracy was investigated to prevent a mismatch between supply [20] and demand in a catastrophic health crisis. As a result, there was no need to be concerned about an imbalance between the supply and demand for food. After several upgrades and accuracy testing, the modified grey Verhulst was discovered to be the most reliable and best suited for short-term forecasting [19]. In Africa, cassava production is severely constrained by disease outbreaks and insect pests. The production of the crop in Africa is endangered by an estimated yearly yield loss of $1.5 billion due to the cassava mosaic virus disease CMVD, which is brought on by the cassava mosaic Gemini viruses and spread by the whitefly Bemisia tabaci Gennadius [32]. The illness is brought on by cassava mosaic Gemini viruses Geminiviridae: Begum. In a six-month field research, cassava, commonly known as Manihot esculenta Crantz, was evaluated to determine if whitefly infections might be decreased by strip cropping Euphorbiaceae. A randomized full-block design was used to plant five rows of cassava, jatropha, and cotton with a 10-row (1-meter-wide) cassava row width around a 25-meter cassava plot [18]. The number of whitefly eggs, nymphs, and adults in each plot was counted and quantified every week. In plots with five rows of cotton and Jatropha curcas on either side, Bemisia tabaci populations were drastically decreased, demonstrating that strip cropping may be utilized as a management strategy to lower Bemisia tabaci numbers in the future [28]. To safeguard cotton crops, they employed pesticides like spirotetramat and buprofuzin [1]. Before using insecticides, as well as 2, 3, and 10 days afterward, whiteflies were counted. A careful analysis of the data showed that each pesticide produces a distinct set of effects. Dia-fenthion and spirotetramat, however, significantly affected the population and brought it to a specified level [16]. They have been demonstrated to be the most effective insecticides for reducing whitefly populations in actual field settings. Under Bahawalpur, the whitefly population was continuously highest in control circumstances following pesticide application, whereas it was lowest in experimental settings [15]. This was the situation in all regions. Compared to the plots treated with Spirotetramat 100 SC, the Dia-Ventura 500 SC-treated plots exhibited the lowest values two days, three days, seven days, and ten days after application. Si… Insecticides were employed to control their number, and whiteflies were counted before, two, three, seven, and ten days after that [14]. A careful analysis of the data showed that each pesticide produces a distinct set of effects. However, the most effective insecticides in field settings for lowering whitefly populations were dia-fenthion and spirotetramat. Despite a sharp fall, the population remained manageable [13]. Under control circumstances, Bahawalpur had the lowest whitefly population, whereas Cotton Research Station-treated regions had the highest. Two days, three days, seven days, and ten days after application, the plots treated with Dia-Ventura 500 SC had the lowest value compared to those treated with Spirotetramat 100 SC. According to earlier research from the farmer's field, the patches treated with Dia-Ventura 500 SC had the lowest population density after 2, 3, and 7 days. Ten days of waiting is also necessary. Dia-Ventura 500 SC may effectively control whiteflies. Hence a range of IPM strategies must be utilized in the field in conjunction with it. Dia-Ventura 500 SC was chosen as the appropriate population check as a result. A Shabbir et al. research Cassava mosaic disease SLCMV, an Asian illness caused by the genus Begomovirus and a member of the Geminiviridae family, has recently posed a danger to the cassava growing industry [12]. SLCMV transmission is essentially unknown where it has just lately been used. According to our examination into the efficacy of transmission by three whitefly species spread across Asia, Asian whiteflies can only effectively transmit SLCMV. It was discovered that the quantity of virus in the whitefly's whole body was positively connected with the amount of SLCMV that could be transmitted by various whiteflies [10]. Because the viral transmission is successful, different whitefly species have different viral transmission capacities. In-depth research on whitefly transmission of SLCMV has been done in this work, which will help understand SLCMV in the field and viral epidemic forecasting [11]. Associates 40 to 50%of the yearly loss in output brought on by infestations before and after harvest can be attributed to insects, pests, and diseases [9].

The loss of potential output is particularly acute in Asia and Africa. Transnational insect pests, a lack of access to crop protection chemicals, and a lack of preparation are the causes of this [8]. It is crucial to develop new, affordable, climate-smart pest control techniques digitally prepared for real-time pest forecasting and decision support systems DSSs [7]. To anticipate the first appearance of pests, the maximum severity/population of pests, and crop age, attempts have been undertaken to develop weather-based forewarning models for pests of Chickpea and Pigeonpea Helicoverpa, Phytophthora blight, and cotton. Stepwise regression and various meteorological indices were used to build the models. Techniques that make use of machine learning artificial neural networks and Bayesian networks. In general, the models provide a good match for all available data, and there is good agreement between the forecasts and the actual condition [6]. We will test the models' predictive accuracy using cross-year and cross-location models. In addition to boosting electronic and mobile engagement, a hybrid mobile app with Microsoft Azure will assist convey information to end users, which is vital for in-the-moment pest monitoring, prediction, and crop protection advice.

3. Materials and Methods

This section consists of an experimental procedure conducted to measure the classification accuracy of the proposed model. This study gathered two sets of greenhouse data, including the whitefly count, temperature, humidity, and light intensity, using a wireless imaging device. Wireless imaging devices equipped with Raspberry Pi 3 B+ embedded devices and Raspberry Pi v2 cameras make up the system. Two CNN models were employed to process the yellow sticky paper images, with the first model filtering out the insect objects and the second model classifying them. The whitefly count data was obtained using an insect pest counting algorithm with 97% accuracy in whitefly identification. In particular, 10 devices were installed in greenhouse No. (Figure 2) in Yunlin, Taiwan, which has a total area of 2208 m2, and 7 machines in greenhouse No. 2 in Chiayi, Taiwan, which has a total area of 529 m2. Tomato seedlings are the principal crop in both greenhouses. The number of whiteflies observed in a day is defined as the daily rise in whitefly count using the raw whitefly count. The data from greenhouse No. 1 was used for model training and testing from May 5, 2018, to December 9, 2018. On the other side, data from 2018/2/22 to 2018/9/10 was used for model validation in greenhouse No. 2 (Figure 3). An approach for fitting time series data to a model to derive future predictions from historical data is Autoregressive Integrated Moving Average (ARIMA). ARIMA has three components in theory: auto-regressive (AR), moving-average (MA), and integrated (I) terms.

ARIMAX transforms ARIMA models into multiple regression models by including exogenous variables X. (Figure 1) shows the flowchart for using the ARIMA (or ARIMAX) model.

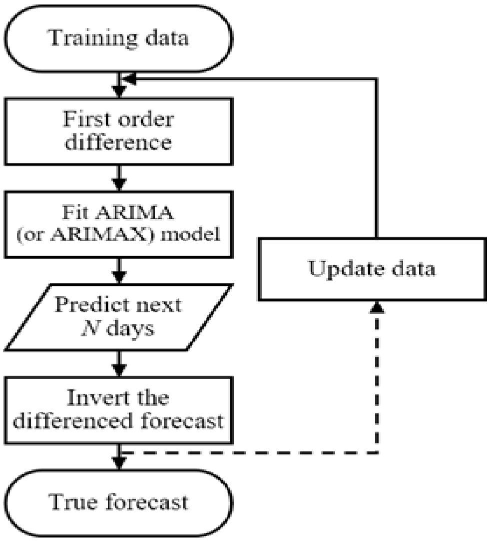

Figure 1. ARIMA (or ARIMAX) model training and forecasting flowchart

3.1. Data Collection

Before fitting the models, the input data should be stationary. This is because the mean and variance of fixed data remain consistent across time, making it easier to forecast. The ADF (Augmented Dickey-Fuller) test (=0.05) is employed in this study to determine whether the data is stationary. The latest 79 data points of greenhouse No. 1 are used for model testing out of a total of 199 data points.

Table 1. ADF Test for Stationary Testing

P-VALUE OF FACTORS				
	Count	Temp. (^{0}C)	Humidity (%RH)	Light intensity (lux)
Raw data	0.22	0.11	0.57	0.04
First order difference	2.13×10^{-6}	3.52×10^{-19}	2.21×10^{-19}	2.83×10^{-7}
	Count	Temp. (^{0}C)	Humidity (%RH)	Light intensity (lux)

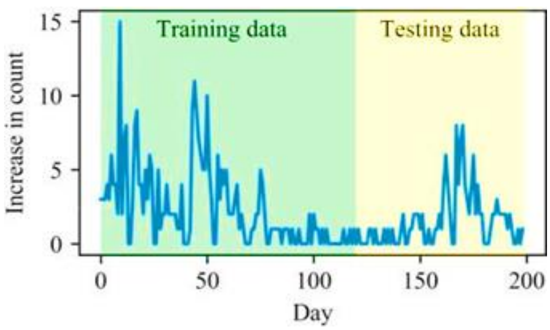

Figure 2. Raw whitefly count data of greenhouse No. 1

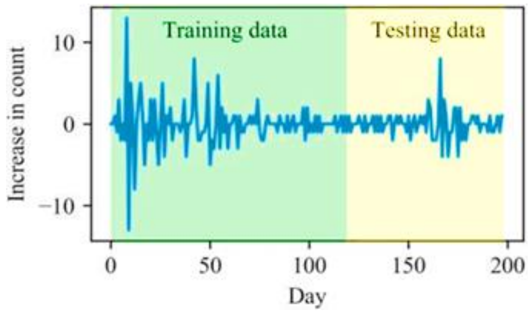

Figure 3. First Difference of Whitefly Count of Greenhouse No. 1

After each day, 7 days of expected data is forecasted in this study. The forecasting model was iteratively retrained to anticipate the next day's whitefly count by adding newly observed accurate data to the training data.

3.2 Data Preprocessing and Normalization

Data preprocessing is followed by data normalization and transformation in the data preparation process. Data preprocessing includes 1) data reduction and 2) null value eradication. The pest monitoring data set chosen for the study serves as the input to the data preprocessing phase. Due to unforeseen circumstances, such as inclement weather or a lack of a scout, the raw data may contain blank entries for specific dates in the study period. An artificial intelligence technique will likely deliver biased results if given data collection with open records. Sometimes, the complete dataset can be reduced to a regularized interval data format. A week's worth of data can be condensed into a single number by averaging the numbers for each day of the week and then summarizing the results. As a result, the reduced data for each week will not contain any open records. The existence of null values in the dataset is another crucial element that influences decision-making. The absence of a scout, inclement weather, or equipment failure could all explain the null values. The reduced dataset's null-valued tuples are deleted. Z-score normalization is applied to the preprocessed data to get a uniform distribution of values across distinct attribute groups. One of the most effective normalization methods is the Z-score approach, which is utilized in situations where the variables' highest and lowest possible values are unknown. To function, it computes the mean value of the attribute data and the standard deviation of that value. Any values in the dataset that were randomly distributed or noisy would be removed throughout the data preparation and normalization procedure. Data must first be handled in a way that can forecast the dynamics of the insect population and the crop. The suggested method uses preprocessed data spanning four to five years, with each year including 52 weeks of data. Training or testing the MLP neural network with this much data would be difficult. After the preprocessed dataset has been altered, fewer records will be considered for further analysis. Transforming (grouping) the data from three months (or twelve weeks) under a single label is the way to do so. A weighted network of input/output units can be considered one approach to conceptualizing a neural network. Both the training and testing processes go into building a neural network. It is possible to train the neural network to anticipate the call label of input samples during the training phase by adjusting the weights. This occurs once the training period has concluded. In addition to being able to recognize patterns in previously learned data, neural networks are highly resistant to input that has been tainted by noise. The proposed intelligent system utilizes a feed-forward MLP network as its neural network.

3.3 Proposed Architecture:

(Figure 4) sensors make up the initial input layer in an IoT design, a perception layer. The Wi-Fi modules in the gateway layer can transport data from the input layer to the server and store data in the storage layer. Finally, there is an application layer where pest predictions are delivered to farmers via the RBFN algorithm, which uses test and train data collected from the storage level to build and generate predictions.

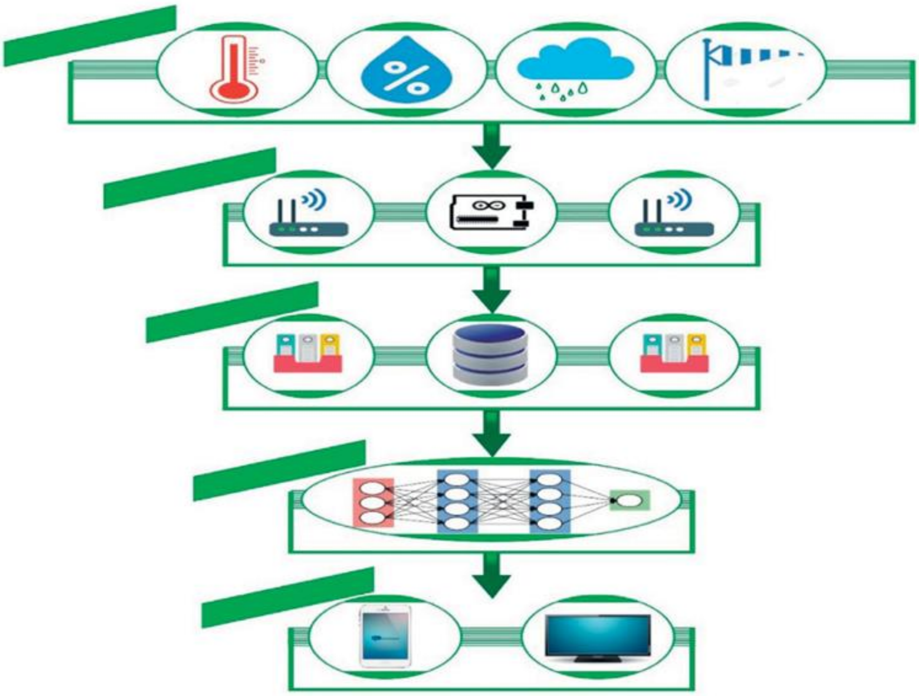

Figure 4. The layered architecture of the proposed model

The sensors could handle the innovative concept of forecasting insect attacks in the field, which is the problem's strength. Another strength of the problem is that it acquired real-time data from the area. Farmers can take safety measures dependent on their environment. RFID technology at the gateway layer, sensor technology at the input layer, communication technology Wi-Fi modules at the gateway layer, and cloud computing at the storage layer is the most essential critical Internet of Things technologies. There are many other major Internet of Things technologies. A method of deep learning known as RBFN was utilized to acquire the stated prediction. The performance of RBFN in binary class prediction while maintaining dataset independence led to its selection as the best option. Deep learning has a significant advantage over machine learning. Even though the machine learning algorithm currently provides a binary classification result, the model that has been suggested uses RBFN for the following five reasons:

1. When trained with a large amount of data, RBFN is the most accurate. The data on environmental characteristics such as temperature, humidity, rainfall, and wind speed has increased as time has passed.
2. When a large volume of data is processed by RBFN, it is more dependable.
3. When the amount of data is increased, accuracy improves.
4. RBFN approaches have an effective prediction decision support system.
5. With a problematic problem, RBFN has a high level of accuracy.

RBFN is made up of many levels. The RBFN Algorithm could be used to tackle the regression or classification problem. Deep learning's ability to automatically extract features is likely its most significant perk. Deep understanding has a massive impact on both the manufacturing and agricultural industries. Figure 3 depicts the two separate portions of a neuron that form a neural network. The activation function and the linear function are two different functions. The activation function is linear in the context of the nonlinear weights function. The RBFN has three layers: an input layer, a hidden layer, and an output layer. RBFN is constructed of three layers.

An input layer, a hidden layer, and an output layer are the three components of the RBFN algorithm.

1. The information that was gathered by the Pest Environment Monitoring System from its sensors can be found in the X input layer (PEMS). The input data consists of the temperature, humidity, precipitation rate, and wind speed.
2. The output layer presents the result as a Yes or No.

An example of RBFN implementation is shown in this section, together with the experiment's settings, experimental area layout, and prototype model deployment. The experiment will be carried out during the second season, which runs from May to November. The investigation will take place on one acre (43,560 ft2) with a length and breadth of 208 and 208 feet. A dataset with 416 rows and 62 columns was utilized to study whitefly assaults. There are 416 plants in each column. Two cotton plants occupy one foot of space. For the whitefly population projection, there are a total of 12,896 cotton plants. A sensor model is built into the design to track temperature, humidity, precipitation, and wind speed variables.

The predictive qualities of the sensing equipment can be described by the properties of the device. The DHT-22 is a temperature sensor that provides the most precise measurements of temperature and humidity in the surrounding environment. Low-cost and low-power, the DHT-22 is an ideal device. We have provided a sensor that measures the amount falling on the ground to detect rainwater. We've come up with this model. This gadget has many advantages, including its low power consumption and affordability. Using the rainfall detection sensor, the user can choose between digital and analog outputs. An easy-to-use rain sensor module can tell if there is rain or other precipitation in the area. The model we've provided uses a wind speed sensor to determine how much wind speed is present in the surrounding area. This is a reliable and constant sensor. An analog signal from the anemometer/wind speed sensor can be interpreted by a computer. The wind speed sensor is an easy-to-use device for measuring wind speed. The microcontroller used in this model is named a WeMos D1 Wi-Fi UNO-based ESP8266 shield for Arduino in our proposal. The predictive model, comprised of a Wi-Fi Arduino and other sensors, may be seen in (Figure 5). The developed model is used to investigate how whitefly infestation is affected by environmental factors such as temperature, humidity, precipitation, and wind speed. The construction of the hardware model and its subsequent deployment in the crop field is depicted in Figure 5.

The web application is built with the programming language PHP, and the database management system MySQL is utilized by the IoT web server. The web application will gather, evaluate, and store environmental information. The data gleaned from sensors is retrieved every four hours on average. In a web application, the libraries "ESP8266WiFi.h" and "DHT.h" are used to collect data from sensors, send it to the server, and then store it in the database (from May to November).

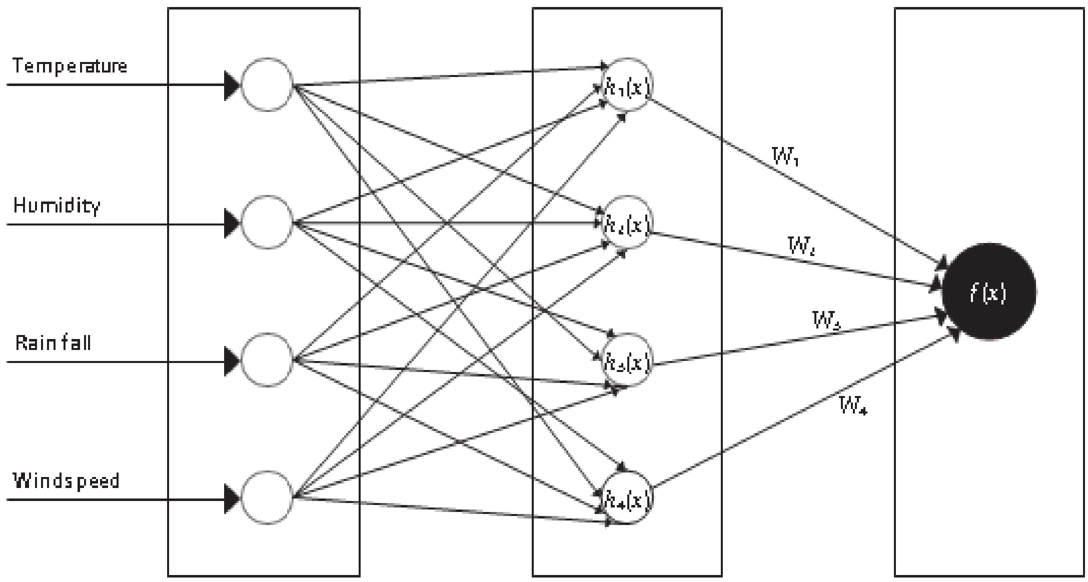

Figure 5. RBFN

Temperature and wind speed have a positive relationship with whitefly, but humidity and rainfall have a negative relationship. The number of whiteflies grows as the temperature and wind speed rise, while the number of whiteflies falls as the relative humidity and rainfall rise.

4. Results and Discussions

The test and training data sets are kept apart from the remaining data. The training data supplied into the neural network during training helps it learn. All three methods considered for pest prediction are made to foresee different crop bug types. The three systems were also created utilizing three various neural networks. The results presented above merely provide an overview of how well each of the three systems performed (prediction accuracy). The results demonstrate that the recommended intelligent system for predicting pest population dynamics of Thrips Tabaco Linde (Thrips) on cotton (Gossypium Arboreum) crop performs quite well compared to the other two systems investigated.

Table 2. Cotton Leaf Worm ML Analysis us Random Sampling

	RF		ExtraTrees		Line	Ar	XGB		Logistic	
	RRMSE	**MAPE**	**RRMSE**	**MAPE**	**RMS**	**MAPE**	**RRMSE**	**MAPE**	**RRMSE**	**MAPE**
Basic	29,11%	23,66%	25,28%	20,56%	46,23%	40,39%	22,77%	24,12%	53,62%	17,65%
Prev IS	29,88%	21,87%	25,93%	19,86%	34,03%	20,26 %	32,32%	19,13 %	26,13%	32,34 %
Temp	29,85%	21,16%	22,10%	19,14%	26,56%	32,12%	23,73%	26,44%	36,26%	22,30%
Thrips	29,42%	23,60%	23,52%	20,61%	45,46%	30,235%	30,56%	20,91%	58,53%	25,66%
Total	29,46%	23,14%	26,10%	18,13%	34,39%	20,26%	26,31 %	19,82%	41,32%	32,56%

Table 2 proposed a model (RBFN) capable of making decisions to predict whitefly attacks. We saw how to use layered designs and prototypes in the last parts. Four sensors (temperature, humidity, rainfall, and wind speed) and a microcontroller were installed in the chosen zone to assess the result.

Table 3. The Confusion Metrics of Results Are Precision, Recall, F1, and Support

Class	F_1	Recall	Precision	Support
0.0	0.52	0.39	0.85	29
1.0	0.94	0.99	0.85	86
Macro avg.	0.73	0.69	0.85	112
Weighted avg.	0.82	0.85	0.85	112

The accuracy of the RBFN algorithm is measured using the Python library "sklearn. metrics," with RBFN, high F1, precision, recall, support, Cohen's kappa, ROC AUC, and log loss, as shown in Table 3.

Table 4. Performance Comparison of Different Classifiers

Classifier	Accuracy	Sensitivity	Specificity	PPA	NPA
SVM	98.56%	98.92%	97.82%	98.90%	97.82%
ANN	90.10%	93.13%	83.39%	90.69%	88.92%
Bayesian	93.90%	98.83%	86.32%	91.96%	97.86%
Binary Deci-sionTree	90.10%	93.12%	83.39%	90.59%	88.92%
k-NN	93.28%	96.64%	87.59%	92.97%	93.43%

Table 4 SVM classifier is trained using several kernel functions in this study. After training, according to the experiment, the SVM classifier may obtain the best accuracy of 98.56% utilizing the Radial Basis Function (RBF) kernel. So, we employed the cross-validated trained classifier for additional testing on new images with RBF kernel. The performance of all five trained classifiers is evaluated using 130 test photos (45 regular and 85 impacted) that were not included in the training set.

The remaining parameters, such as dropout = 0.1, must be changed for the model to perform better after the basic LSTM network structure has been built. Unlike conventional machine learning algorithms, LSTM networks allow network parameters to be changed in response to new data without restarting feature selection or recreating networks. Real-time changing of the network settings depending on the most recent input data can also be used to forecast the occurrence of other pests.

Table 5. Predictions of Different Kinds of Pests and Diseases with the LSTM Network

Metrics	Bollworm	Whitefly	Jassid	Leaf Blight
ACC	0.9315	0.9332	0.9432	0.9607
AUC	0.9768	0.9776	0.9834	0.9889
F1-score	0.8948	0.9334	0.9222	0.9281

Table 5 compares the performance of LSTM networks on various datasets. This demonstrates our model's great generalization capacity and ability to accurately forecast diseases and pests. The results indicate the LSTM network's suitability for predicting conditions and problems that affect cotton, providing the theoretical foundation for its potential practical application.

His idea consists of three essential parts: (i) abiotic variables that affect population net growth rate, (ii) biotic variables that regulate fitness, and (iii) variables relating to spatial movement. In the proposed system, the ARIMAX technique was used, and several combinations of input data were tested to establish the essential input data for forecasting. C stands for the number of whiteflies found each day, T for the ambient temperature (0°C), H for relative humidity (% RH), and L for light intensity (lux). In addition, the minimal number of input training data or days to the model was evaluated to see if the method could be used effectively on different greenhouses.

Different training days, such as 15, 30, 60, 90, and 120, were investigated in this study. Table 6 shows the performance of ARIMA and ARIMAX models trained with various combinations of input data.

Table 6. RMSE ± STD Of Different Factors with a Different Number of Training Days where C, T, H, and L are The Daily Increase in Whitefly Counts, Temperature, Humidity, and Light Intensity, Respectively.

Factors	**Number of the training day**				
(difference)	**15**	**30**	**60**	**90**	**120**
C	3.41+3.62	2.16+1.36	2.21+1.43	2.21+1.43	2.21±1.43
C+T	2.66+1.95	2.21+1.43	1.58+1.04	1.69+1.09	1.66±1.09
C+H	2.61+1.69	1.98+1.26	1.89+0.82	1.82+1.18	1.63±0.99
C+L	2.21+1.43	1.74+1.08	1.70+1.09	1.84+1.26	1.70±1.14
C+T+H	3.14+2.30	2.21+1.48	1.62+0.72	1.84+1.26	1.56±0.91
C+T+L	3.26+2.53	2.09+1.31	1.74+1.15	1.53+1.24	1.75±1.14
C+H+L	3.65+2.87	1.95+1.27	1.82+1.26	1.92+1.27	1.72±1.13
C+T+H+L	10.56+11.88	2.02+1.30	1.82+1.10	1.67+0.95	1.62±0.96

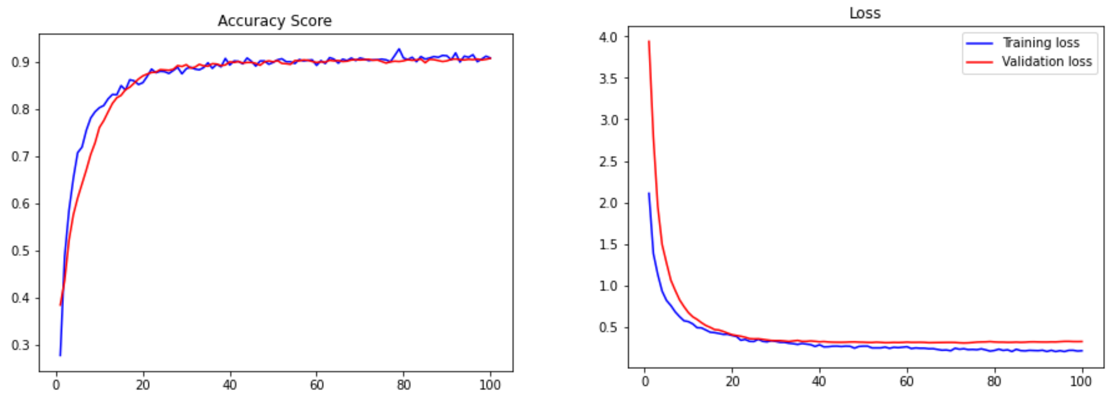

Figure 6. Accuracy and Loss Curves

The results (Figure 6) suggest that C+T+H was the best combination of input data. It also demonstrates that when training data is 60 days long, most models have the lowest RMSE, and the RMSE of the ARIMAX models is lower than the ARIMA model. Regarding the minimal number of input data, the findings were satisfactory at roughly 30 days because the RMSE values were less than 3. The curve in Fig 5 shows the accuracy and loss of the proposed system.

5. Conclusion and Future Work

Using machine learning tools, this research forecasts the whitefly population on cotton. The weather is taken as a parameter, and the whitefly's people are predicted. Earlier it was done by the use of statistical methods. With temperature and humidity as exogenous inputs, the results of this study suggest that ARIMAX is the best model for forecasting whiteflies on cotton. The accuracy of this model was almost equal to the statistical models used for forecasting. The proposed methodology is faster than the statistical models and saves energy. So ARIMAX can also be used for forecasting purposes because it is a statistical model.

The data is gathered manually in the present work, which takes more human effort and time. In the future, sensors can be used to collect data, saving human energy. In the end, a mobile will be developed that will be connected to the model, and that will send alerts to the farmers on their mobile phones indicating the current population of the whiteflies in the field, and it will also send the prediction of whiteflies. So that the farmer will be updated about the current state or population of whiteflies and can take precautions. In the future, this system can also be used for forecasting other insects and pests in cotton or other plants.

Funding: No funding was received.

Data Availability Statement: The authors declare that all data supporting the findings of this study are available within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kumar, S., Jain, A., Shukla, A. P., Singh, S., Raja, R., Rani, S., ... & Masud, M. (2021). A comparative analysis of machine learning algorithms for detection of organic and nonorganic cotton diseases. Mathematical Problems in Engineering, 2021.
2. Ali, S., Badar, N., & Fatima, H. (2015). Forecasting production and yield of sugar cane and cotton crops of Pakistan for 2013-2030. Sarhad Journal of Agriculture, 31(1), 1-10.
3. Anwar, W., Javed, M. A., Shahid, A. A., Nawaz, K., Akhter, A., Ur Rehman, M. Z., ... & Haider, M. S. (2019). Chitinase genes from Metarhizium anisopliae for the control of whitefly in cotton. Royal Society open science, 6(8), 190412.
4. Aswathi, V. S., & Duraisamy, M. R. (2018). Comparison of Prediction Accuracy of Multiple Linear Regression, ARIMA and ARIMAX Model for Pest Incidence of Cotton with Weather Factors. Madras Agricultural Journal, 105(September (7-9)), 313-316.
5. Chen, W., Hasegawa, D. K., Kaur, N., Kliot, A., Pinheiro, P. V., Luan, J., ... & Fei, Z. (2016). The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. BMC biology, 14(1), 1-15.
6. Chi, Y., Pan, L. L., Bouvaine, S., Fan, Y. Y., Liu, Y. Q., Liu, S. S., ... & Wang, X. W. (2020). Differential transmission of Sri Lankan cassava mosaic virus by three cryptic species of the whitefly Bemisia tabaci complex. Virology, 540, 141-149.
7. Chiu, L. Y., Rustia, D. J. A., Lu, C. Y., & Lin, T. T. (2019). Modelling and forecasting of greenhouse whitefly incidence using time-series and ARIMAX analysis. IFAC-PapersOnLine, 52(30), 196-201.
8. Dey, A., Bhoumik, D., & Dey, K. N. (2016). Automatic detection of whitefly pest using statistical feature extraction and image classification methods. Int. Res. J. Eng. Technol, 3(09), 950-959.
9. Fenu, G., & Malloci, F. M. (2021). Forecasting plant and crop disease: an explorative study on current algorithms. Big Data and Cognitive Computing, 5(1), 2.
10. Guo, R. (2021, June). Research on Prediction Accuracy of Mathematical Modeling Based on Big Data Prediction Model. In Journal of Physics: Conference Series (Vol. 1952, No. 4, p. 042122). IOP Publishing.
11. Amjad, A., Javed, M. S., Hameed, A., Khan, A. A., Amjad, M. R., Ali, S. A., ... & Raza, A. (2021). Improving the Probiotics Viability and Quality Characteristics of Yoghurt Enriched with Barley Bran.
12. Swathi, M., Gaur, N., & Singh, K. Virus Vector Relationship of Yellow Mosaic Virus and Whitefly, Bemisia tabaci (Gennadius) in Soybean. Legume Research-An International Journal, 1, 5.
13. Kalkal, D., Lal, R., Dahiya, K. K., Singh, M., & Kumar, A. (2015). Population dynamics of sucking insect pests of cotton and its correlation with abiotic factors. Indian Journal of Agricultural Research, 49(5), 432-436.
14. Kataria, S. K., Pal, R. K., Kumar, V., & Singh, P. (2019). Population dynamics of whitefly, Bemisia tabaci (Gennadius), as influenced by weather conditions infesting Bt cotton hybrid. Journal of Agrometeorology, 21(4), 504-509.
15. Keswani, B., Mohapatra, A. G., Mohanty, A., Khanna, A., Rodrigues, J. J., Gupta, D., & De Albuquerque, V. H. C. (2019). Adapting weather conditions based IoT enabled smart irrigation technique in precision agriculture mechanisms. Neural Computing and Applications, 31(1), 277-292.
16. Khan, M., & Damalas, C. A. (2015). Farmers' knowledge about common pests and pesticide safety in conventional cotton production in Pakistan. Crop Protection, 77, 45-51.
17. Xiao, Q., Li, W., Kai, Y., Chen, P., Zhang, J., & Wang, B. (2019). Occurrence prediction of pests and diseases in cotton on the basis of weather factors by long short term memory network. BMC Bioinformatics, 20(25), 1-15.
18. Khuriyati, N., Nugroho, D. A., & Wicaksono, N. A. (2020). Quality assessment of chilies (Capsicum annuum L.) by using a smartphone camera. In IOP Conference Series: Earth and Environmental Science (Vol. 425, No. 1, p. 012040). IOP Publishing.
19. Kumar, S., Jain, A., Shukla, A. P., Singh, S., Raja, R., Rani, S., ... & Masud, M. (2021). A comparative analysis of machine learning algorithms for detection of organic and nonorganic cotton diseases. Mathematical Problems in Engineering, 2021.
20. Xiao, Q., Li, W., Kai, Y., Chen, P., Zhang, J., & Wang, B. (2019). Occurrence prediction of pests and diseases in cotton on the basis of weather factors by long short term memory network. BMC Bioinformatics, 20(25), 1-15.

21. Ahmad, S. (2022). Population assessment of Bemisia tabaci (Gennadius.) and disease occurrence of Begomovirus in okra, Abelmoschus esculentus L. Journal of the Saudi Society of Agricultural Sciences, 21(2), 77-86.
22. Haarith, D., Kim, D. G., Chen, S., & Bushley, K. E. (2021). Growth chamber and greenhouse screening of promising in vitro fungal biological control candidates for the soybean cyst nematode (Heterodera glycines). Biological Control, 160, 104635.
23. Russia, D. J. A., Lin, C. E., Chung, J. Y., Zhuang, Y. J., Hsu, J. C., & Lin, T. T. (2020). Application of an image and environmental sensor network for automated greenhouse insect pest monitoring. Journal of Asia-Pacific Entomology, 23(1), 17-28.
24. Li, Y., & Yang, J. (2020). Few-shot cotton pest recognition and terminal realization. Computers and Electronics in Agriculture, 169, 105240.
25. Liu, S., Yang, Z., & Volodymyr, V. (2022). Importance of Bemisia tabaci Forecasting Technology: A Review. Advances in Entomology, 10(2), 149-158.
26. Patil, J., & Mytri, V. D. (2013). A prediction model for population dynamics of cotton pest (Thrips tabaci Linde) using multilayer-perceptron neural network. International Journal of Computer Applications, 67(4).
27. Saleem, R. M., Kazmi, R., Bajwa, I. S., Ashraf, A., Ramzan, S., & Anwar, W. (2021). IoT-Based Cotton Whitefly Prediction Using Deep Learning. Scientific Programming, 2021.
28. Sseruwagi, P., Sserubombwe, W. S., Legg, J. P., Ndunguru, J., & Thresh, J. M. (2004). Methods of surveying the incidence and severity of cassava mosaic disease and whitefly vector populations on cassava in Africa: a review. Virus Research, 100(1), 129-142.
29. Tageldin, A., Adly, D., Mostafa, H., & Mohammed, H. S. (2020). Applying machine learning technology in the prediction of crop infestation with cotton leafworm in greenhouse. bioRxiv.
30. Vyas, M., Raza, A., Ali, M. Y., Ashraf, M. A., Mansoor, S., Shahid, A. A., & Brown, J. K. (2017). Knockdown of whitefly gut gene expression and mortality by orally delivered gut gene-specific dsRNAs. PLoS One, 12(1), e0168921.
31. Wang, X. W., Luan, J. B., Li, J. M., Su, Y. L., Xia, J., & Liu, S. S. (2011). Transcriptome analysis and comparison reveal divergence between two invasive whitefly cryptic species. BMC genomics, 12(1), 1-12.
32. Xiao, Q., Li, W., Kai, Y., Chen, P., Zhang, J., & Wang, B. (2019). Occurrence prediction of pests and diseases in cotton on the basis of weather factors by long short term memory network. BMC Bioinformatics, 20(25), 1-15.
33. Meena, R. S., Ameta, O. P., & Meena, B. L. (2013). Population dynamics of sucking pests and their correlation with weather parameters in chilli, Capsicum annum L. crop. The Bioscan, 8(1), 177-180.