International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-06, Issue-11, FEB 2021

Recommendation System: Pest Prediction using Machine Learning

Rutuja Mohekar, B.E. Student, Sipna College of Engineering and Technology, Amravati, Maharashtra, India, rutujamohekar@gmail.com

Rishant Mohekar, B.E. Student, Dr. D. Y. Patil Institute of Engineering, Management and Research, Pune, Maharashtra, India, morish02@hotmail.com

Abstract - In addition to the previous work in crop prediction, we have proposed a crop recommendation system as well. An intelligent prediction model for pests is proposed by combining artificial neural network and genetic algorithm and the recommendation is based on various environmental factors such as temperature, rainfall, soil type, sunlight, pH value, and season of cultivation. In order to overcome the disadvantages of a single backpropagation network, the network structure is constructed with the historical pest's data, the global optimization characteristic of genetic algorithm is obtained, and the weights and thresholds of the neural network are trained to overcome the main shortcomings of the single back propagation network. The model output is quickly and accurately forced into the target sample. The prediction method is more accurate than that of the traditional backpropagation artificial neural network. The prediction accuracy and degree of prediction have been greatly improved and is as high as 91.67%. Based on this, the intelligent model is used to predict the insect pests of plants, and a new way of thinking is established in the field of prediction of pests, which greatly promotes the rapid development of agriculture. This system is available for everyone, no need to even register, anyone can use the system, and the structure and flow of the project is maintained simple, to be understood by each and every one.

Keywords — Prediction model, Naive Bayes algorithm, Neural Network, Precision Farming, Recommendation system, Structural Similarity Index (SSIM).

I. INTRODUCTION

The pest on crops causes serious harm to crop growth, resulting in a serious reduction of production. Therefore, rapid, and accurate prediction of pest damage is the main method to make correct prevention measures and reduce pests. Traditionally, pest prediction methods include observation method, statistical method, and mathematical ecological model method. In rural areas, the establishment of basic insect forecasting stations, but these methods have some limitations due to their shortcomings, such as empirical prediction of human factors are obvious, the accuracy of the correlation coefficient used is low. Therefore, by using Deep Learning technology, we can accurately detect the pests and disease in the farms. In recent years, some researchers have put the Back Propagation Artificial Neural Network compared with the traditional method Back Propagation Neural Network improves the prediction accuracy and ease of use greatly, but there are two serious defects: slow convergence speed and easy to fall into a minimum point. Considering the effects of multivariable, time-varying, and uncertain factors on insect pests, it is necessary to establish a pest model with high prediction efficiency, accuracy, and precision. Firstly, based on the Back Propagation neural network, the existing weather is used.

A. Automation in agriculture

From the AI point of view, Agriculture offers numerous application areas. All kinds of AI core technologies such as Mobile, autonomous agents operating in uncontrolled environments, stand-alone or in collaborative settings, allow to investigate, test, and exploit technologies from robotics, computer vision, sensing, and environment interaction. Integrating multiple partners and their heterogeneous information sources leads to the application of semantic technologies. The complexity of the agricultural production seeks progress in modeling capabilities, handling of uncertainty, and in the algorithmic and usability aspects of location and context-specific decision support. The growing interest in reliable predictions as a basis for planning and control of agricultural activities needs interdisciplinary cooperation with domain experts e.g., from agricultural research. Modern agricultural machines shall use self-

 DOI : 10.35291/2454-9150.2021.0037

configuring components, thus be able to collaborate and exhibit aspects of self-organization and intelligence.

The prediction model is established by using the historical data of insect pests, and then the genetic algorithm is used to search for the special characteristics of population optimization. By optimizing the weights and thresholds of the network, the pest occurrence degree can be predicted quickly and accurately. As it is said, Prevention is better than cure, our project aids framers to accurately predict pest as well recommend crop suitable to their specifications. Automation technology is the most focused technology by the Indian start-ups. Automated Drones and Bots are deployed in farms for monitoring purposes. Day by day the integration of technology is swapping from normal spraying to specified target spraying of pesticides and fertilizers on the crops. Artificial Intelligence, Machine Learning, and Deep Learning algorithms are deployed to monitor the crops precisely, hence spray corrective solutions in that specific target area for precision farming.

B. Need

It causes economical loss to framers as weather changes pests and insects develop on the crops and destroy the yield. Farmers can inspect manually but these can take time as well as accuracy concern is present, so there arises a need to modernize the traditional agriculture system and with the assistance of the prediction system these problems can be minimized. With the facility of the recommendation system, it will be best to plant crops suitable to the climate conditions, land specifications, and other environmental factors. Data mining is the rehearsal of examining and developing purposeful knowledge from the data. Data mining discovers its application in various fields like finance, retail, medicine, agriculture, etc. Data mining in agriculture is used for investigating the various biotic-abiotic factors. Agriculture in India plays a major role in the economy as well as in employment. The joint problem existing among the Indian farmers is they don't pick the right crop based on their soil requirements. Due to this, they encounter obstacles in productivity. This problem of the farmers has been talked about through precision agriculture. Precision agriculture is a modern farming method that uses research data of soil characteristics, soil types, crop yield data collection, and proposes the farmers the right crop based on their site-specific parameters. It visibly lessens the wrong decision on a crop and promotes precision in framing. [6].

C. Motivation

The motivation of the project is to digitalize our daily life as well as agriculture in our country. In many countries, this prediction and recommender system is available and popular. And we want to apply this system in our country, as a part of achieving and stepping towards modernization of Agricultural techniques.

D. Objective

The objectives are:

- Optimize Yield: With easy and fast prediction, no time is wasted in the pest prediction and we can ensure no crop gets damaged by early detection and prevention of the pest. This increases the yield of crops and a lesser loss to the farmer.
- Loss Reduction: As the pest is detected before any damage to the crop there are very few chances of any kind of loss to the yield. Prevention of pests and recommendations can help to reduce any kind of loss of crops.
- Appropriate cultivation: Based on the recommender system we can analyze the suitable crops as per various factors like weather, sunlight, humidity, etc. and manage the appropriate cultivation of only those crops suitable.
- Increase Accuracy: Most of the pest prediction techniques don't ensure great accuracy as they might develop certain errors during prediction. But the recommender system uses proper algorithms to ensure complete accuracy without any errors.
- Integrate technology in the Agriculture system: Most of the traditional agriculture systems don't have any kind of modern prediction system. This kind of technological system integrates all the modern techniques and helps in greater yield of the agriculture system.
- No or Minimum Cost Solution: The cheapest method for accurate pest prediction can be done using the recommender system with machine learning. This project once built can lead to only minimal costs and greater efficiency as compared to other methods.

II. LITERATURE REVIEW

Many approaches are given for predicting the agriculture crop outcomes which given in this section Pratheepa.et al [1] has developed a Classification model for forewarning about the cotton crop pest and also identifying the factors influencing the pest population density which will further help the farmer to apply pest control strategies on time to reduce crop loss. The review has charted some capable techniques that have been used to recognize the relationships of various climate and other factors on crop production. This review suggests that further examinations are needed to understand how these techniques can be used with complex agricultural datasets for crop yield prediction integrating seasonal and spatial factors by using GIS technologies. Cintra.et al [2] used a fuzzy decision tree model for giving coffee plant rust warnings where the author has constructed the six datasets according to two

 DOI : 10.35291/2454-9150.2021.0037

distinct infection rates. while estimated monthly disease rates reach one of the two thresholds then this model can be used to triggers warnings and alerts. The author also showed the compression of all decision tree algorithms and showed that the fuzzy decision tree model gives better accuracy power and predictability. Hong.et al [3] has given Data-Driven Approach for collecting soil moisture where he develops a framework to predict the soil moisture using SVM and RVM that is relevance vector machine where they used historic field-collected sensor data where author achieved low error rates and the highest correlation between predicted values and actual values. Rahman.et al [4] given approach to predict the yield of crops using machine learning model classifiers, where this model is first trained by the author on the correlation between the past environment patterns and crop production rates, then the models are compared to measure their effectiveness on unknown climatic variables. The clue is to pool spatial information mining/decision tree strategies with master framework strategies and apply them to shape a savvy agriculture arrive reviewing data framework. Tripathy.et al [5] experimented with four consecutive agriculture seasons to understand the relation between crop-weather-pest/diseases using wireless sensor and field level surveillance data using Association Rule mining and multivariate regression mining technique were made to information about the relationship, then author develops web-based decision support system using the information.

Data mining is the rehearsal of examining and developing purposeful knowledge from the data. Data mining in agriculture is used for investigating the various biotic-abiotic factors. The joint problem existing among the Indian farmers is they don't pick the right crop based on their soil requirements. Due to this, they face a serious obstacle to productivity. This problem of the farmers has been talked about through precision agriculture. Precision agriculture is a modern farming method that uses previous data of soil, crop yield data. Then proposes the right crop based on their site-specific parameters. This decreases the wrong decision on a crop and increase productivity [6].

III. IMPLEMENTATION

The System developed would allow users to detect the pest on their crop within a minimum amount of time and effort. The user needs to upload a photo of their infected crop and the system will predict the pest or disease-infected their crop. This would help them identify and implement suitable techniques to overcome the condition of the crop. To implement the model we require Python IDLE, Tkinter as GUI, Python as programming language, and OpenCV. Python IDLE is a basic programming Integrated Development and Learning Environment. Tkinter is the fastest and easiest tool to model GUI applications. OpenCV is an open-source computer vision python library which is mainly deployed in real-time computer vision. This library is used for Image Processing, it performs all the operations related to Image processing.

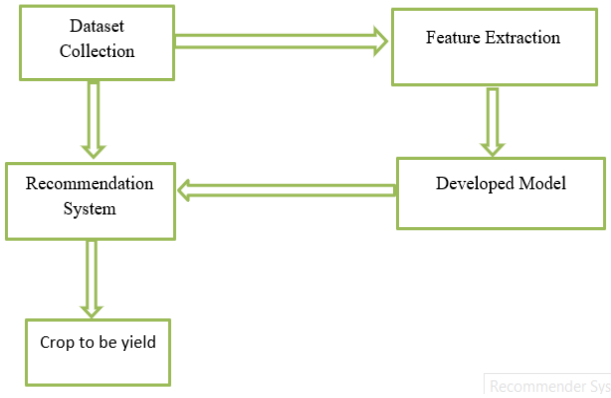

Fig.3.1. Overall methodology of the developed system.

A. Proposed System

In Metropolitan areas of numerous countries prefer the use of technology in agriculture. Finding out manually requires a deep sense of knowledge and accuracy rate is the main concern. Hence there is a need for assistive technology, which would ease the task for agricultural producers. The System developed would allow users to detect the pest on their crop within a minimum amount of time and effort. The user needs to upload a photo of their infected crop and the system will predict the pest or disease-infected their crop. This would help them identify and implement suitable techniques to overcome the condition of the crop.

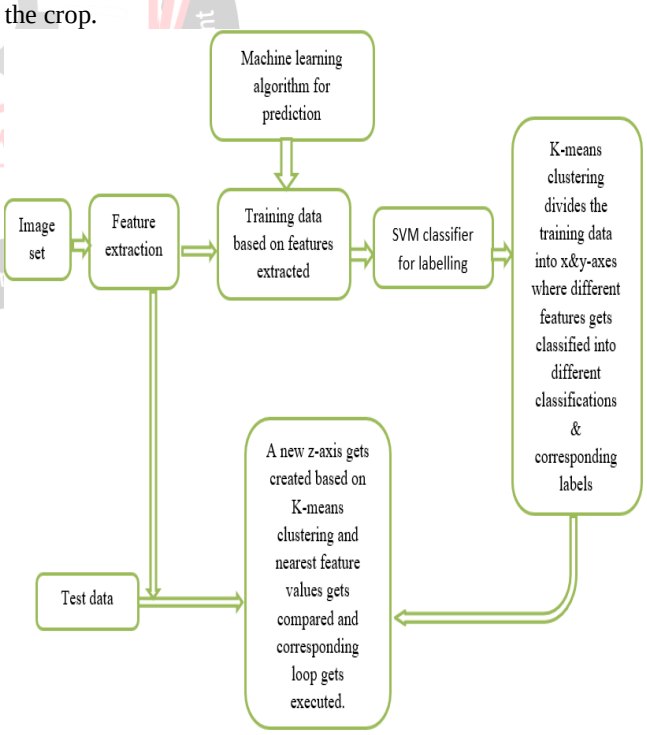

Fig.3.2. Overall Working of the system

The proposed system consists of phases. Each of the phases is explained below:

1.Development of Prediction and Recommendation System

2. Training and Testing the model

 DOI : 10.35291/2454-9150.2021.0037

3. Feature Extraction from Images
4. Classify Images
5. Getting to the detail of Infected Pest or Disease
6. Questionnaire-based on Recommendation System

The complete system consists of these 6 phases, the user can directly access and make use of the recommendation as well as the prediction system without authentication, hence no hassle of login. Whereas the admin needs to get authentication via admin login and password. Admin can upload more datasets for training as well as for testing and can also add a questionnaire for the recommendation system. The admin also has the provision to all the functionality of the system developed.

B. System Inputs and Outputs

The Pest Prediction and Recommender System provides different inputs and outputs services. These services will be explained in detail according to their use.

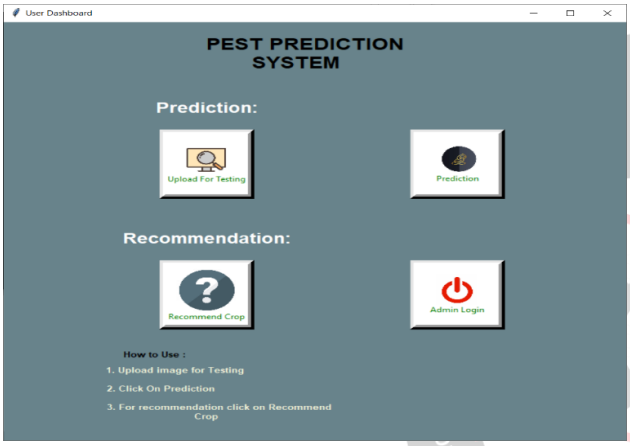

Fig.3.2.1. User Dashboard

- Users

Users can make use of the application to predict pest and diseases infected or check the suitable crop according to their environmental factors wherever they are.

- Administrator

The administrator is responsible for adding, updating, and maintaining the system. They also can use the functionalities of the system.

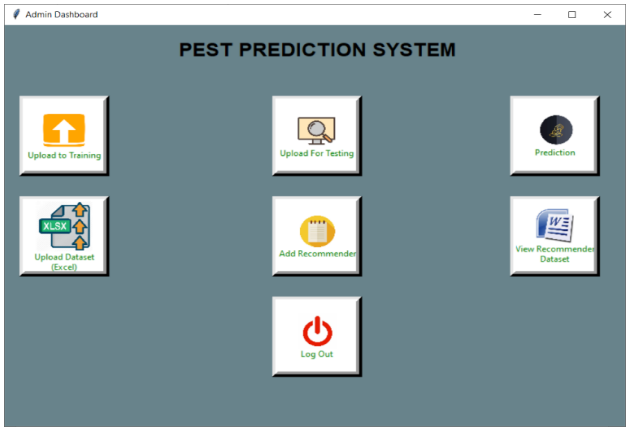

Fig.3.2.2. Admin Dashboard

C. System Architecture

The design of the system architecture includes the structure, behavior, and more views of the system. The goal of design is to provide a module of the system which is employed to create the system within the proposed system. Initially, once the user uploads a picture into the developed system, the user can detect the pest and disease that occurred to their crop and get a solution to it. The user can upload images and get predictions and recommendations of crops. The system is developed in stages, such as Data collection, Data Cleaning and preprocessing, building a prediction and recommender model, training, and testing.

a. Dataset Collection

Dataset is taken from Kaggle for prediction and the Indian government agricultural portal www.data.gov.in. for recommendation. The prediction dataset contains images and corresponding disease name and the solution to it. This is an advancement in the existing system to give a proper solution to the predicted disease. on the other hand, the crop recommendation dataset contains temperature, rainfall, soil type, sunlight, pH value, and cropping season. Data Attributes such as Rainfall, Temperature, Sunlight availability, and Soil type are referred from [11] whereas data attributes such as Fertilizers, Nitrogen, Phosphorous, and Potassium Values, Organic Carbon value, and Soil pH value are taken from [12].

b. Data cleaning and preprocessing

The basic data cleaning and preprocessing tasks such as data integration, attribute selection, attribute extraction, and data reduction are implemented in this stage. Data integration refers to data collection from different resources and then combining them into a single dataset to build the prediction and recommender model. Based on the correlation measure, all attributes are evaluated and selected. Their correlation value signifies crop. All these attributes are used in developing machine learning models. Data reduction is implemented to reduce the noise in the data and the repetition of data. The selected features and instances are extracted and stored as .csv files depending on the results of feature selection.

The system receives an image from the user in RGB format. The provided image may contain noise and removing noise image is cleaned and preprocessed. To carry out pre-processing the image undergoes various modifications such as segmentation, conversion into greyscale, and application of threshold. The three-channel image is then converted into a single channel, i.e., from RGB to binary or greyscale and it is applied threshold for better analysis.

 DOI : 10.35291/2454-9150.2021.0037

Fig.3.3.1. Image preprocessing

c. Building a prediction and recommender model

Naive Bayes and K-means clustering and SSIM are considered in this research work. The final recommendations are generated by association rules which depend upon the results of the classifiers. Assigning an entity to a predefined class by analyzing the features is called as classification.

Algorithms used in the development of the prediction model are:

- Association: Used to identify the relationship between different dimensions.
- Clustering: Used to group together similar attributes to form a cluster.
- Regression: Used to obtain Value dependent variable as a function of the independent variable and set of parameters.
- Classification: Used to classify different items into class.
- K-means clustering: Usage is like clustering only difference is that items will be clustered into k groups.
- Naive Bayes: Used to predict the probability of different classes based on various attributes.

For recommendation, we have used the Structural Similarity Index (SSIM) to compare similarity.

- *Naive Bayes algorithm*

  The Naive Bayes algorithm is based upon the Bayes rule of conditional probability. It uses the attributes defined in the dataset. All the attributes are analyzed individually because these are equally important and independent of each other. It uses the probabilistic components of a structure, as contrasting to deterministic comparisons to describe the connections among variables.

- *Support Vector Machine (SVM)*

  A support vector machine (SVM) is a supervised machine learning model in machine learning. Two-group classification problems are solved using classification algorithms. An SVM model is provided with a set of labeled training data for each category which is then ready to categorize images. The application reported for SVM was the modeling of urban land use conversion. This research derived the association between rural-urban land-use change and various factors. SVM was also applied to deliver visions into crop response patterns. Affected by climate conditions by providing the structures contribution analysis for agricultural yield prediction. Thus, analyzed different possible deviations of the weather scenarios using SVMs. We have used SVM to categorize images in our model.

- *Structural Similarity Index (SSIM)*

  The Structural Similarity Index (SSIM) is a metric that evaluates the quality of an image. Image degradation can be caused by processing such as data compression or by losses in data transmission. SSIM is used for measuring the similarity between the testing image provided by the user and the images in the database. It is a full reference metric that requires two images from the same image to capture a reference image and a processed image. SSIM has applications in a variety of different problems such as Image Compression, Image Restoration, and Pattern Recognition.

  We have used SSIM to compare the given image with the training dataset images. The value of SSIM ranges from 0 to 1, 1 being the highest. The value is set to greater than 0.8, which means if the given image matches with the dataset then it shows the details related to the disease and solution to it.

d. Training the model

The dataset is divided into two training dataset and the testing dataset. The dataset is split into 80-20%, where 80% is used for training purpose.

e. Testing the model

The developed model is tested for various performance measures. 20% of the dataset is used as a testing dataset. Accuracy is the primary parameter for evaluating classification models. The best accuracy of 100% indicates that all the predictions are correct. It is mathematically calculated as the number of correct predictions by the total number of predictions.

IV. RESULT AND DISCUSSION

The results of this system are stated in two sections. The first section shows the pest prediction result, and the second section shows the Recommendation result. For prediction, user have to provide an image for testing and then click on prediction to get pest details and solution to it.

 DOI : 10.35291/2454-9150.2021.0037

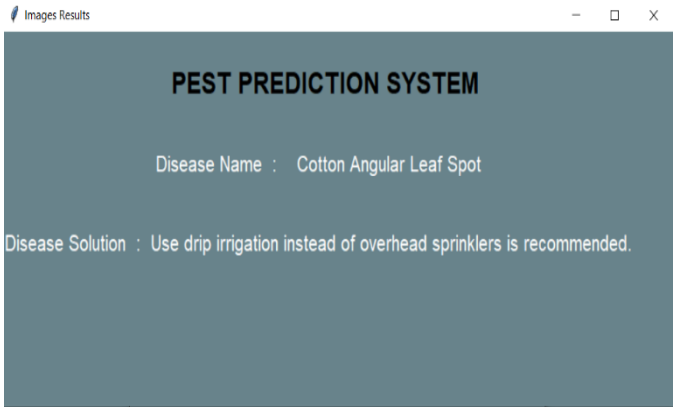

Fig.4.1. Predicted Output

Requirement for Prediction System:

- The image provided must be clear enough to extract its features.
- The infected area on the crop must be in focus to evaluate accurately.
- The image should not be taken from a longer distance.
- Single image should be provided at a time, Collaged images will not give proper results.

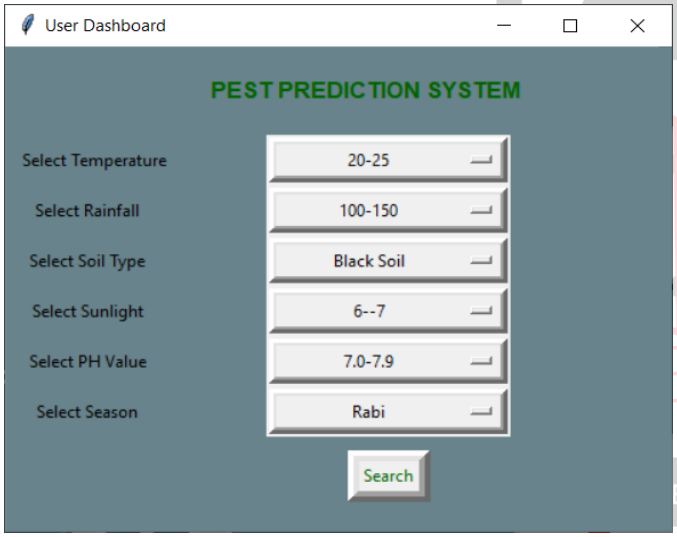

Fig.4.2. Recommendation Parameters

For recommendation, users have to select all mandatory fields such as temperature, rainfall, soil type, sunlight, pH value, and season of cultivation. The recommended crop name will appear in the pop-up window.

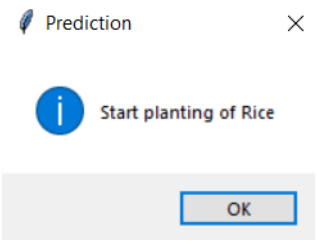

Fig.4.3. Recommended Output

In fig. 4.2, All mandatory parameters are selected, and corresponding to the inputted fields the system generates output. Here it starts planting of rice as seen in fig.4.3.

Requirement for Recommendation System:

- All mandatory fields must be selected.
- Select the options in range available.
- If the specification value is not enlisted in the system, please select the range in which it follows.

V. CONCLUSIONS

This project focused on how Raw Data in the field and past data set were used to predict the pattern of plant diseases and pests using the Naive Bayes algorithm. Agriculture needs a continuous and sustainable increase in productivity and efficiency in agricultural production. While resources like water, energy, fertilizers, etc. need to be used carefully and efficiently to protect and sustain the environment and the soil quality of the arable land. The complexity is boosted by other short-term events which are difficult to predict, such as epidemics, financial crisis, or price volatility for agricultural raw materials and products. Weather plays a vital role in the development or generation of various pests. Pest infections and disease attacks cause major damage and injury to the crops. The transmission of the bacteria or virus to the crops and due to which there is a downfall of the production of crops at the time of harvest. This results in a major loss to the Farmer community. To overcome this problem, a weather-based forewarning pest prediction model is proposed. The proposed model uses the Multiple Regression model and Generalized Linear model techniques.

VI. FUTURE SCOPE

Further, we are planning to transform the project from prototype to a complete end-user product. This can be done using the TensorFlow library function in Python IDE with high processors (recommended using NVIDIA). The model would be accurately predicting disease/pest attacks along with giving preventive solutions. A larger set of data would be provided for the training network. The whole algorithm would be developed using TensorFlow for better processing. OpenCV is used for Image analytics similar to the Image Processing Toolbox in MATLAB. Therefore, the farmer has to just take a picture of the infected area, upload it and the back-end processing will do predict/detect analysis, and give corrective measures for preventing the yield from infection.

REFERENCES

[1] Sun Fan, Chen Hanrong, *"BP network prediction model for occurrence trend of rice gall midge"* [J], Computer Simulation, 2000, 17 (3): 61-63.

[2] Chen Guanhao, Zhang Yiu-chung, Yang Guofeng, "*Prediction model of occurrence degree of main damage generation of late rice gall midge"* [J], Journal of Biosafety, 2006, 15 (2): 137-139.

 DOI : 10.35291/2454-9150.2021.0037

[3] Wang Xuan, Lu Jiake, Xie Deti, et al. *"Application of BP neural network integrated with GASA hybrid learning strategy in rice pest prediction"* [J], China Agronomy Bulletin, 2007, 23 (4): 404-409

[4] Xing Zhenxiang, *"Improvement of genetic algorithm and BP Model and its Application in Water Resources Engineering"* [D], Northeast Agricultural University, 2004.

[5] Zhang Lei, Huo Zhiguo, Wang Li, et al. *"Effects of Climate change on crop pests in China"* [J], Journal of Ecology.

[6] Wu Congcong, "*Application Research of pest prediction based on rough set and artificial neural network"* [D], Jilin University, 2011.

[7] Gao Ning, *"Prediction and prediction of crop insect condition based on BP neural network and its MATLAB realization"* [D], Anhui Agricultural University, 2003.

[8] Xie Xiao, *"Research on Railway passenger Volume Prediction based on genetic algorithm and BP Neural Network"* [D], Southwest Jiaotong University, 2010.

[9] Sheep Wenqi, "*Research on Regional Highway Traffic forecasting based on genetic algorithm and BP Neural Network"* [D], Southwest Jiaotong University, 2015.

[10] Cheng Yugui, Dawn, Lin Mingyu, "*Middle and long-term power load forecasting and analysis based on genetic algorithm and BP neural network"* [J], Computer applications, 2010, 30 (1): 224-226.

[11] Dr. P. Chandra Shekara, Dr. N. Balasubramani, Dr. Rajeev Sharma, Dr. Chitra Shukla, Dr. Ajit Kumar, Bakul C. Chaudhary, "*Farmer's Handbook on Basic Agriculture*", German Federal Ministry for Economic Cooperation and Development (BMZ), Second Edition: August 2016.

[12] Season and Crop Report TamilNadu 2009-10, Department of Economics and Statistics, Chennai - 600 006

 DOI : 10.35291/2454-9150.2021.0037