Review paper on Prediction of Crop Disease Using IoT and Machine Learning

Ms. Supriya S. Shinde, *SYMTech, MIT Academy of Engineering Pune*
Mrs. Mayura Kulkarni, *Sr.Asst.Prof., MIT Academy of Engineering Pune*

***Abstract*—Environmental parameters like humidity, temperature, rainfall, wind flow, light intensity, soil pH are main factors for precision agriculture. Fluctuations in weather parameters like humidity, temperature and so on along with the inappropriate management result into a decrease in crop productivity. Therefore disease prediction is more important to beat these problems. The real-time update will alert the farmer by indicating which crop is in trouble, so the expenses on insecticides, pesticides will reduce and overall economic condition of farmers will improve. The proposed system gives more emphasis to predict diseases of the crop with the use of the Internet of Things and machine learning algorithms. Different sensors collect the real-time data of environmental parameters like temperature, humidity, rainfall, light intensity. Utilizing these data, crop diseases are predicted using machine learning algorithms. Such predictions would warn the farmers about crop diseases through text message or web browser. This work can be extended in the future to help farmers in other ways like which fertilizer can be used to overcome this disease problem.**

***Index Terms*—Agriculture, Internet of Things, Machine Learning, Prediction algorithms, Sensors.**

I. INTRODUCTION

Agriculture is one of the main economic activities of India. It contributes amply to the gross national income. Nowadays crops face many traits or diseases. Many crops are affected by the attack of insects or pests. Hence crop disease prediction is an important task to the agricultural development in India.

Insecticides or pesticides are not always proved effective because they may be toxic to some kind of birds and animals. It also damages the natural animal food web and also food chains. Crop disease results in subjectiveness and considerably low throughput.

In recent years, there has been a large-scale infection area of crop diseases and insect pest, which caused heavy economic losses to the farmers. Crop production losses due to pests and diseases are quite substantial, particularly in the Indian weather semi-arid conditions. Weather plays an extremely important role in agricultural production. Most of the crops are more common in the weather based frangible agriculture systems.

Internets of Things (IoT) data services are usually defined to be available to devices or users on request at any place and at any time. Using the Internet of Things you can make anything like a computer and can connect it to the internet. Trust, reliability, quality, latency, availability, and continuity are the key parameters that have effects on an easy approach to IoT technology. There are some cons of the Internet of Things technology such as privacy, security, trust. The wireless sensor network is one of the component parts of the IoT. There are many sensors which are simple in design mechanism and easy to use their functionality. Sensor technology can be easily integrated with the IoT. To measure the weather parameters different sensors can be used to give accurate readings [1].

In the current scenario, agricultural data are being harvested along with the cultivation of crops and then collected and stored in databases. As the volume of the data stored increases, the gap present between the amount of the data collected and the amount of the data analyzed for processing increases. Data of the farms, crops and other related data are days by day increasing. These data should be used for optimization. Such data may be used in fruitful decision making if proper machine learning algorithms are applied.

Machine learning allows to draw the most important properties from such a vast data collection and to uncover patterns which are until unknown and the hidden relationships present in the data that may be related to present agricultural problems. A prediction algorithm uses the information from past experiences to predict future events. Prediction systems are based on a hypothesis about the pathogen's interface with the host, environment and the disease forming a triangle. The main objective is to accurately predict when the above three factors - host, environment, and pathogen - all are observed. According to that dataset, the whole model is trained.

Prediction system for crop diseases using IoT and machine learning comprises four modules. The first module contains the data acquisition using a wireless sensor network which collects the data like temperature, humidity using wireless sensor network. Different microprocessors like Raspberry pi, Arduino, ESP8266 can be used as an efficient gateway in a wireless sensor network. The second module is about the cloud storage which stores the collected data to the cloud so that data can be accessed from anywhere. The third module contains the machine learning prediction algorithm which predicts the diseases of crops using the training dataset. The last module contains the notification system in which the alert message is sent to the farmer through a text message.

This work is supported in part by the Vigyan Ashram Pabal.

Ms. Supriya Shinde, SYMTech, is with the Department of Computer Engineering, MIT Academy of Engineering Alandi, Pune, MH 412105 INDIA (e-mail: supriyashinde20@ gmail.com).

Mrs. Mayura Kulkarni, Sr.Asst.Prof., is with the Department of Computer Engineering, MIT Academy of Engineering Alandi, Pune, MH 412105 INDIA (e-mail: mukinikar@comp.maepune.ac.in).

In the remaining article, the related work is described which contains the literature review. System model gives the idea of the proposed system. The conclusion is written at the end of the paper.

II. Related Work

In this section, some of the methods related to this work are discussed. These methods are explained as follows.

A. Forecasting Wheat Powdery Mildew by Integrating Remotely Sensed Observations and Meteorological Data at a Regional Scale

This paper uses weather data and observations taken by the sensors jointly for predicting the probability of Powdery Mildew (PM) disease. They have taken precipitation, temperature, humidity, the solar radiation as meteorological factors. They also took remotely sensed data, the reflectance of red band and its temporal dynamics.

This work is carried out at a regional scale. Forecasting model is constructed for calculating the probability of disease occurrence. Different types of data were taken into consideration in forecasting the disease. First, meteorological data is the weather data like temperature, humidity taken from weather stations. Second is the remotely sensed data from the sensors. The third is the survey data of the corresponding field in relation to the PM disease. Dataset taken for this work contains the data collected from 2010 to 2012.

The steps involved in their work can be explained as follows. They extract the remotely sensed features. These features are identified and according to that input variables are selected. Disease forecasting model is constructed using Logistic Regression on the selected input variables. Using these steps wheat PM disease occurrence probability is calculated.

Findings of the paper state that the combination of meteorological and remotely sensed data is a simple, easy and efficient way to predict the disease occurrence. Image processing techniques and Machine learning algorithms easily integrated for PM disease forecasting model [2].

B. Automatic Detection of Disease Powdery Mildew and Tomato Spotted Wilt Virus in Greenhouse

This work is carried out on greenhouse bell peppers. They developed an automatic disease detection system. The system is aimed to detect Powdery mildew (PM) disease occurred on bell peppers and also to detect another common disease occurred on bell peppers named Tomato spotted wilt virus (TSWV) disease. They have used two detection algorithms namely the coefficient of variation algorithm (CV) and principal component analysis algorithm (PCA). These algorithms help in disease control, reduce pesticide application and increase yield.

For disease detection, they used a sensory apparatus for sensing the disease affecting parameters. The sensory apparatus consists of an RGB camera and a laser sensor (DT35, SICK). They used a dataset of 36 plants. These plants were divided into two groups. The first part contains 24 plants and the second part contains 12 plants for both disease detections.

The accuracy of the CV-based algorithm is 84%-87% and accuracy of PCA-based Algorithm is 90%. CV-based algorithms are better than PCA-based algorithms in changing lighting conditions. Therefore the combination of these two algorithms may lead to higher disease detection accuracy.

This paper states that use machine learning algorithms in the IoT for better crop disease prediction [3].

C. Detection of Leaf Rust Disease Utilizing Hyperspectral Measurement with Regression Techniques in Machine Learning

In this work, they consider the symptoms of the different wheat diseases. They measured both, the scale of infected leaves and also the scale of noninfected leaves in these disease symptoms.

They collected all data about the wheat leaves. They take healthy and inoculated wheat leaves and created a spectral dataset from that data. The sequence of RGB digital pictures was gathered from disease infected leaves. Three algorithms are studied using these datasets. First is the Gaussian Process Regression (GPR), the second one is Support Vector Regression (SVR), and third is Partial Least Square Regression (PLSR).

They used multivariate techniques such as Support Vector Regression, Gaussian Process. The main focus of this paper is to use the machine learning algorithms for detection of leaf rust disease. The size of samples used for training and the impact of the symptoms of the disease are also considered during this study. In the last, performances of all algorithms are compared with each other.

Findings of the paper show an investigation in techniques of machine learning regression for the purpose of detecting the leaf rust disease with the use of hyperspectral measurement. GPR's performance shows higher accuracy when the training dataset is smaller. This paper also states that the regression technique in machine learning minimizes the hurdles in forecasting leaf rust disease because of minor changes [4].

D. Combine Approach of Machine Learning and Wireless Sensor Network for Parameter Assessment in context with Gas Source

This paper is about the detection and estimation of gas source parameters. This work uses machine learning techniques very efficiently in a wireless sensor network. It uses different parameters like gas release location, release rate, the concentration of gases, etc. This study is done in both the cases, parameter estimation for single source and also for parameter estimation for more than one source.

They carried out their work in the 5000 meters squared area. This area is split into two same size clusters along with its separate cluster head. Due to separate cluster head, processing of information is carried out at the regional level. Sensors are placed dispersedly in the region such that density should be $0.05 \text{sensor}/m^2$.

The methods implemented in this paper are Advection-diffusion model which gives a description of detection and Kernel-based regression models for estimation phase. They propose a new clusterized framework in wireless sensor network for the parameter assessment in context with the gas source. The methods were manifested to be extremely effective in modeling nonlinear hassles. In this way, the proposed approach demonstrated that it gives accurate results. This approach also deals with the noise occurred in the data. This technique gives an efficient result even with the multiple gas sources.

Failures can be vigorously handled by this technique. This approach requires comparatively less energy. Wireless sensor network can be easily merged with this technique. The paper suggests that machine learning algorithm in a wireless sensor network are for better prediction in different real-time applications [5].

E. Event Forecasting with Bayesian Technique using Preliminary Stage Longitudinal Information

The main focus of this paper is to forecast the events which are going to happen in upcoming time point. It uses the details regarding a little count of event occurrences at the starting phase of a longitudinal study. The paper proposes an Early Stage Prediction (ESP) technique for forecasting system.

First, they develop an advanced practice Kaplan-Meier estimator to manage examined data. For prior probability, they fit time-to-event information using Weibull and Log-logistic distributions. Then they extend the Naïve Bayes algorithm and generate new algorithm Early Stage Prediction – Naïve Bayes. In a similar way, they extend the Tree-Augmented Naive Bayes algorithm and generate new algorithm Early Stage Prediction – Tree-Augmented Naive Bayes. They also extend the Bayesian Network algorithm and generate new algorithm Early Stage Prediction – Bayesian Network. These newly developed algorithms can effectively predict event occurrence using a little number of incidences happened during the starting stages of a work. The probability of event occurrence is forecasted by proposing an Early Stage Prediction (ESP) framework.

This paper proves that the Event Forecasting with Bayesian Technique using Preliminary Stage Longitudinal Information is proved efficient [6].

F. Disease Forecasting for Healthcare Sector by Machine Learning in context with Big Data

This work describes the restructured machine learning algorithm for better forecasting of recurring epidemic diseases in communities which are disease susceptible. This work uses the realistic medical data of Central China collected from 2013 to 2015.

The experiment with the modified prediction models is carried out over that data. Some of the incomplete entries in the data are treated with a latent factor technique. The epidemic disease cerebral infraction is considered for their work. Structured and unstructured medical data both are utilized in the disease forecasting. Data consists of many elements like sufferers primary data, doctor's interrogation records and diagnosis.

In this way, they introduce the data imputation for disease forecasting. They present two approaches multimodal and unimodal. They develop Convolutional Neural Network - based Multimodal Disease Risk Prediction (CNNMDRP) algorithm for multimodal disease forecasting. In a similar way, Convolutional Neural Network - based Unimodal Disease Risk Prediction (CNN-UDRP) algorithm is developed for unimodal disease forecasting approach.

Thus, the paper findings state that the performance of CNN-MDRP is better than CNNUDRP. Unimodal approach and multimodal approach both algorithms shows the effective results with a little difference. Convolutional Neural Network works efficiently on big data in context with the disease forecasting [7].

III. SYSTEM MODEL

The actual problem statement of this work is to predict the diseases of crops using Support Vector Regression and Adaboost algorithm by taking the real-time input of wireless sensor network and developing a notification system for farmers to give them real-time updates and alerts about the unusual situation about their farms.

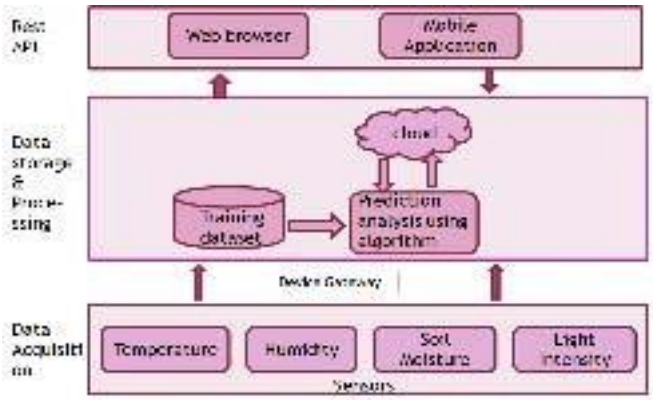

Fig. 1. The Architecture of the system

The figure shows the architecture of the system. In that input data is taken from the wireless sensor network, that data include humidity, temperature, soil ph, light intensity, etc. Different sensors are used to collect the real-time data like the DHT22 sensor will collect the temperature and humidity, Soil moisture sensor will collect the water content in the soil. All collected data is stored on the cloud like Amazon Web Service so that data can be accessed from anywhere. Then this data is passed over the prediction algorithm and by using training data prediction is done. And according to that predicted real-time updates about the crops will be sent to the farmers through a text message or web browser, so that they can take precautions and will able to save their crops.

Here this work uses an algorithm which is a combination of two algorithms namely, Support vector regression and Adaboost. Both Support vector regression and Adaboost algorithm are supervised learning algorithms. So it can be used for classification as well as regression. Support vector

regression is same as Support vector machine but it used for regression. Support vector regression is memory efficient algorithm. Selecting hyperplane is an important issue in Support vector regression and also it requires more training time. The algorithm can effectively handle high dimensional data. The second one is Adaboost, 'Boosting' means a group of algorithms which help to convert weak learners to strong learners for better prediction. For making a strong learner, it combines the prediction result of each weak learner by taking their weighted average or higher vote. It is time efficient and improves the accuracy of time.

This system will give more accurate crop disease predictions and will be easy to access the farmers. So that farmers can improve their crop production. The real-time update will alert the farmer by indicating which crop is in trouble, so the expenses on insecticides, pesticides will reduce and hence the economic condition of farmers will improve.

IV. Conclusion

This review paper gives information about Machine Learning and IoT implementations used for crop disease predictions. The listed papers describe different Machine Learning and IoT techniques for Precision agriculture. But, the more reliable and cheap system has not been developed.

There is no such method is developed to predict diseases of different crops which is easy to implement, cheap and simple to use. In this way, the system model suggests that 'prediction of crop diseases using IoT and Machine learning' will implement efficiently. IoT network will help in real-time data collection. Machine learning algorithms give more precise predictions. The proposed work can also be extend to act as a guide for farmers like, which fertilizer can be used to overcome the disease problem and which crop is beneficial to sow in these weather conditions.

Acknowledgment

The authors would like to acknowledge the advice of the professors and those who previously worked on this area. Authors would also like to thanks the reviewers for their suggestions to improve the quality of the paper.

References

[1] P. Barnaghi, A. Sheth. (2016, Nov-Dec). On Searching the Internet of Things: Requirements and Challenges. IEEE Intelligent Systems [Online]. 31(6), pp. 71-75. Available:https://ieeexplore.ieee.org/document/7742299/

[2] J. Zhang, R. Pu, L. Yuan, W. Huang, C. Nie and G. Yang. (2014, Nov). Integrating Remotely Sensed and Meteorological Observations to Forecast Wheat Powdery Mildew at a Regional Scale. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing[Online]. 7(11), pp. 4328-4339. Available:https://ieeexplore.ieee.org/document/6814816/

[3] N. Schor, A. Bechar, T. Ignat, A. Dombrovsky, Y. Elad, and S. Berman. (2016, Jan). Robotic Disease Detection in Greenhouses: Combined Detection of Powdery Mildew and Tomato Spotted Wilt Virus. IEEE Robotics and Automation Letters[Online]. 1(1), pp. 354-360. Available:https://ieeexplore.ieee.org/document/7383246/

[4] D. Ashourloo, H. Aghighi, A. Matkan, M. Mobasheri, and A. Rad. (2016, Sept). An Investigation Into Machine Learning Regression Techniques for the Leaf Rust Disease Detection Using Hyperspectral Measurement. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing[Online]. 9(9), pp. 4344-4351. Available:https://ieeexplore.ieee.org/document/7533506/

[5] S. Mahfouz, F. Mourad-Chehade, P. Honeine, J. Farah, and H. Snoussi. (2016, July). Gas Source Parameter Estimation Using Machine Learning in WSNs. IEEE Sensors Journal[Online]. 16(14), pp. 5795-5804. Available:https://ieeexplore.ieee.org/document/7470596

[6] M. Fard, P. Wang, S. Chawla, and C. Reddy. (2016, Dec). A Bayesian Perspective on Early Stage Event Prediction in Longitudinal Data. IEEE Transactions on Knowledge and Data Engineering[Online]. 28(12), pp. 3126-3139. Available:https://ieeexplore.ieee.org/document/7564399/

[7] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang. (2021, Apr). Disease Prediction by Machine Learning over Big Data from Healthcare Communities. IEEE Access[Online]. 5, pp. 8869-8879. Available:https://ieeexplore.ieee.org/document/79

[8]

[9] 12315/.