Arecanut Crop Disease Prediction using IoT and Machine Learning

*SharathKumar KR[1], Mohan K[2], Nirisha[3]

1, 3(BE, Department of Computer Science & Engineering, Srinivas Institute of Technology, Mangaluru, India)
2(Assistant Professor, Department of Computer Science & Engineering, Srinivas Institute of Technology, Mangaluru, India)
Corresponding Author: sharathkr606@gmail.com

Abstract: *A prevailing recession in the agricultural goods sector is evident from the present scarcity and lack of food supplies. A major reason for this scarcity is the inherent growth of diseases in essential crops. A major development is thus required in this field for avoiding these problems in the future. This development is intended to simplify the management tasks of different roles in agricultural industries. A proper intimation of the importance of disease prediction and environmental factors must be done to the less aware farmers. To address these challenges, we have proposed a disease prediction system that takes into consideration temperature (°C), humidity(%), rainfall(cm), wind flow(m/s) and soil moisture (%) around the region of crop and developed a model to predict the occurrence of disease. This system will provide information prior to the occurrence of disease by analyzing different relationships among environmental factors.*

Keywords: *Arecanut, Disease Prediction, Crop Diseases, Difference Algorithm, IoT, Koleroga.*

1. Introduction

Agriculture is the most important aspect in India. But many problems arise in agriculture which includes the diseases that occur in plants which decrease the production of crops. This work is based around the Areca crops. It is cultivated largely around the southern part of India and is commercially available in dried, cured and fresh forms. It is also known as betel nut, as it is commonly used for mastication with betel leaves. It is economically important and is consumed by hundreds of millions of people worldwide. India is one of the largest producers of arecanut in the world and it is mainly concentrated in the states of Karnataka, Kerala and few eastern states.

As the spread increases, the effects of different environmental conditions vary widely and results in few diseases being attacked. The diseases like 'Koleroga', Yellow Leaf Disease, Foot Rot are few of them. One among these is the Mahali/Koleroga/Fruit Rot. The characteristic symptom of Fruit Rot is the rotting and excessive shedding of immature nuts, loss of natural lustre and usually lighter weight and presence of vacuoles in the nuts. The disease is a reason for huge losses of productivity and it is important to determine the spread of these to avoid the possibilities of infections [3].

A proposal is established where the innate relationship between crop disease and surrounding environmental conditions are identified and monitored continuously to predict the possibilities of a disease infection. The collected data can be used to mine further information and provide suitable counter measures to accomplish the aforementioned tasks.

II. Material and Methods

Prediction of these diseases is challenging and is considered to be expensive. Machine learning is used for performing several tasks in computers without the need of human help. We can use algorithms like Regression, Random Forest algorithm (RF), Gradient Descent (GD), or Support Vector Machines (SVM) for predicting and

quantifying the disease impact [4]. In this project work we have used the Support Vector Machine Regression Algorithm (SVMR) and the Random Forest Classifier (RFC) algorithm to perform the prediction task. Environmental condition sensors like DHT-11 and soil humidity sensor is used to collect environmental data values that include temperature, rainfall, humidity, soil, and speed of the wind. These collected data should be used by the farmers when they reckon the need to use technology to simplify their tasks and aspire to accomplish better outcomes than intended.

Study Design: Environmental data collection favorable for arecanut disease prediction model

Study Location: The data is collected from areca fields in the region of Vittal, Dakshin Kannada and the region surrounding this place.

Study Duration: January 2018 to April 2019 along with historic weather data.

Study Disease: Mahali/Fruit Rot/Koleroga

Conditions suitable for Disease:
1. Frequent splashes of rain
2. Usually 15-20 days after the onset of monsoon is considered prime spot
3. Humidity higher than 90%
4. Low temperatures in the range of 20 °C to 23 °C
5. Intermittent sunshine and rain hours.

1.1 Procedure methodology

We have considered two forms of data, one from a thorough research on disease conditions and another a real-time data collected from sensors. The system is trained with data values collected from historical research modeled as a dataset. Manual data values are monitored using sensors and recorded as tables before modelling it into a dataset. In order to get the data values, we have used DHT-11 and soil humidity sensors. DHT-11 sensor is used to get the temperature and humidity values and soil humidity sensor is used to get the soil moisture value. We have considered seven days difference between these data values to obtain a relation among the consecutive days and mine a pattern or trend inherent in these. These values are compared with the training data values, in order to check the similarity in trend among these values. We have used regression algorithms to find the same. The above made observations are scored and we can get the best results. [2][5]

Below mentioned algorithm is used to mine a pattern from the data to predict the disease. Raw data is fed to the algorithm and the actual difference between the consecutive values of the same are calculated. This difference is then used to make observations and predict occurrence of disease.

Algorithm 1: Basic steps describing the proposed algorithm.

---

1. Obtain the necessary range of data values as a list.
2. Traverse through the list of data.
3. Calculate the difference between 2 consecutive values.
4. Append the difference to a new list.
5. Return the set of differences as a list.
6. Repeat steps 2 through 5 for both train and test data.
7. Identify and compare the pattern in the two obtained lists.
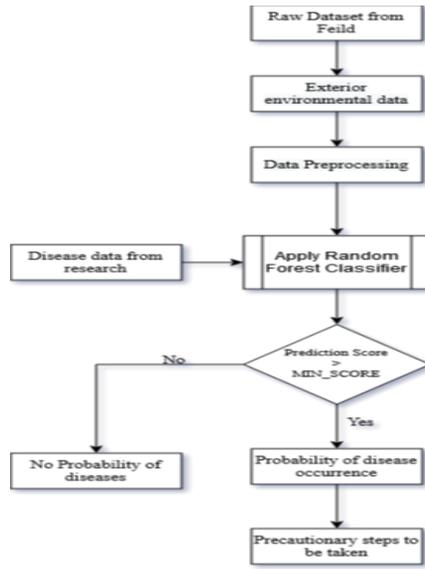8. Specify the results based on the observations.

Figure 1: A Flowchart for Crop Disease Prediction using IoT and Machine Learning

A model representation of the system is shown in Figure 1. The flowchart shows the procedure of work to be done to obtain the data and predict the values.

Here, the sensors installed in the field are used to collect the data like temperature, humidity, rainfall, wind speed etc. This raw data needs to be processed and converted into a tabular format to develop the input dataset for the model. In addition to the data collected from the fields, the historical data of the region was also collected from the respective sources. The data was classified based on the identified disease conditions to extract the actual conditions of disease. Table 1 shows the data obtained from research centers at CPCRI, Vittal that provides the environmental characteristics for a specified amount of time in a year. This constitutes the environmental features like maximum and minimum temperature, humidity, wind, sunshine, evaporation and rainfall. These data have been grouped categorically to represent a monthly view of the conditions. The historic data collected from the region is then analyzed and the time period of disease occurrence is identified to get the probable disease conditions.

Table 1. An overview of the Historic data of environmental conditions of the fields

Month	Temperature		Humidity		Wind	Sunshine	Evaporation	Rainfall	Rainy Days
	Max[1]	Min[2]	FN[3]	AN[4]					
Jan-18	33.7	16.3	94.5	39.6	1.9	6.0	3.3	0.0	0
Feb-18	35.3	17.5	92.8	34.0	2.5	6.7	4.1	0.0	0
Mar-18	35.1	20.1	92.8	48.1	2.8	4.8	4.4	79.4	2
Apr-18	35.3	22.0	92.5	57.2	3.1	5.1	4.5	88.3	3
May-18	33.6	22.7	93.9	62.6	3.3	3.2	3.7	334.1	9
Jun-18	28.7	22.3	97.7	87.8	2.8	1.0	1.7	1216.3	24
Jul-18	28.3	22.0	97.7	86.0	3.8	1.1	1.5	1149.4	28
Aug-18	28.3	20.9	97.7	84.5	4.0	1.5	1.7	858.5	29
Sep-18	32.0	21.5	95.1	63.7	2.5	6.5	3.4	51.8	5
Oct-18	33.4	21.9	93.8	59.6	2.0	6.4	3.1	258.4	10

---

[1] Maximum
[2] Minimum
[3] Fore Noon
[4] After Noon

Nov-18	34.3	21.4	92.0	49.5	2.1	6.8	3.5	84.6	2
Dec-18	33.4	20.4	94.2	50.2	1.8	6.0	2.9	0.0	0

III.   Result

In this section, we present the Machine Learning algorithms implemented and the constraints set on the data for achieving the desired results. The data collected from research work along with the generated dataset was used to train the Random Forest Classifier model to define the scores for the observations. The Random forest uses hundreds of decision trees to predict the class of an instance, sometimes even thousands. Now every tree in this forest will predict a class for some random set of the input data. This results in an overall prediction with low variance. The temperature distribution of the data collected is given below in Figure 2. This shows that the temperature range lies within the probable range of 20 °C to 23 °C as obtained from the research work. Thus, the further processing was carried out to determine any patterns in the corresponding data.
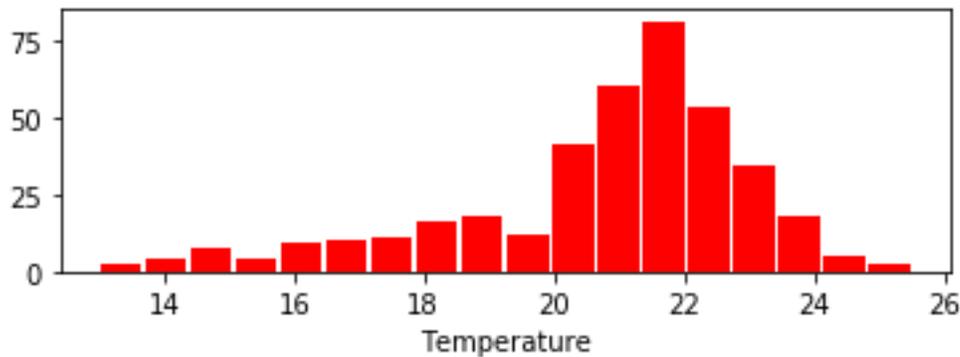

Figure 2: A histogram representing the distribution of temperature in the collected data favorable for disease

Later, a Support Vector Machine(SVM) classifier is also applied to the data that was obtained after the application of *Algorithm 1* andprocessed accordingly to check if any common pattern or trend is found in the data[1]. A similar trend identification is performed over the dataset obtained from the field as well. If both the trends are found to be identical, then there are possibilities of the disease infection. This is then notified to the concerned person to take necessary precautionary steps. If not identical, then there is no possibility for the disease to occur. A sample output data for a test classification is shown in Table 2 below.

Table 2.A classification report on Support Vector Machine (SVM) Classifier on the model over the training data

	Precision	Recall	F1-Score	Support
Micro avg	0.32	0.32	0.32	50
Macro avg	0.54	0.34	0.32	50
Weighted avg	0.56	0.32	0.32	50

To have a better understanding of the result prediction, we applied the Random forest classifier to the data. The prediction is achieved by averaging all the outputs of the decision trees. Here we have used 300 decision trees in the Random Forest and trained a model to implement the prediction system. The scores were in a range from 1-10 and as per the data collected, MIN_SCORE was considered as 4 for the classification model. A score greater than 8 was set as the threshold for high probability in infection. The prediction was then carried out over the real-time data which was gathered using sensors from the fields. The outputs were recorded and prediction was intimated to the user. The results obtained indicated greatly that the model was able to predict the pattern and classify the new inputs accordingly.

IV. Discussion

Table3 shows a set of identified environmental conditions that is scored manually based on inputs from the farmer and related research work.As described by Table 3, the real-time data entered by the user is fed into the model and a score is depicted as explained below.

Table 3. A sample data to train the model for disease prediction.

Days	Temperature	Humidity(%)	Wind(m/s)	Soil Moisture	Rainfall(cm)	Sunshine	Result	Prediction
1	21	88.9	7.1	93.4	135	0	8	highly yes
2	22	89.1	7	94.8	149	0.3	8	highly yes
3	22.3	91.1	6.95	94.9	150	1	8	highly yes
4	22.8	90	6.9	93	164	0.6	9	highly yes
5	21.6	94	7.1	91	165	0.4	9	highly yes

The score is then used to classify the input into either a disease favourable or not so favourable condition. Different patterns are observed in the environmental conditions, but most of the time the variations are the reason to consider working on, as it leads to adverse changes in the crops being cultivated. The approach followed here mainly lacks the strong credibility in the data availability part. It is very difficult to collect the historic data of every place and find a pattern in the data. We have defined a considerable approach to mine the pattern in occurrence of the Mahali/Koleroga/Fruit Rot disease in Arecanut crops and the obtained results are adequate to identify and warn the farmers of a possibility of infection spread in the region. Under certain assumptions, the project work construed a methodology to predict the disease spread. However, it is a limitation here that the research work is restricted to a particular region and time period. To overcome this set back, there is a necessity to obtain widespread research data for the arecanut plantation and strengthen the model.

There has not been any conflict of interest to the idea described in the paper and it stands firm in achieving the goal of predicting the Mahali/Koleroga/Fruit Rot disease in Arecanut crops.

V. Conclusion

In this paper, we have exhibited a system to predict if the arecanut plant is susceptible to disease infection or not by comparing the data values which we got by using some IoT sensors like DHT-11 and soil humidity with data values that caused diseases in plants that were found by doing historical research. Comparison between these data values are done by using a regression algorithm to find similarities in the pattern of conditions. A range is determined for the scoring of results and a value greater than 8 (on a scale of 1 to 10) indicates that there is a high possibility of disease infection in the region. The output obtained is manually confirmed with experience. This helps to evaluate the model efficiently. The farmers are warned about the possibility of disease so that they can take necessary precautions. As a scope for the future work on this model, it can be extended to multiple crops and for multiple diseases based on thorough research over various crops and conditions.

VI. Acknowledgement

References

1. Yun Hwan Kima, Seong Joon Yooa, *, Yeong Hyeon Gua, Jin Hee Limb, Dongil Hana, Sung WookBaik, "Crop Pests Prediction Method using Regression and Machine Learning Technology: Survey*", International Conference on Future Software Engineering and Multimedia Engineering, 2013.*
2. Dengwei Wang1,2, Tian'en chen1. * and Jing Dong Research of the early Warning Analysis of Crop Diseases and Insect Pests 1 *National Engineering Research Center for Information Technology in Agriculture (NERCITA), Beijing 100097, China, school of Mathematics and Computer Science Ningxia University, Yinchuan, 750021*, China, 2014.

3. HemantKumar Wani, Nilima Ashtankar "An Appropriate Model Predicting Pest/Diseases of Crops Using Machine Learning Algorithms", *International Conference on Advanced Computing and Communication Systems (ICACCS-2017),* Jan 06-07, 2017.
4. Huu Quan Cap, Satoshi Kagiwada, Hiroyuki Uga, Hitoshi Iyatomi "A Deep Learning Approach for on-site Plant Leaf Detection", *International colloquium on Signal Processing and its applications (CSP A 2018)*, 9-10 March 2018.
5. G. Prem Rrishi Kranth, M. Hema Lalitha, Laharika Basava, Anjali Mathur "Plant Disease Prediction Machine Learning Algorithms", *International Journal of Computer Applications (0975-8887), Volume* 182-*No.25,* November-2018.
6. Adithya Khamparia, Gurinder Saini, Deepak Guptha, Ashish Khanna, Shrasti Tiwari, Victor Hugo C. de Albuquerque "Seasonal Crop Disease Prediction and Classification Using Deep Convolutional Encoder Network", *Circuits, Systems and Signal Processing,* Jan 15, 2019
7. R. D. Devi, S. A. Nandhini, R. Hemalatha and S. Radha, "IoT Enabled Efficient Detection and Classification of Plant Diseases for Agricultural Applications," 2019 *International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, Chennai, India, 2019.