iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 12 **Issue:** XI **Month of publication:** November 2024

**DOI:** https://doi.org/10.22214/ijraset.2024.65307

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 12 Issue XI Nov 2024- Available at www.ijraset.com*

# Weather-Wise CropAlert: Predicting and Preventing Plant Diseases

Shraddha Patel[1], Prachi Shrirao[2], Sumit Naik[3], Sumit Raina[4], Aniruddha Sutawane[5], Dnyaneshwar Kanade[6]
[1, 2, 3, 4, 5, 6]*Department of Electronics and Telecommunication Engineering, Vishwakarma Institute of Technology, Pune, 411037, Maharashtra, India*

***Abstract:*** ***This paper provides a system which consists of three components namely, weather parameter input system, plant disease prediction system and disease prevention system. The system will help the farmers to efficiently use the weather data, and to take preventive measures in time and helps establishing a direct relationship between the weather and plant diseases. The system currently provides information for three crops, namely, wheat, cotton, rice. The proposed system used machine learning to identify the patterns and predict the plant disease. Various clustering and classification models were used in the system such as K means, Random Forest and XGboost.***
***Keywords: plant disease prediction, weather parameters, K means, Random Forest, XGboost***

## I. INTRODUCTION

Plant diseases can significantly lower crop production and undermine food security, which is why the agricultural sector has experienced enormous challenges in recent years. One of the main factors influencing the occurrence and spread of plant diseases is the weather. Climate variables such as temperature, humidity, rainfall, and others can create settings that either encourage or inhibit the spread of illness. Therefore, predicting and avoiding illness requires an understanding of the link between weather patterns and plant health. The goal of the project "Plant Disease Prediction and Prevention Based on Weather Conditions" is to forecast plant disease outbreaks and suggest preventive actions by utilising machine learning and advanced data analytics techniques. This project will create predictive models that can determine the probability of illness occurrences under particular weather scenarios by combining historical weather data, real-time climate monitoring, and disease incidence records.

Furthermore, the initiative will concentrate on formulating practical approaches for farmers and other agricultural stakeholders to lessen the effects of these illnesses. According to the anticipated risk of illness, recommendations will be made regarding the best times to plant, when to water, and whether to use pesticides or fungicides. The ultimate goal of this project is to improve agricultural sustainability and productivity by taking a proactive approach to managing plant diseases, protecting crop health and guaranteeing a steady supply of food. All the models that are available today only detects the plant disease based on image datasets and then it provides the deterministic measures on how to improve plant health. There are no model to prevent diseases based on disease predictions. The model we are providing establishes a clear correlation between weather and plant disease so that one can predict the disease that are likely to happen due to the weather conditions and provides the preventive measures.

## II. LITERATURE REVIEW

The nineteenth century saw a small advancement in meteorology as weather prediction shifted from being a guessing game to an actual science. The precision of weather forecasting has been rising ever since. Additionally, there has been an improvement in sensor accuracy [2]. An Arduino, Raspberry Pi, or an ESP wifi module are a few examples of development boards that can be used to connect several sensors [1]. The primary purpose of a weather station is to collect meteorological data from several sensors and store it in databases using various cloud services or APIs. [3].

Similar to this, for the past ten years, there has been a concept to accurately identify plant diseases that are present in the plant through the use of cameras and detection algorithms [7]. There are systems setup at present that anticipate crop production based on weather forecasts and comparative study to determine which algorithm is best suited [4][6]. Certain systems have employed soil analysis and weather forecasts to determine when to plant which crops [5][9].

The algorithms that we are using to predict the plant disease is based on clustering and classification model. The purpose of clustering, a sort of unsupervised learning, is to arrange a collection of objects so that the objects in the same group (or cluster) resemble one another more than the objects in other groups [6]. To find hidden patterns or groupings in data, exploratory data analysis makes extensive use of it.

©IJRASET: All Rights are Reserved 

Classification is done to put a test case in a particular cluster having similar properties [7]. The proposed system was also done by using high spatial resolution in a small town in Korea [8], where the weather forecast was done for a particular farmland. Hourly or daily warning of the predicted disease was made. That system received an accuracy of 0.44 as the leaf wetness event were not taken in consideration. We also referred to another algorithm specially developed for crops grown in India to increase the transparency and interpretability of the machine learning models [10].

## III. METHODOLOGY

### *A. Dataset*

For the proposed system we required two datasets, plant disease dataset and preventive measures dataset. Both are in the form of comma separated values (csv) file. The plant disease dataset includes diseases of 3 different crop variety namely, cotton, wheat and rice and their respective diseases, and their relationship with different weather parameters. Temperature, humidity, precipitation, wind speed are added in the used dataset, one can also use different weather parameters. The dataset included the numerical data for the weather conditions, which was then formulated as a range.

A variety of graphs have been produced to show the dataset so as to make it easier and faster to understand. Patterns, trends, and abnormalities that might not be obvious from raw data become simple to identify. The following graph shows the number of disease present in which range of that particular weather parameter. This graph was plotted to analyse roughly that during which weather conditions is the plant most likely to be affected. This graph accounts for all the diseases irrespective of the crop.

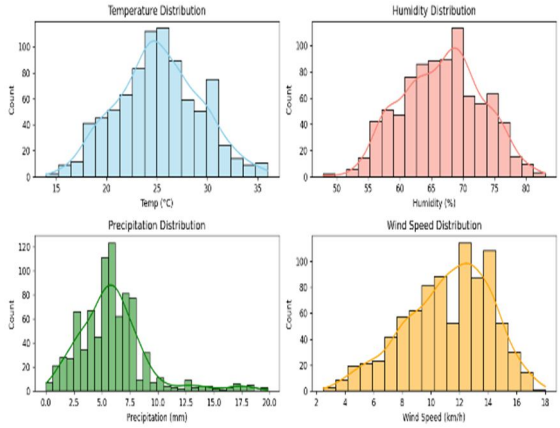



Figure 1. Count of the weather parameter vs weather parameter in the dataset

The following graph is a boxplot which shows the range of the weather condition in which that particular disease is present. The earlier graph only provided with the count of the disease for that range. But the following boxplot provides the specific disease which is present in that weather condition. Also providing the outliers and median value of that range.

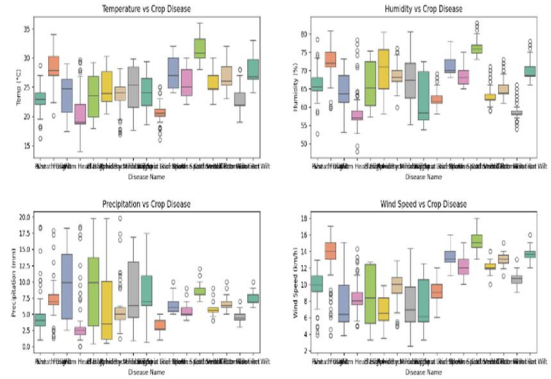



Figure 2. Boxplot showing the range where the disease lies with respect to a particular weather parameter.

The dataset was then standardized to show more accurate results. There were various outliers to be accounted, hence a violin plot was plotted.

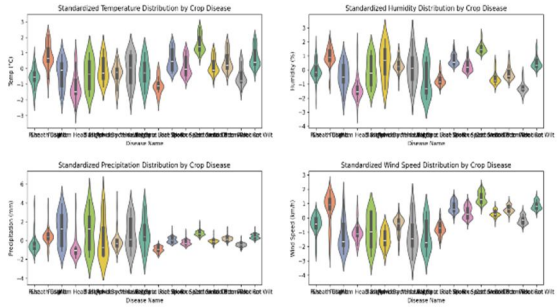
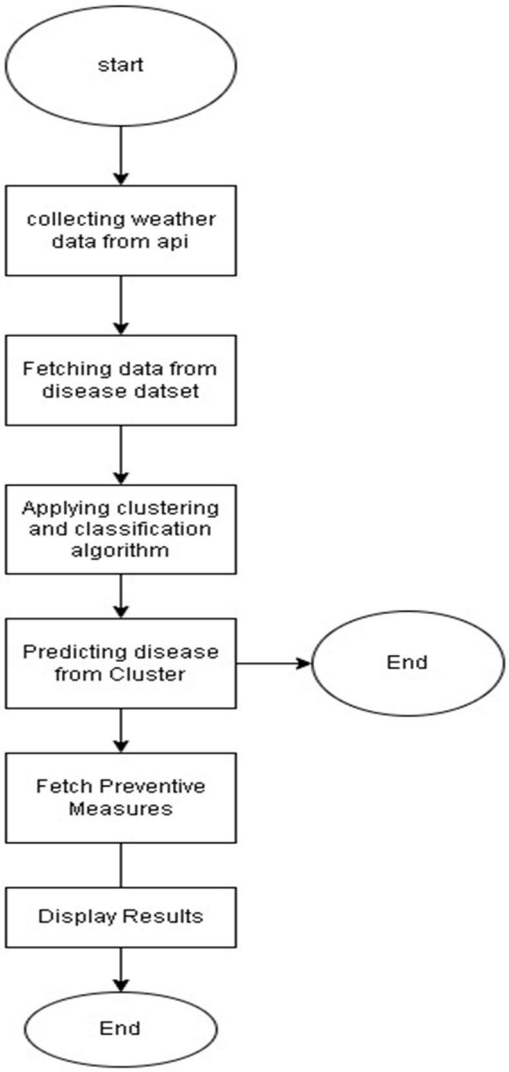



Figure 3. Graph showing standardized result of weather distribution with respect to disease name

*B. Flowchart*




*C. Algorithm*

*1) Clustering*

The data is divided into K clusters, each of which serves as a prototype for a cluster and to which each data point belongs. Until convergence, the method moves the cluster centres, or centroid positions, iteratively, allocating points to the closest centre. Selecting optimum number of clusters was an important parameter too, which was done by Elbow Method or Silhouette Analysis. The following graphs shows different number of clusters for comparison purpose.

Euclidean Distance:

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - c_{jk})^2}$$

Where,

- $x_i$ is the data point
- $c_j$ is the centroid
- $n$ is the number of dimensions(weather parameters)

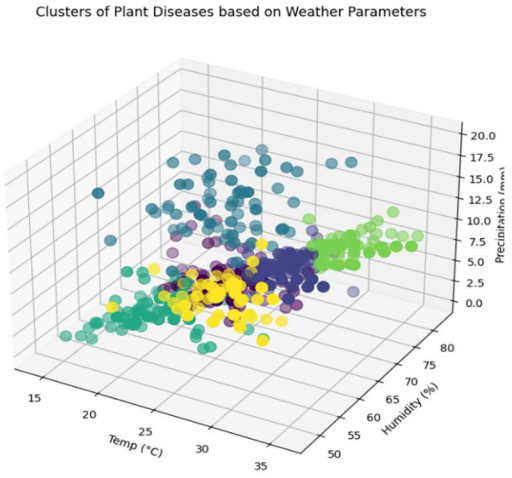



Figure 5. Graph showing six clusters with respect to three weather parameters

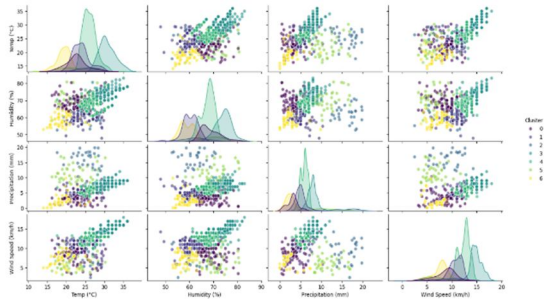


Figure 6. Graphs showing 6 clusters with respect to various parameters

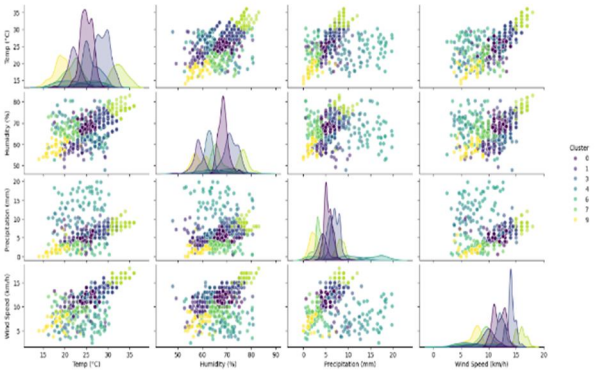


Figure 7. Graph showing ten clusters

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

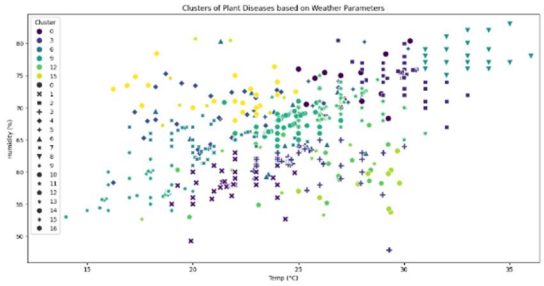



Figure 8. Graph showing clusters formed with respect to unique disease name

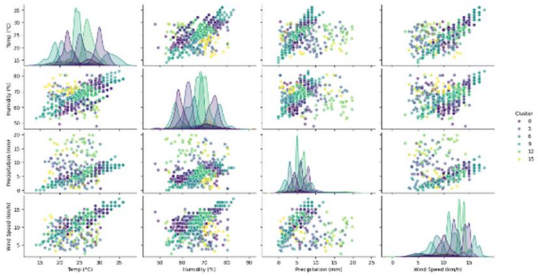


Figure 9. Graph with 16 clusters unique to disease name

Comparing multiple clusters allow us to select optimal cluster identification which helps in understanding the data distribution in the dataset. It also helps in visualizing various other parameters such as intra-cluster distance and inter-cluster distance which facilitates in cluster quality.

To evaluate the clustering quality, the Silhouette Coefficient s(i) was used to measure how similar a point is within its own cluster compared to others. It is calculated as:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Where:

- *a(i)* is the average distance from point i to all other points in the same cluster.
- *b(i)* is the minimum average distance from point i to points in a different cluster.

Furthermore, predicting a new observation's categorical label based on previously seen data with known labels is the aim of classification, a sort of supervised learning. Utilising labelled data to train a model, additional data points are tagged using the model. For this purpose, we have used two algorithms, namely, Xgboost and Randomforest.

*2) XGBoost*

XGBoost is a powerful and efficient gradient boosting solution for issues needing supervised learning. By combining the predictions of multiple weak learners—typically decision trees—the procedure creates a potent predictive model. Being very accurate, XGBoost has a number of benefits, one of which is better prediction performance. It can handle big datasets since it is very efficient and well-optimized. L1 (Lasso) and L2 (Ridge) regularisation techniques are also used into XGBoost to prevent overfitting, and it enables parallel and distributed computing to further improve performance and scalability.

The objective function for XGBoost with regularization controls overfitting:

$$Obj(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Where:

- $l(y_i, \hat{y}_i)$ is the loss function (e.g., logistic loss)
- $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda||w||^2$ is the regularization term

*3) Random Forest*

Comparably, A collaborative learning method called Random Forest creates many decision trees during training and provides the mode of the classes for classification tasks or the mean prediction for regression tasks. This technique reduces overfitting by averaging several decision trees, which results in excellent accuracy. In addition, Random Forest is more resilient to noise and outliers than single decision trees. In addition, it offers important feature estimates, that are useful for choosing features and understanding the model's behaviour. The features which were chosen for this algorithm were all the meteorological parameters.

In decision trees, Gini impurity measures how often a randomly chosen element would be incorrectly labeled:

$$G = 1 - \sum_{i=1}^{n} {p_i}^2$$

Where $p_i$ is the probability of a class being chosen.

For classification models, the confusion matrix is an essential performance monitoring tool. This makes it possible to calculate a number of metrics, including F1 score, recall (sensitivity), accuracy, and precision. The confusion matrix allows for a more thorough understanding of the model's effectiveness, which offers comprehensive insights into the classification performance. It does this by showing not only the overall accuracy but also the precise regions in which the model is incorrect.

To evaluate your classification models (Random Forest, XGBoost), you should use these formulas:

- Accuracy:

$$\frac{TP + TN}{TP + TN + FP + FN}$$

- Precision:

$$\frac{TP}{TP + FP}$$

- Recall:

$$\frac{TP}{TP + FN}$$

- F1 Score:

$$2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

Table 1. Comparison of accuracy in different clusters.

| No. of clusters | Accuracy in random forest | Accuracy in xgboost |
|---|---|---|
| 3 | 0.77 | 0.77 |
| 6 | 0.80 | 0.79 |
| 10 | 0.78 | 0.77 |
| 16 | 0.76 | 0.77 |

*D. Software Implementation*

We made a website which was also integrated with a free online weather API which provided weather conditions from different locations.

Tools used:

- Programming language: Python
- Machine learning libraries: XGboost, Randomforest
- Visualisation tools: numpy, seaborn, matplotlib
- Database: comma separated value file
- Integrated development environment: Python IDLE, Jupyter notebook, google colab
- Weather API

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

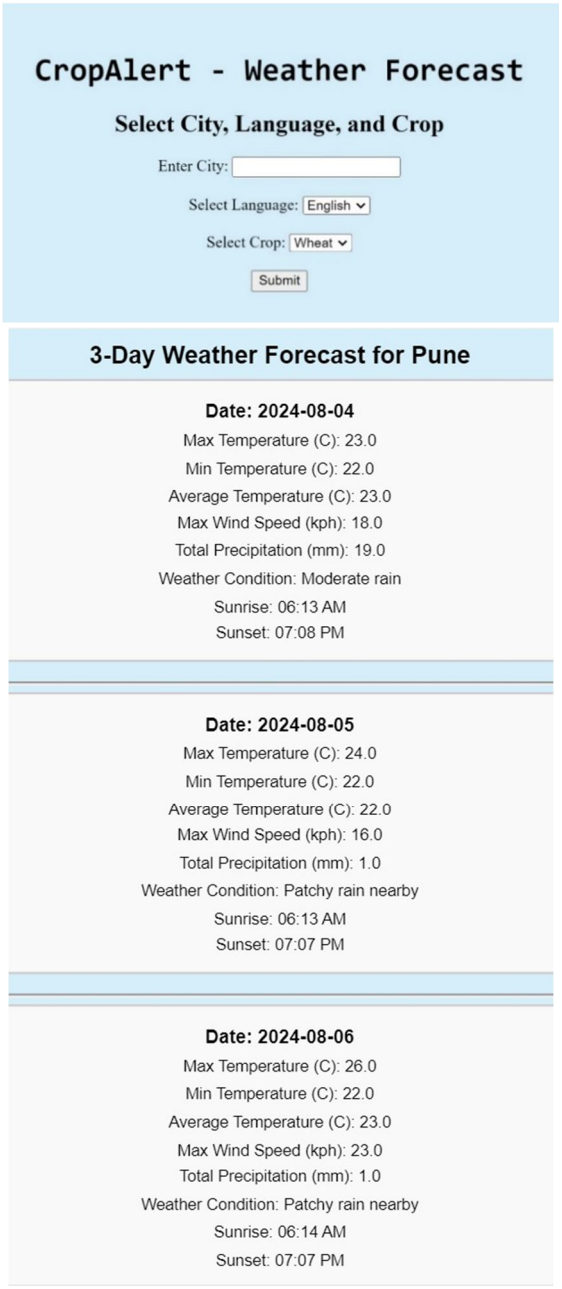


Figure 12. Screenshot of the website created

*E. Hardware Implementation*

Tools used:

- DHT11 sensor
- Soil moisture sensor
- ESP8266
- Arduino IDE

In addition, a hardware prototype was made as part of the project to incorporate meteorological characteristics straight from sensors. For a specific location, like a farm, this will provide precise meteorological parameters. so that more precise forecasts about plant diseases can be produced. Because weather APIs will provide standard characteristics for a broad region, this could result in imprecise forecasts for more localised areas.

The DHT11 sensor will be used for temperature and humidity parameters and the soil moisture sensor is for measuring precipitation, the wind speed is measured by a device called an anemometer. The wind speed was instead acquired from the API because we were unable to use it in the prototype due to its high cost. ESP8266 wifi module was used to operate all of the sensors. All the results are then get uploaded to the cloud connected with the ESP8266 wifi module.

## IV. RESULTS

The website was able to successfully show the predicted disease and its preventive measure.




Figure 13. website showing prediction and prevention

An accuracy of 0.8 was achieved using Randomforest and 0.79 was achieved using XGboost. As there are no such dataset present currently, hence the accuracy will increase more when more accurate dataset regarding this field be generated.

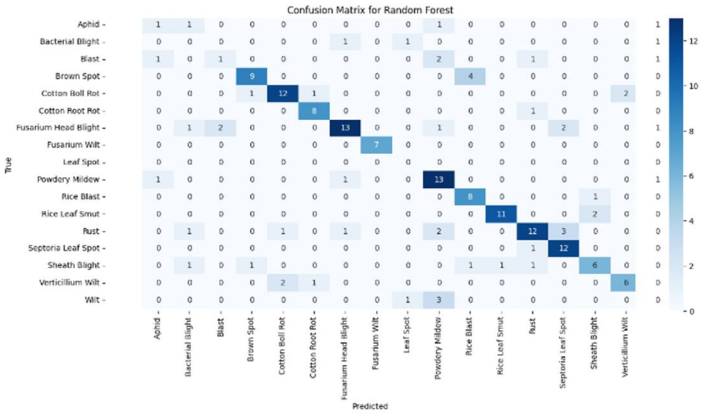



Figure 10. Confusion matrix for Random Forest

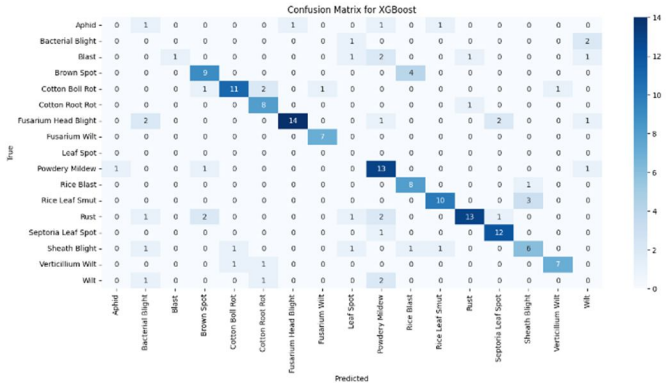



Figure 11. Confusion matrix for XGboost

## V. CONCLUSION

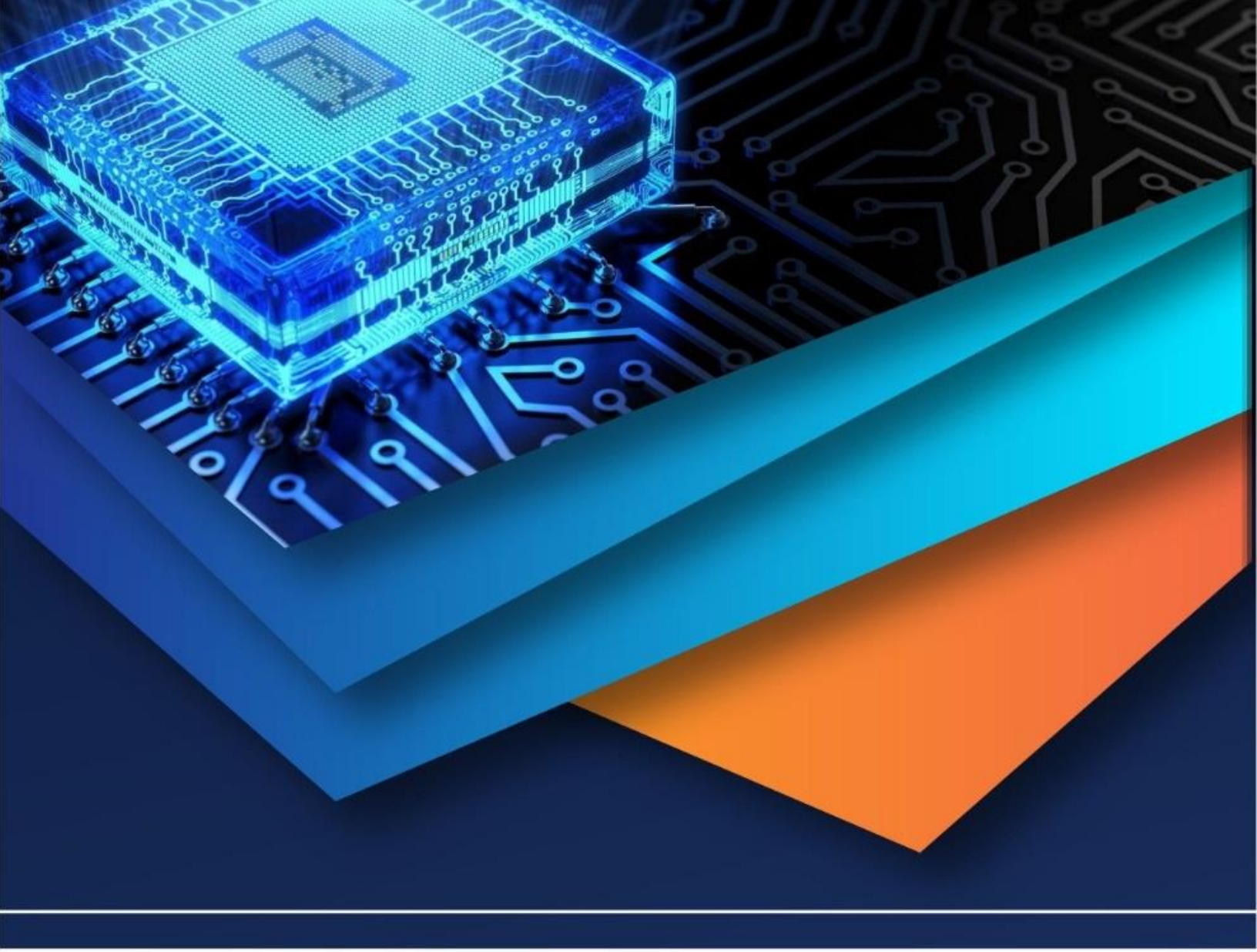
In conclusion, our study showed that using smart computer methods like XGBoost and Random Forest can help predict plant diseases based on weather. We can also use unsupervised and supervised learning techniques together in order to predict the diseases. By understanding these connections, farmers can better protect their crops, ensuring we have enough food for everyone. This research opens doors to more effective ways of safeguarding our food supply in the future.
Furthermore, the accuracy of the proposed system can be increased with more accurate and bigger dataset.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

## REFERENCES

[1] P. Kapoor and F. A. Barbhuiya, "Cloud Based Weather Station using IoT Devices," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, pp. 2357-2362, doi: 10.1109/TENCON.2019.8929528.
[2] P. G. Krishna, K. Chandra Bhanu, S. A. Ahamed, M. Umesh Chandra, N. Prudhvi and N. Apoorva, "Artificial Neural Network (ANN) Enabled Weather Monitoring and Prediction System using IoT," 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2023, pp. 46-51, doi: 10.1109/IDCIoT56793.2023.10053534.
[3] M. Sadhukhan, S. Dasgupta and I. Bhattacharya, "An Intelligent Weather Prediction System Based on IOT," 2021 Devices for Integrated Circuit (DevIC), Kalyani, India, 2021, pp. 528-532, doi: 10.1109/DevIC50843.2021.9455883.
[4] Agrawal, R. and Mehta, S.C., 2007. Weather based forecasting of crop yields, pests and diseases-IASRI models. J. Ind. Soc. Agril. Statist, 61(2), pp.255-263.
[5] Soil Analysis and Crop Prediction Shubham Prabhu, Prem Revandekar, Swami Shirdhankar, Sandip Paygude Department of Electronics Engineering, KJ Somaiya Institute of Engineering and Information Technology Sion, Maharashtra, India.
[6] N, Banupriya & Tejasvi, D & Vaishnavi, P. (2022). CROP YIELD PREDICTION BASED ON INDIAN AGRICULTURE USING MACHINE LEARNING. 2021.
[7] Shafik, Wasswa & Tufail, Ali & Namoun, Abdallah & Silva, Chandratilak & Anna, Rosyzie & Haji, Awg & Apong, Mohd. (2023). A Systematic Literature Review on Plant Disease Detection: Motivations, Classification Techniques, Datasets, Challenges, and Future Trends.
[8] Kang, Wee Soo, Soon Sung Hong, Yong Kyu Han, Kyu Rang Kim, Sung Gi Kim, and Eun Woo Park. "A web-based information system for plant disease forecast based on weather data at high spatial resolution." The Plant Pathology Journal 26, no. 1 (2010): 37-48.
[9] Rao, Madhuri & Singh, Arushi & Reddy, N V Subba & Acharya, Dinesh. (2022). Crop prediction using machine learning. Journal of Physics: Conference Series. 2161. 012033. 10.1088/1742-6596/2161/1/012033.
[10] Shams, Mahmoud Y., Samah A. Gamel, and Fatma M. Talaat. "Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making." Neural Computing and Applications (2024): 1-20.
[11] Tilva, V., Patel, J., & Bhatt, C. (2013). Weather based plant diseases forecasting using fuzzy logic. Proceedings of the 2013 Nirma University International Conference on Engineering (NUiCONE).
[12] Fernandes, J. M. C., Pavan, W., & Sanhueza, R. M. (2011). SISALERT - A generic web-based plant disease forecasting system. In M. Salampasis & A. Matopoulos (Eds.), Proceedings of the International Conference on Information and Communication Technologies for Sustainable Agri-production and Environment (HAICTA 2011), Skiathos, 8-11 September, 2011.
[13] Liu, Z., Bashir, R. N., Iqbal, S., Shahid, M. M. A., Tausif, M., & Umer, Q. (2022). Internet of Things (IoT) and machine learning model of plantdisease prediction—Blister blight for tea plant. https://doi.org/10.1109/ACCESS.2022.3169147
[14] Gokulnath, B. V., & Devi, G. U. (2020). A survey on plant disease prediction using machine learning and deep learning techniques. Inteligencia Artificial, 23(65), 136-154. https://doi.org/10.4114/intartif.vol23iss65pp136-15
[15] Khattab, A., Habib, S. E. D., Ismail, H., Zayan, S., Fahmy, Y., & Khairy, M. M. (2019). An IoT-based cognitive monitoring system for early plant disease forecast. Computers and Electronics in Agriculture, 166, 105028
[16] ED, Madden LV. 2019 Predicting plant disease epidemics from functionally represented weather series. Phil. Trans. R. Soc. B 374: 20180273. http://dx.doi.org/10.1098/rstb.2018.0273
[17] Rosenzweig, Cynthia; Iglesius, Ana; Yang, X. B.; Epstein, Paul R.; and Chivian, Eric, "Climate change and extreme weather events - Implications for food production, plant diseases, and pests" (2001). NASA Publications. 24

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)