Mango Diseases Prediction Using Machine Learning Algorithms

Lavanya. K [1[0000-0001-7097-7149]], Dr. A. Packialatha[2[0000-0001-7374-1262]]

K.S. Archana [3[0000-0002-2596-3941]]

[1]Research Scholar, Department of CSE, Vels Institute of Science, Technology and Advanced Studies, Pallavaram,Chennai

[2]Assistant Professor, Department of CSE,Vels Institute of Science Technology and Advanced Studies, Pallavaram, Chennai

[3]Assistant professor, Department of Data science and Business systems, School of Computing SRM Institute of Science and Technology,Kattankulathur, Chengalpattu, District - 603203

lavanyasanat@gmail.com[1],packialatha.se@velsuniv.ac.in[2], archanak1@srmist.edu.in[3].

Abstract. The country's economy relies heavily on agriculture, and the health of the crops is essential to its success. Crop diseases that go undetected can cost the agricultural industrially detection and identification are crucial. It is possible to avoid crop diseases from destroying the harvest if they are accurately diagnosed and detected. Because healthy and diseased plants appear identical in their early stages, farmers cannot distinguish between the two by watching the crop leaf. India exports vast mangoes, making it an economically and environmentally significant fruit. About 1500mango species are grown in India, with over 1000 commercial types. There are a lot of diseases that harm mangoes, affecting their look, taste, and economy. Mango trees in India are plagued by a fungus called Anthracnose, which is the most frequent disease of its kind. Anthracnose, a highly contagious fungus, requires a quick and accurate method of diagnosis. As a result, an in-depth examination of the plants is essential before initiating any control measures. The Prognosis of Disease in the Mango Fruit Crop IoT and machine learning are used in a complex warning system. One of the main objectives is to develop a system that can forecast disease outbreaks on mango fruit harvests using historical weather information and crop yield. The field sensors collected current weather information to detect disease immediately. A regression method that makes use of a random forest is the random forest regression. This work investigates machine learning (ML) methods for identifying and categorising illnesses in mango plants. In this paper, the effectiveness of ML-based classification models for mango crops, as well as their datasets and feature extraction methods, are assessed. Finally, many challenges related to identifying plant diseases are examined

Keywords : Internet of Things, Temperature, prediction, Machine Learning, Sensors, Prediction.

1 Introduction

Throughout the world, agriculture became the primary duty for enhancing the country's economic contribution. Nowadays, most agricultural lands are underdeveloped because ecosystem control methods have not been widely adopted. These issues prevent the mango crop output from increasing, which has an impact on the agricultural sector. Considering the plant yield prediction, agricultural productivity has increased. [1]. It would be beneficial for farmers to know the mango crop yield before it is harvested so they can take the necessary marketing and storage precautions. The Prognosis of Disease in the Mango Fruit Crop IoT and machine learning are used in a complex warning system The concept of agricultural productivity was developed using plant yield predictions. Using a dataset, agricultural sectors must anticipate the mango crop using machine learning techniques. [2]. This research will help before cultivating on the agricultural area, farmers can determine the production of their mango crop, assisting them in making the best choices. By creating a working prototype of an interactive prediction system, it tries to fix the issue. There are various methods or algorithms for data analysis in mango crop prediction, and we can predict mango crop production with the use of those methodologies. It employs the random forest algorithm. There aren't any suitable solutions or technologies to deal with the scenario we are in right now, despite the analysis of all these problems, including weather, temperature, humidity, rainfall, and moisture. In India, there are numerous approaches to boost agricultural economic development. Predicting mango crop yield production uses data mining. Data mining, in general, is the process of dissecting data from several perspectives and distilling it into useful knowledge. The most well-known and effective supervised machine learning algorithm, known as random forest, can perform both classification and regression tasks. It builds an infinite number of decision trees during training, and it outputs classes based on the average prediction (for regression) or mode of the classes (classification) of each tree. Mango crop yield forecasting uses machine learning as a decision support tool, helping with decisions about what to plant and what to do during the growing season. To facilitate studies on mango crop yield prediction, several algorithms have been used.

Machine learning (ML) techniques are utilized in a variety of industries, from supermarkets to analyze consumer behavior to phone usage forecasting. The development of agriculture in modern times is influenced by numerous technologies, environments, and techniques. Additionally, the use of information technology may alter the circumstances of decision-making, allowing farmers to produce in the optimal way [7]. Indians have been engaged in agriculture for a long time, but the yield of the mango crop is never satisfactory due to several issues [10]. Agriculture has long employed machine learning in this industry. Predicting mango crop production is one of the challenging issues in precision agriculture, and numerous models have been presented and proven up to this point. Because mango crop yields rely on several variables, including climate, weather, soil, application of fertilizer, and cultivar, this problem necessitates the use of numerous datasets. This demonstrates that predicting mango crop yields is a difficult undertaking that requires several intricate steps. Currently, models for predicting the yield of the mango crop may reasonably approximate the actual yield, but improved yield prediction performance is always desired [11] [12]. The foundation of the Indian economy is agriculture. Weather

conditions play a major role in agricultural productivity. Rainfall is mostly needed for rice farming. To help farmers maximize the agricultural production of their mango crops, timely advice is required for forecasting future crop yields and analysis. Forecasting agricultural yields is a crucial topic [13] [14]. Farmers used to forecast their yields based on the yield experience from the prior year. Therefore, there are various methods or algorithms for this kind of data analysis in mango crop prediction, and we can forecast the yield of the crop using these algorithms [15]. It employs a random forest algorithm. Big data analysis techniques in agriculture are finding more and more uses and roles thanks to all these algorithms and the linkage between them. The field of agriculture has been slowly deteriorating with the development of new technology and processes. These caused individuals to concentrate on growing fake, hybrid items, which resulted in a life that was unhealthy. People in the modern world nowadays are unaware of agriculture. Due to these methods, seasonal climatic conditions are also changing at the expense of resources like land, water, and air, which results in food insecurity. There are no suitable solutions or technology to deal with the situation we are facing despite analyzing all these issues including weather, temperature, and several other elements. There are various approaches to quicken the agriculture sector's economic growth in India. Mango crop productivity and quality can be increased and improved in several ways. The yields of mango crops can also be predicted using data mining. The article concludes that the agriculture sector would benefit from cost-effective solutions due to the quick development of sensor technology and ML approaches.

2 Literature Review

Experiments by Aruvansh Nigam, Saksham Garg, and Archit Agrawal [3] on a dataset from the Indian government have demonstrated that the Random Forest machine learning algorithm has the best yield forecast accuracy. Here, we have put the theoretical aspect of risk forecasting into practise. The Jupyter Notebook is one technology that any information scientist should be familiar with or utilise. Jupyter Notebooks are robust, adaptable, shareable, and give you the freedom to carry out data visualisation in the same setting. Data scientists may create and share documents, ranging from codes to comprehensive reports, using Jupyter Notebooks. Jupyter notebooks use a highly step-by-step approach to compose items like code, graphics, text, output, etc. to demonstrate the analytical process step-by-step. Each algorithm's output is recorded in a Jupyter Notebook and compared to one another. Performance is then evaluated using Accuracy [5].

Table 1. Recent research works on Mango disease classification

References	Model/Methodology	Disease/Description	Data set	Metrics
[1] (2021)	FrCNet	Heterogeneous	Self-acquired Dataset (Mango research institute)	Accuracy= 99 FNR=0.8%
[2] (2020)	VGG-V2Incepnet	Anthracnose	1198 self-acquired images and 1070 images from the plant villages dataset	Accuracy= 92%
[3] (2020)	ANN AlexNet, VGG16, ResNet-50	The authors discovered tiny disease blobs on plant leaves using higher solution images.	Self-acquired dataset with four categories	Accuracy, precision, Recall,F1, Score
[4] (2020)	RF, SVM Feature extraction (PCA,LDA)	Anthracnose S	Self-acquired dataset	Accuracy= 91%
[5] (2020)	VGN 6	Using a VGG-16-based deep learning network, designed a system that can classify pests into multiple categories.	Self-acquired Dataset with data augmentation	Accuracy= 73%

Reference Model/Methodology Disease/Description Data Set Metrics

3 RESEARCH METHODOLOGY

Any machine learning system requires data. The dataset has now been provided to the machine learning model, which utilized this collection of data to train. Every new detail entered into the application form acts as a test set of data. Following the testing process, the model makes predictions based on the conclusions drawn from the training data sets. A lot of people have used satellite imagery to forecast mango crop yield [8]. It is vital to collect data at the district level since local climates vary. To put the system into place, historical information on the culture and climate is required. This information is gathered from a few websites. Climatic parameters that most affect mango crops are Ph, K, N, P, Temperature, humidity, rainfall, label. Therefore, data on these climate parameters were collected on a monthly level.

3.1 Data

This stage involves gathering data from multiple sources and setting up the dataset. And analysis is used to offer the data set. The entire dataset is split into two halves; for

instance, 20% of the data are used to test the model, and the remaining 80% are used to train the model. To anticipate upcoming events Machine Learning Methodologies
supervised education: The data needed to train the system, which was previously labelled with the pertinent right responses, is necessary for supervised learning. Classification problems can be used to better categorise problems with supervised learning [4]. Machine learning algorithms under supervision can use examples to apply what has been discovered in the past to fresh datalabelled .The system can offer a target for any new entry after proper training. The learning algorithm can also tell its outputs apart from the accurate and anticipated output and can detect flaws, which may then be corrected in the model.

3.2 Unsupervised Learning

When the data used for training cannot be categorized, unsupervised machine learning techniques are utilized. When describing a hidden structure from unlabeled data, unsupervised learning examines how a system may handle a function. The system estimates the output to explain hidden structure from unlabeled data, but it also analyses the data and may make inferences from the data sets.

3.3 Classification Algorithms

It is necessary to compare the performance of multiple different machine learning algorithms. Each model will have different performance characteristics [9]. In this study, we utilised logistic regression. It is commonly a supervised kind set of rules. For a given collection of functions, X, the goal variable, y, in a class hassle, can accept great discrete values. Contrary to popular misconception, logistic regression IS a type of regression. The version creates a regression model to anticipate the possibility that a set of facts will be admitted to the elegance designated as "1." Similar to how linear regression makes the assumption that the data follows a linear function, logistic regression uses a sigmoid function to represent the data. Even when a choice criterion is included, logistic regression excels as a class technique. The placement of the edge price, which depends on the class problem itself, is a crucial component in logistic regression. The accuracy of the result was 95.22%t.
In this we carried out SVM set of rules. Support Vector Machine (SVM) is a supervised device studying algorithmic software used for every class. Although we're pronouncing regression problems but its excellent acceptable to class. Searching for a hyper plane in an N-dimensional area that categorizes the data points with certainty is the aim of the SVM set of rules. The number of options affects how big the hyper plane is. The hyper plane is a line if the number of enter functions is two. The hyper plane becomes a 2-D plane if the number of enter functions is three. Once there are more than three functions, it is difficult to think clearly. We got SVM with the accuracy of 10% Utilizing Random Forest Classifier, the random forest maximum famous and effective supervised device studying set of rules able to act each class and regression tasks, that function through building a mess of selection bushes on the time of schooling and producing outputs of the elegance this is the mode of the classes (class) or imply prediction (regression) of the man or woman bushes. The greater bushes in a woodland the greater sturdy the prediction, showed an accuracy of 99.09%

We performed Decision Tree which is a Supervised studying method that may be used for each classification and Regression problems, however in most cases it's far desired for fixing Classification problems. The choices or the look at are carried out based on the notion that the supplied dataset's functions. It is a graphical representation of the process of purchasing all potential solutions to a problem or selection made entirely based on predetermined criteria. There are diverse algorithms in machine learning, so selecting the excellent set of rules for the given dataset and hassle is the primary factor to don't forget whilst developing a device studying version. The Decision Tree classifier showed an accuracy of 90%. The probability of features is assumed to be-

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

Finding the mean and standard deviation in each label, which is all that is required to determine such a distribution, will suffice to change this model. By providing climate data of this site, a user-friendly website is developed to predict mango crop yields that can be used by any user according to their mango crop selection. It showed an accuracy of 99.09%.

4 Result

With the use of the study's data, we first tested the Decision Tree, yielding the results displayed in the table. An accuracy of 90% was shown by a decision tree classifier using the data examined in this research. Based on the findings of our investigation, we performed Gaussian naive bayes, and the classifier yielded a 99.09 percent accuracy rate. Then we carried out SVM with an accuracy of 10%. Later we tested with the logistic regression classifier, obtaining the result with accuracy 95.22%. Finally, using Random Forest to analyze the study's data revealed accuracy of 99.09%, which was on par with Gaussian naive bayes accuracy of 99.09%.

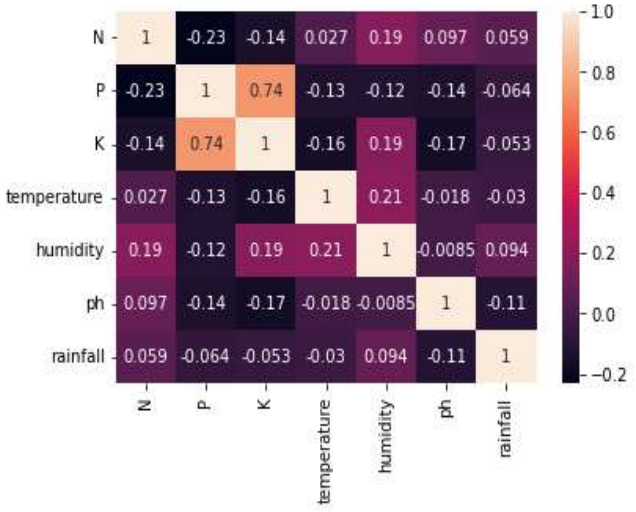

Fig.1 Heat map

Table 2. Classification Report for Decision Tree

precision	recall	f1-score		support
accuracy	0.90			440
macro avg	0.84	0.88	0.85	440
weighted avg	0.86	0.90	0.87	440

Table 3. Classification Report for Gaussian naïve bayes

precision	recall	f1-score		support
accuracy	0.99			440
macro avg	0.99	0.99	0.99	440
weighted avg	0.99	0.99	0.99	440

Table 4. Classification Report for SVM

precision	recall	f1-score		support
accuracy	0.11			440
macro avg	0.66	0.13	0.14	440
weighted avg	0.66	0.11	0.13	440

Table 5. Classification Report for Logistic Regression

precision	recall	f1-score		support
accuracy	0.95			440
macro avg	0.95	0.95	0.95	440
weighted avg	0.95	0.95	0.95	440

Table 6. Classification Report for Random Forest

precision	recall	f1-score		support
accuracy	0.99			440
macro avg0.99	0.99	0.99		440
weighted avg	0.99	0.99	0.99	440

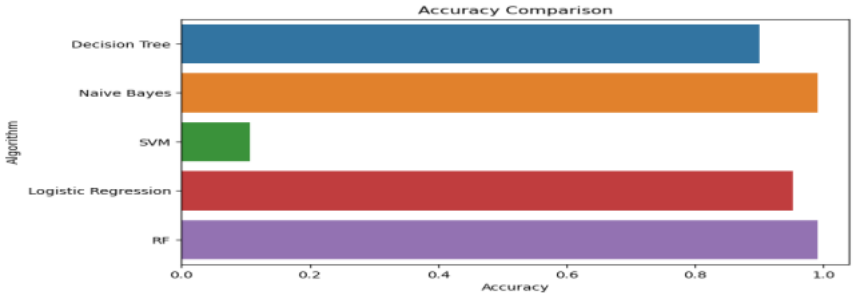

Fig.2 Accuracy comparison of all algorithms

5 Conclusion

Based on climate inputs, this study provides a demonstration of the possibility of using data mining methods to predict mango crop yields. The forecast accuracy is greater than 75% in all mango crops and districts chosen for the study, demonstrating a higher level of forecast accuracy, and the website is designed to be user-friendly. The suggested approach considers information about the soil, weather, and previous year's productivity and makes recommendations on the most lucrative mango crops that may be grown in the ideal environmental conditions. The system will cover the most varieties of mango crops, allowing farmers to learn about mango crops that may have never been grown before. By listing all conceivable mango crops, this system aids farmers in choosing which mango crops to produce [6]. This website's climate data is utilized to create a user-friendly website that forecasts mango crop yields and may be used by any user in accordance with their choice of mango crop.

The farmer was able to reduce the loss caused by trip outbreaks because to disease prediction. It aided farmers in taking preventative action that lessened production loss. Mango quality decline is also avoided, aiding farmers in raising yields and improving fruit quality. The random forest algorithm has been shown to be reasonably good at predicting the chance of trips. This method has aided farmers in taking preventative actions that have improved farm productivity and decreased the use of chemical pesticides on crops. It will eventually stop instances of alien flies in fruit lot shipments. With more accurate on-field sensors, the system can be expanded to be implemented in mango orchards. The low implementation cost will help in micro farming in India. The already developed accurate models of disease detection and classification can also be used in hand with this system to increase throughput.

6 References

[1] Sravan, Vemishetti, et al. "A deep learning-based crop disease classification using transfer learning." Materials Today: Proceedings (2021).

[2] Saleem, Rabia, et al. "Mango Leaf Disease Identification Using Fully Resolution Convolutional Network. "CMC COMPUTERS MATERIALS & CONTINUA 69.3 (2021): 3581-3601.

[3] Kusrini, Kusrini, et al. "Data augmentation for automated pest classification in Mango farms." Computers and Electronics in Agriculture 179 (2020): 105842.

[4] Pham, Tan Nhat, Ly Van Tran, and Son Vu Truong Dao. "Early disease classification of mango leaves using feed-forward neural network and hybrid metaheuristic feature selection." IEEE Access 8 (2020): 189960-189973

[5] Nagaraju, Y., et al. "Transfer learning based convolutional neural network model for classification of mango leaves infected by anthracnose." 2020 IEEE International Conference for Innovation in Technology (INOCON). IEEE, 2020

[6] Ardila, Carlos Eduardo Cabrera, Leonardo Alberto Ramirez, and Flavio Augusto Prieto Ortiz. "Spectral analysis for the early detection of anthracnose in fruits of Sugar Mango (Mangiferaindica)." Computers and Electronics in Agriculture 173 (2020): 105357

[7] Aruvansh Nigam, Saksham Garg, Archit Agrawal Mango crop Yield Prediction using ML Algorithms ,2019.

[8] An Overview of Internet of Things (IoT) and Data Analytics in Agriculture: Benefits and Challenges by Olakunle Elijah, Tharek Abdul Rahman, IgbafeOrikumhi, Chee Yen Leow, ISSN:2319-7242, Volume 5| Issue 6 |NO. 5, OCTOBER 2018

[9] Mango crop Prediction System using Machine Learning by Prof. D.S. Zingade, Omkar Buchade, Nilesh Mehta Shubha Ghodekar, Chandan Mehta, International Journal of Advance Engineering and Research Development , Volume 4 | Special Issue 5 | p-ISSN : 2348- 6406 Dec.-2017.

[10] E. Park, Y. Cho, J. Han, and S. J. Kwon, ―Comprehensive approaches to user acceptance of Internet of Things in a smart home environment, ‖ IEEE Internet Things J., vol. 4, no. 6, pp. 2342–2350, Dec. 2017

[11] R.K.Ritika, A. Srinivasulu, et al.,"State-of-the-Art Review on IoT Threats and Attacks: Taxonomy, Challenges and Solutions", Sustainability, vol. 13, Issue. 16, pp. 9463, 2021, doi: 2071-1050/13/16/9463

[12] IERC. (Mar. 2015). European Research Cluster on the Internet of Things-Outlook of IoT Activities in Europe. Accessed: Sep. 20, 2017. [Online]. Available: http://www.internetof-things-research.eu/pdf/. IERC_Position_Paper_IoT_Semantic_Interoperability_Final.pdf.

[13] N. Gandhi L.J. Armstrong O. Petkar "Predicting Rice Mango crop Yield using Bayesian Networks" communicated 2016.

[14] N. P. Sastra and D. M. Wiharta, ―Environmental monitoring as an IoT application in building smart campus of University as Udayana, ‖ in Proc. Int. Conf. Smart Green Technol. Elect. Inf. Syst. (ICSGTEIS), Oct. 2016, pp. 85–88.

[15] C.Ravariu, N.Jalaja, A. Srinivasulu,"Modeling of the DNA viruses invasion by logic circuits with finite states', Advanced Nano-Bio-Materials and Devices, vol.5, Issue.2, pp. 666-674, 2021.

[16] Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, ―Internet of Things for smart cities, ‖ IEEE Internet Things J., vol. 1, no. 1, pp. 22–32, Feb. 2014.

[17] L. Sanchez et al., ―Smart santander: IoT experimentation over a smart city testbed, ‖ Comput. Netw., vol. 61, pp. 217–238, Mar. 2014.

[18] S. Dahikar S. Rode "Agricultural mango crop yield prediction using artificial neural network approach" International Journal of Innovative Research in Electrical Electronics Instrumentation and Control Engineering vol. 2 no. l pp. 683-6862014.