IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 1, 2022

IDENTIFICATION OF PLANT DISEASE DETECTION USING MACHINE LEARNING

[1]Dr. M. APPARAO, [2]Dr. M. RAGHAVA NAIDU

[1]Professor & HOD, Department of ECE, Guntur Engineering College, Guntur, Andhra Pradesh, India
[2]Assistant Professor, Department of CSE, Krishna University College of Engineering and Technology, Machilipatnam, Andhra Pradesh, India

ABSTRACT: The Indian economic system relies heavily on productivity in farming, and disease detection is crucial for sustainable agriculture. Monitoring crops and identifying diseases requires extensive knowledge and processing time. Machine Learning algorithms can help predict increased agricultural yields, addressing a significant challenge in the industry. Applying machine learning algorithms to crop yield prediction with an emphasis on crop yield prediction is essential given the rising importance of agricultural output prediction. Machine learning supports the automatic diagnosis of plant illness based on the visual symptoms of the plant using methods such as random forest, decision tree, support vector machine, and Bayesian network, among others. The results show that this method improved the accuracy of disease progression level detection, diffusion pattern analysis, and sickness diagnosis. Plant diseases may therefore be precisely categorized, as demonstrated by machine learning techniques with a focus on identifying plant leaf diseases in crop yield prediction data. Hence, this model shows better results interms of accuracy and precision.

KEYWORDS: Support Vector Machine (SVM), Random Forest (RF), Bayesian Network, Decision Tree (DT), Machine Learning (ML).

I. INTRODUCTION

Farming is one of the most common professions in India. Agriculture is a major contributor to the economy of all nations. A growing worldwide population has led to food scarcity or shortages, hence modernizing the agricultural sector is necessary to maintain the status quo [1]. Growing population, inherent climatic unpredictability, soil loss, and demand practices that adapt to the environment all have linked effects that must be taken into consideration in order to ensure Additionally, it ought to increase the viability of agricultural food production [2]. These needs show that agricultural yield estimation, crop protection, and land value are becoming more crucial to the world's food production. As a result, accurate crop output estimation will be dependent on the country's policymakers obtaining desirable export and import valuations in order to improve national food security.

Both bacterial and fungal infections can harm crops. The impact on farmers output is significant. Crops need to be healthy in order to yield their maximum. Diagnosing diseases with our eyes always remains a complicated process. The farm must be continuously watched in order to achieve success. When the farm is big, it is also very expensive. Agricultural experts are also struggling to diagnose diseases and find solutions due to this challenge. Farmers would benefit greatly from a machine that could recognize plant diseases. This method can inform farmers in a timely manner so they may take the appropriate procedures [3].

Therefore, predicting agricultural production is difficult because of many complex factors. Crop output is affected by landscapes, soil quality, insect infestations, genotypes, water quality and availability, meteorological conditions, harvest scheduling, and other variables. Processes and methods for predicting crop yields are fundamentally nonlinear and time-specific [4]. These approaches are

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 1, 2022

also complex due to the inclusion of a wide range of associated parts that non-arbitration and outside sources can specify and impact.

Plant diseases are caused by many organisms such as infections, germs, and fungus. The automated disease detection procedure improves in the early detection of plant pathology. Early disease diagnosis has a positive impact on plant health. Most disease symptoms show on the leaves, branches, and fruits. Plant leaf indicators are used to diagnose illness faster, more accurately, and at a cheaper cost. In general, farmers use visual inspection to identify plant diseases, which enables disease detection and diagnosis. This requires a huge number of experts and constant plant monitoring, both of which are expensive when dealing with large farms [5]. Farmers in certain countries, however, without the necessary resources or knowledge on how to approach experts. Professional consultation is therefore both expensive and time-consuming.

It seems like the best way to deal with this problem is through Machine Learning. There are several ML algorithms that can classify and diagnose plant diseases automatically. The leaves, stem, seeds, and fruits are only a few of the elements of a plant that are affected by diseases. The diseases have varying effects on different areas of the plant's body [6]. The process of photosynthesis can only be accomplished with the help of leaves, despite the fact that the central component of a plant is commonly referred to as its leaves.

A plant's life cycle is immediately affected when it suffers from a foliar disease.To fight these diseases effectively, a method for automatically detecting and classifying diseases must be established. ML is a potential solution to this problem. Several ML methods have recently been proposed for the identification and classification of plant diseases.

II. LITERATURE SURVEY

R. Reda, B. Ilham, T. Saffaj, O. Saidi, L. Brahim, K. Issam, and E. M. El Hadrami et al. [7] Soil samples were taken from four agricultural fields in Morocco, and ML algorithms were used to estimate the soil organic carbon (SOC) and total nitrogen (TN). In contrast to conventional chemical procedures, near-infrared spectroscopy data are used since they require less processing time and use less resources. Compared to other regression models and back-propagation neural networks (BPNN), the ensemble learning modelling approach performs the best.

Y. Cai, D. Lobell, T. Xu, K. Guan, J. Peng, S. Wang, B. Potgieter, S. Asseng, A. Y. Zhang, L. You, and B. Peng et al. [8] Over fourteen years, ML algorithms were used to estimate wheat yield in Australia utilizing two data sources, meteorological data and satellite data (NN, SVM, and RF). According to the simulation's outcomes, when compared to satellite data for yield prediction, meteorological data provide different information. Crops planted on time are more productive and suffer less financial loss, which is exactly timing is so crucial.

Mohsen Azadbakht, Soheil Radiom, Hossein Aghighi, Davoud Ashourloo, and Abbas Alimohammadi et al. [9] Performance was examined for the high, medium, and low levels of the Leaf Area Index (LAI) in order to detect wheat leaf rust. Four methods namely random forest regression (RFR), v-support vector regression (v-SVR), gaussian process regression (GPR), and Boosted Regression Tree (BRT) were used to describe the performance of rust detection. To evaluate performance, an experiment was carried out on 7000 hectares of wheat farming in North West Iran. For the same purpose, hyperspectral reflectance data collected

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 1, 2022

under various environmental circumstances is used. The results of a comparison between ML and the Spectral Vegetation Index (SVI) demonstrate that ML exceeds SVIs.

Kerim Karada˘g, Mehmet Emin Tenekeci, Ramazan Ta¸saltın, and Ay¸sin Bilgili et al. [10] A technique has been proposed to aid in the detection of fusarium wilt in pepper. The spectroradiometer can learn more about the plant by observing how it reflects light. Between 350 and 2500 nm ought to be the wavelength. There are two layers of processing. 1. A feature vector; 2. A categorization of feature vectors. Three methods are used for classification purpose namely Artificial Neural networks (ANN), KNN and Naïve Bayes (NB). KNN has a success rate of 100 percent, while ANN has a success rate of 97.5 percent and Naıve Bayes has a success rate of 90 percent. Performance in classification is hampered by large data sizes. Large data dimensions decrease the performance of categorization. Reduced dimension is achieved through the use of Wavelet Transformation (WT). For this, Sym5, db2, and Haar are utilized. In ANN, the backpropagation algorithm is trained for classification. Using cross-validation on 10 fields, its performance is evaluated. For the purpose of implementing the classification method, Mat lab R2015b is utilized. Sym5, Haar, and db2 have success rates of 100%, 100%, and 98% for KNN when K=2. KNN produces better results in comparison to other algorithms. The characteristics are reduced through wavelet decomposition from 2150 to 75.

Cramer S., Freitas A. A., Kampouridis M.and A. K. Alexandridis et al. [11] shown that seven ML systems were efficient in predicting rainfall. According to statistical outcomes, radial-basis function neural network (RBFNN) performs better than other advanced algorithms.

Vijai, S., A.K.Misra et. al., [12] utilized genetic algorithm for plant disease detection. The only disease detection technologies now in use include plant disease specialists inspecting leaves in the field. This action requires a significant team of specialists and ongoing plant observation, which is quite expensive for large farms. In some countries proper agricultural facilities are not available. Hence, it is imperative for them to hire professionals. However, this results in time-consuming and costly activities. In such cases, this advice strategy is useful in preventing crop overproduction. Plant illness can be automatically detected by examining the state of the leaves, which is less expensive and more efficient. This machine vision provides robot guidance, inspection and automatic process control based on images. It takes a lot of effort and work to visually identify leaf diseases. However, this is not particularly reliable and can only be done in limited locations. However, it is quicker, faster reliable, and takes less time to detect plant diseases automatically rather than manually.

Chong K. L., Kanniah K. D., Pohl C. and Tan K. P. et al. [13] It includes a thorough analysis of the many remote sensing uses for palm oil plantations, such as tree counting, age estimate, change detection, pest and disease detection, estimation of above-ground biomass (AGB) and carbon output, and yield computation. They offered proposed fixes and pointed out potential research holes. The emphasis on palm oil yield by utilizing machine learning algorithms is not, highlighted in this assessment.

A. Morellos, G. Tziotzios, X. T. Alexandridis, E. Pantazi, R. Bill, D. Moshou, R. Whetton, J. Wiebensohn, and A. M. Mouazen et al. [14] Visible and infrared spectroscopy were also used to calculate TN, SOC, and Moisture Content (MC) in the Premslin, Germany field. A predictive Machine Learning (ML) model

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, Volume 11, Iss 1, 2022

is constructed using the spectroscopy dataset to predict the three soil properties. In terms of RMSE and residual prediction deviation (RPD), principal component regression and partial least square regression multivariate approaches are outperformed by least square support vector machine (LS-SVM) and cubist ML algorithms.. With RMSEs of 0.062 and 0.457 and RPDs of 2.24 and 2.20, LS-SVM can predict SOC and MC.

S. K. Singh, A. Porwal, R. Saxena, and S. S. Ray et al. [15] Hail storms in February and March 2015 alone resulted in an 8.4% drop in the country's wheat production, according to an evaluation of the impact of hail rains on India's wheat harvest. Correct weather forecasting can be a lifesaver for farmers, especially those who are economically disadvantaged and live in countries like India where it is difficult to store harvested products for long periods of time. When frequently used on a system, ML models serve as feedback controls. We are able to predict the variables that impact crop production with the use of precise ML models. Therefore, preventative measures can be taken before a crop production imbalance ever happens.

III. A Machine Learning Approaches with Special Emphasis For Identification Of Plant Leaf Diseases In Crop Yield Prediction

The block diagram of identification of plant diseases in crop yield by utilizing Machine Leaning approaches with special emphasis is represented in fig.1.

Gather pictures of healthy and diseased plants. The training time was calculated using the Open Source Computer Vision Library (OpenCV) framework using a Python script that automatically cropped the photos. A sample collected in the initial step and then sample is pre-processed and labelled. The labelled part is transferred to the input image. Then the segmentation is done with the feature extraction is done. Augmentation process is done after the feature extraction. The augmentation process is divided into regression and classified. Then finally get the output.

A Python script that makes use of the OpenCV framework included photos that were automatically scaled in the dataset to reduce training time. Scaling the data points from [0, 255] (the image's minimum and maximum Red Green Blue (RGB) values) to [0, 1] pre-processes the input data. The plant's leaf's picture is photographed. The picture is in RGB format.

A training part and a testing segment have been created from the dataset. The dataset has 20% for testing and 80% for training. 16,511 photos make up the training dataset, whereas 4,128 images make up the testing dataset. The test data set is hidden from view to test the model's accuracy while the model is trained using the training data set. The division of a picture into pixel-like characteristics is known as image segmentation. It mostly supports the visual interpretation process. It moves an image from a lower level to a higher level. Its capacity to successfully analyse an image is largely dependent on the precise the segmentation process. Contextual and non-contextual factors both have a role in segmentation. The segmentation procedure makes use of a number of algorithms.

A method to expand the quantity of photographs in a database is called data augmentation. In order to diversify our collection, several procedures such as moving, rotating, stretching and rotating are applied to the image datasets. Overfitting can be minimized during the training phase by expanding the dataset and introducing distortion to the images. In-place or real-time data augmentation are supported by Keras ImageDataGenerator class. They can make sure that our network

ISSN PRINT 2319 1775 Online 2320 7876

Research paper Volume 11, Iss 1, 2022

learns about new deviations by providing this type of data. This allows us to come up with higher results using a smaller dataset.

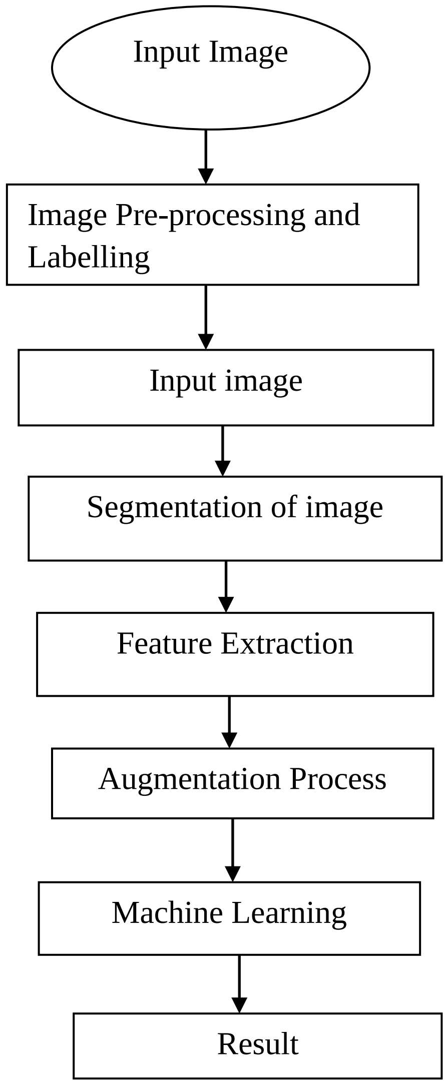

Fig.1 Block Diagram Of Identification Of Plant Diseases In Crop Yield By Using Machine Leaning Approaches With Special Emphasis

Crop production has been predicted using a wide range of regression- and classification-based prediction techniques. k-nearest neighbours (K-NN), linear Regression (LR), support vector machine (SVM) and decision tree (DT), support vector regression (SVR), random forest (RF), and Extremely Randomised Trees (extra tree) (ERT) are utilized in agricultural yield estimate.

The interaction between an independent variable and one or more dependent variables is represented by the LR model. Regression algorithms are used in a machine learning framework to minimise loss or error (RMSE or MSE), which is the way learning is achieved. Multiple Multiple Linear Regression (MLR) implementations, analysis have shown that multiple independent variables are more effective and dependable than a single individual element

In order to compute regression from a large number of decision trees, it creates a wide range of logistic regression. The RF performs more accurately than any other Decision Tree, and the randomization allows the single decision tree to make up for the bias. After classification then the result is evaluated.

IV. RESULT ANALYSIS

The performance of plant foliar disease detection in agricultural yield prediction by ML methods.

Table 1: Performance Comparison Table

Classifiers	**Precision**	**Accuracy**
SVM	**97**	**99**
RF	**98**	**98.6**
CNN	**70**	**75**

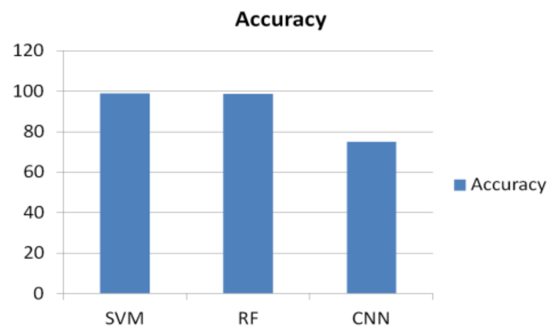

Fig.2: Accuracy Comparison Graph

In Fig.2 accuracy comparison graph is observed between SVM, RF and CNN.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, Volume 11, Iss 1, 2022

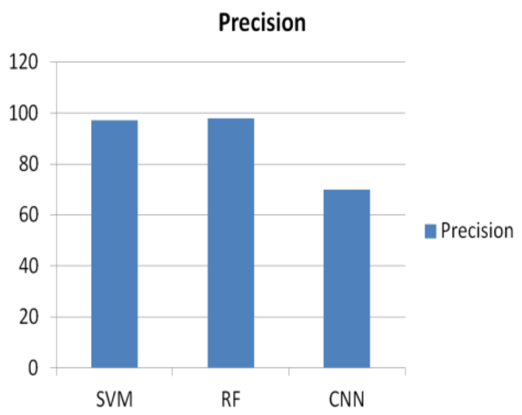

Fig.3: Precision Comparison Graph

Precision comparison graph is observed in Fig.3 between SVM, RF and CNN.

V. CONCLUSION

The detection of plant leaf disease in agricultural output prediction using machine learning approaches is presented in this inquiry. An ML classifier, in this case utilizing SVM and RF, is used to predict plant leaf disease. Although it is challenging to pinpoint the most effective strategies due to the state of the research, some machine learning algorithms are in use and show a lot of promise. Searching for illnesses and keeping an eye on plant health are essential for sustainable agriculture. This necessitates rapid processing times, a large amount of manpower, and understanding of plant diseases. In conclusion, a leaf is regarded as sick if a test for plant leaf disease is positive; if not, it is regarded as normal. The outcomes demonstrate that the presented system outperforms other classifiers in terms of accuracy and precision.

VI. REFERENCES

[1] Arivazhagan, S., Newlin S., Ananthi, S., Vishnu V, (2013). "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features", Agric Eng. Int CIGR, 15, pp.211-217.

[2] J. Huang, J. L. Gómez-Dans, H. Huang, H. Ma, Q. Wu, P. E. Lewis, S. Liang, Z. Chen, J.-H. Xue, Y. Wu, F. Zhao, J. Wang, and X. Xie, (Oct. 2019). ''Assimilation of remote sensing into crop growth models: Current status and perspectives,'' Agricult. Forest Meteorol., vols. 276–277, Art. no. 107609.

[3] Revathi, P., Hemalatha, M. (2012). "Classification of cotton leaf spot diseases using image processing edge detection techniques".169- 173.10.1109/ INCOSET.2012.6513900J.

[4] Camargo, A., Smith, (2009). "Image pattern classification for the identification of disease causing agents in plants", COMPUT ELECTRON AGR, pp 121-125.

[5] R. Kaundal, A. S. Kapoor, and G. P. S. Raghava, (Nov. 2006). "Machine learning techniques in disease forecasting: a case study on rice blast prediction," BMC Bioinformatics, vol. 2006, no. 7, pp. 1–16.

[6] J. Liu and X. Wang, "Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network", *Front. Plant Sci.*, vol. 11, [online] Available: https://www.frontiersin.org/articles/10.3389/fpls.2020.00898.

[7]R. Reda, T. Saffaj, B. Ilham, O. Saidi, K. Issam, L. Brahim, and E. M. El Hadrami, (Dec. 2019,). ''A comparative study between a new method and other machine learning algorithms for soil organic carbon and total nitrogen prediction using near infrared spectroscopy,'' Chemometric Intell. Lab. Syst., vol. 195, Art. no. 103873.

[8] Y. Cai, K. Guan, D. Lobell, A. B. Potgieter, S. Wang, J. Peng, T. Xu, S. Asseng, Y. Zhang, L. You, and B. Peng, (Aug. 2019). ''Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches,'' Agricult. Forest Meteorol., vol. 274, pp. 144–159.

[9] Mohsen Azadbakht, Davoud Ashourloo, Hossein Aghighi, Soheil Radiom, and Abbas Alimohammadi, (2019). "Wheat leaf rust detection at canopy scale under different lai levels using machine learning techniques".

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 1, 2022

Computers and Electronics in Agriculture, 156:119–128.

[10] Kerim Karadağ, Mehmet Emin Tenekeci, Ramazan Taşaltın, and Ayşin Bilgili, (2019). "Detection of pepper fusarium disease using machine learning algorithms based on spectral reflectance". Sustainable Computing: Informatics and Systems.

[11] S. Cramer, M. Kampouridis, A. A. Freitas, and A. K. Alexandridis, (Nov. 2017). ''An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives,'' Expert Syst. Appl., vol. 85, pp. 169–181.

[12] Vijai, S., A.K.Misra, (2017). "Detection of plant leaf diseases using image segmentation and soft computing techniques", Information Processing in Agriculture, vol 4, pp 41-49.

[13] K. L. Chong, K. D. Kanniah, C. Pohl, and K. P. Tan, (Apr. 2017). ''A review of remote sensing applications for oil palm studies,'' Geo-Spatial Inf. Sci., vol. 20, no. 2, pp. 184–200.

[14] A. Morellos, X. E. Pantazi, D. Moshou, T. Alexandridis, R. Whetton, G. Tziotzios, J. Wiebensohn, R. Bill, and A. M. Mouazen, (Dec 2016). ''Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy,'' Biosyst. Eng., vol. 152, pp. 104–116.

[15] S. K. Singh, R. Saxena, A. Porwal, and S. S. Ray, (2015). ''Assessment of hailstorm damage in wheat crop using remote sensing,'' Current Sci., vol. 112, no. 10, pp. 2095–2099.