International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5563
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

Machine Learning based Classification of Diseased Mango Leaves

A. Selvakumar
[1]School of Computer Science and Engineering,
Vellore Institute of Technology (VIT),
Chennai, India
Email: selvacsecit@gmail.com

A. Balasundaram
[2]School of Computer Science and Engineering,
Center for Cyber Physical Systems,
Vellore Institute of Technology (VIT),
Chennai, India.
Email: balasundaram.a@vit.ac.in

Abstract-The preponderance of population depends on agriculture to produce crops which would be their primary subsistence for their livelihood. So, agriculture is considered the backbone of any nation. Mango (Mangifera indica Linn), belonging to a family Anacardiaceous, is a conspicuous fruit that captivates all ages because of its meticulous taste, delicious flavor, ampleness variety, and highly lustiness. Mangoes are generally rich in minerals, vitamins, fibers, and negotiable fat. Mango plants are exposed to many micro-organisms. If these are not detected and treated in the initial developing stages, it would affect peculiar parts of the mango plant and result in loss of overall productivity. Several factors like biotic and abiotic always ensue in the decrease in the overall productivity of mango plants. Self-regulated Detection of mango plant disease is imperative, and it must be detected at the preliminary stages of the growing period of the mango plant. This paper discusses the existing methodology to classify diseases in mango plant leaves by implementing ensemble technique (Stack) which includes algorithms like Decision Tree (DT), Support vector machine (SVM), Neural Network (NN), and Logistic Regression (LR). The developmental results validate that the disease classification methodology can successfully classify a higher percentage in predicting whether mango plant leaf is healthy or diseased.

Keywords- Machine learning; classification; decision tree; support vector machine; neural network; Adam optimizer; logistic regression

1 Introduction

Agriculture is considered a nation's backbone. In India, 70 percent of the total country's population relies on agriculture as the primary source of income for their livelihoods. In general, farmers face several issues in agriculture just because of plant diseases mainly triggered by changes in climatic conditions. Crop production is primarily affected by bacteria, fungi, and viruses, which causes various plant diseases that result in the decline of the overall output. Detecting plant disease manually may result in false perception, and this process is time-consuming. In disease detection, we lack professionals to make a reliable diagnosis; thereby, insecticides could be wrongly applied to the plants [1]. Mango, which is treated like a king of fruit, has a distinctive feature that can be matured and cultivated in various parts of the country, which can acclimate to several abiotic factors like rainfall, moisture, temperature, rainfall, and nutrient deficiency [2].

As India's national fruit, Mango is still afflicted by micro-organisms such as fungi, bacteria, parasites, and algae, lowering our productivity. Powdery Mildew, Anthracnose, Bacterial Canker, Phoma Blight, Red Rust, Die Back, Scab, gummosis, and malformation are all caused by these bacteria. Despite the evidence that these diseases are affecting the productivity of mangoes, India is still the primary contender in the production of mangoes, competing with the rest of the world's output. If the infection in mango plant leaves is identified early and treated, no country in the world will be able to compete with India's mango production.

Several machine learning-based disease diagnostic models have been presented, requiring less training and diagnosis time; nevertheless, these approaches have several drawbacks.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5563
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

Machine learning methods rely on smaller data sets, which leads to overfitting; also, these algorithms aren't fully automated because feature extraction requires human intervention, which makes diagnosis much more difficult [4]. Deep Learning (DL) techniques, which rely on a very large dataset to overcome the problem of overfitting, and Deep Learning-based models that can recover comparable traits on their own, have overcome the overfitting problem on their own. Multiple linked processing units (neurons) interact to solve issues in DL, which is a subset of machine learning and artificial intelligence. Deep learning (DL) has been successfully applied in a variety of sectors, including bioinformatics, agriculture, and drug design [4],[3]. When compared to machine learning techniques, DL reduces processing time, error rate, and
achieves a high precision rate in classification problem. The main goal of our research is to detect plant diseases and propose treatments to help them recover, using the CNN model [5].

2 Related Works

Table 1: Survey of algorithms for leaf disease classification

Reference	Approach used	Achieved	Performance Metrix
[1]	SVM, Logistic Regression and Random Forest	To classify plant infection	Able to achieve 87.6%.
[3]	SVM, ANN	To determine health status of plant	- SVM performed better than ANN.
[4]	CNN, AlexNet	Detect and classify plant disease	Able to achieve 96.5%
[5]	CNN	Plant disease detection for varieties of plant.	Able to achieve 86%
[7]	Red Rust, Bacterial canker Leaf and fruit	Self-acquired Dataset	BPNN has an accuracy of 99.68%.
[12]	Multiple Disease	Self-acquired Dataset	SVM has an accuracy of 80.00%.
[13]	CNN	Able to train the machine	Able to achieve 88.8% and 12.2% can be also filled.
[8]	AlexNet and CNN	Detect disease in mango and potato leaf.	Able to achieve 98.33%
[14]	AlexNet and CNN	Leaf disease detection.	Simple AlexNet – 96.34%. Hybrid of AlexNet and SVM layers – 99.98%.

Tab. 1. shows the list of the survey of algorithms for the classification of disease with their performance metrics.

3 Disease Classification Methodology

This section describes a method for detecting plant infection that is currently in use. We can classify the leaf as having some disease infection using the image and different image alteration algorithms. To achieve the rigor of disease diagnosis, the framework is divided into multiple stages, as indicated in Figure 1. Our system's four steps are outlined below.

Step 1: The image dataset is imported from Kaggle's 'Plant-Village' archive.

Step 2: The archived images are pre-processed, including image resizing, rescaling, and arrangement conversion.

Step 3: Next, the image dataset is spilled in the ratio of 80:20 for training and testing.

Step 4: Finally, it was decided to create a standard for images. The training dataset is the input to the CNN model, and the weights are changed to accurately identify the diseases in the mango plant leaf.

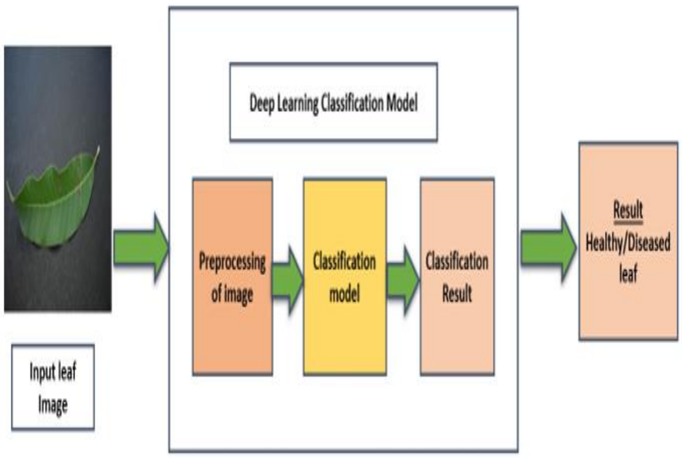

Figure 1: Block diagram of leaf disease classification methodology

In this section, we will be seeing about the basic steps of identification and classification of mango plant leaf disease using the processing on the data collection are shown in the Fig.1.

3.1 Image Acquisition and Pre-processing

The images (healthy/ diseased) of mango plant leaves are taken Kaggle plant village dataset. These images are represented in different colors like Red, Blue, and Green. The dataset includes images that can vary with resolutions and can be in another format; because the image files were downloaded from the plant village dataset. We resize the images as 256X256 pixels to achieve high accuracy, improve extraction of features, and reduce training time.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5563
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

3.2 Image Augmentation

Image augmentation is used to maximize the image size in the dataset. Image rotation, affine transformation, strength transformation, and perspective transformation are skill sets used in the image dataset. Finally, we obtained an image database that contains 4335 images.

3.3 Image Classification

Database that contains images is divided into different sets like validation, training, and test. By dividing the images that we obtained into 70:30, i.e., 70% for processing of images(training) and 30% for testing and validation. To avoid overfitting problems, the validation set of images is not the same as the trail set of images, but it is included in the testing set of images.

4 Performance Metrics

The following are the performance metrics used to measure the machine learning model that has been applied in this work.

4.1 Classification Accuracy

It can be defined as the ratio of the total number of successfully classified samples to the total number of samples that have been given.

$$\text{Classification Accuracy (CA)} - \frac{TP + TN}{TP+TN+FP+FN} \quad (1)$$

4.2 Precision

It can be defined as the number of positive samples that is taken out from the total number of samples.

$$\text{Precision (P)} - \frac{TP}{TP+FP} \quad (2)$$

4.3 Recall

It can be defined as the number of true positive samples that is taken out from the total number of positive declared patterns.

$$\text{Recall (R)} - \frac{TP}{TP+FN} \quad (3)$$

4.4 F1 Score

Precision and Recall are the building blocks of F1 Score. F1 Score is the combination of precision and recall metrics into a single metric that works well on unbalanced data.

$$\text{F1 Score} - \frac{2*P*R}{P+R} \quad (4)$$

5 Experimental Results

This test aims to determine whether there is a disease in mango plant leaf. We have used several algorithms like Decision Tree, support vector machine, neural network, and logistic regression to classify disease in leaf images.

5.1 Experimental Results Using Logistic Regression (LR)

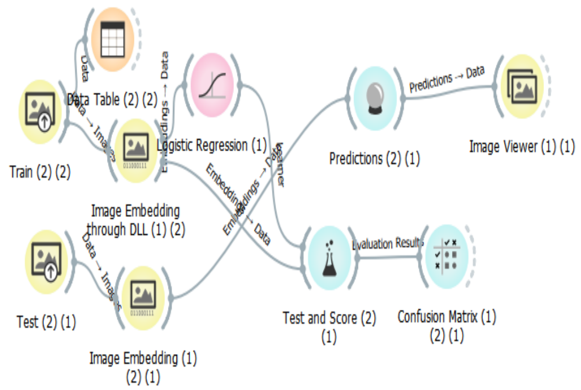

Figure 2: Logistic regression for classification of disease in mango plant leaf

Flow diagram for the classification of disease in mango plant disease is shown in Fig.2.

The typical method for generating prediction models for a binary outcome is logistic regression, extended to illness categorization using microarray data. By implementing logistic regression to classify disease in the mango plant, we can attain a classification accuracy of 94.2%.

5.2 Experimental Results of Decision Tree Classification (DT)

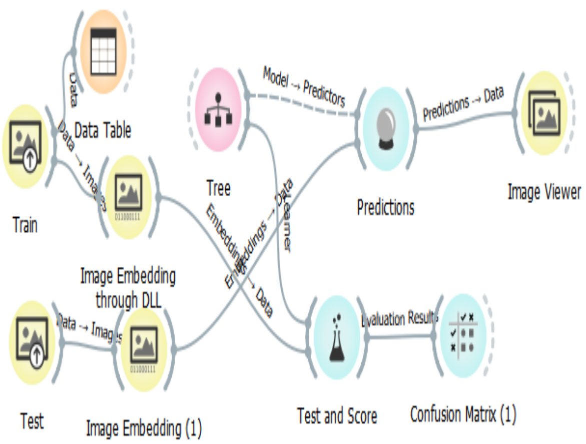

Figure 3: Decision tree for classification of disease in mango plant leaf

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5563
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

Flow diagram for the classification of disease in mango plant disease using Decision Tree is shown in Fig.3.

Decision trees are created using an algorithmic method that allows the information to be split in a variety of ways based on certain conditions. In the field of supervised algorithms, decision trees are the most powerful algorithm. Decision tree analysis is a type of predictive modelling that can be used to a variety of circumstances. Decision trees are created using an algorithmic method that allows the information to be split in a variety of ways based on certain conditions. We can achieve a classification accuracy of 92.3 percent by using a Decision Tree for disease classification in the mango plant, which is lower than that of logistic regression.

5.3 Experimental Results of Support Vector Machine Classification (SVM)

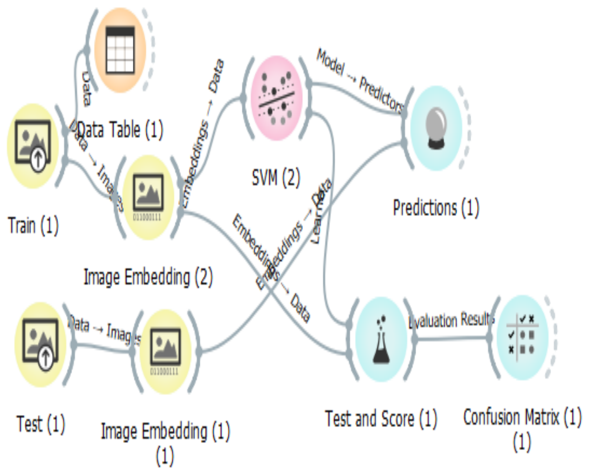

Figure 4: Support vector machine for classification of disease in mango plant leaf

Flow diagram for the classification of disease in mango plant disease using Support Vector Machine is shown in Fig.4.

In multidimensional space, an SVM model represents various classes in a hyperplane. SVM will iteratively create the hyperplane to reduce error. SVMs are widely used because they handle both continuous and categorical variables. Their purpose is to classify datasets to identify a maximum marginal hyperplane. By implementing a Support Vector Machine for the classification of disease in mango plants, we can attain a classification accuracy of 93.3%, which is better than the performance of the Decision Tree.

5.4 Experimental Results of Neural Network Classification (NN)

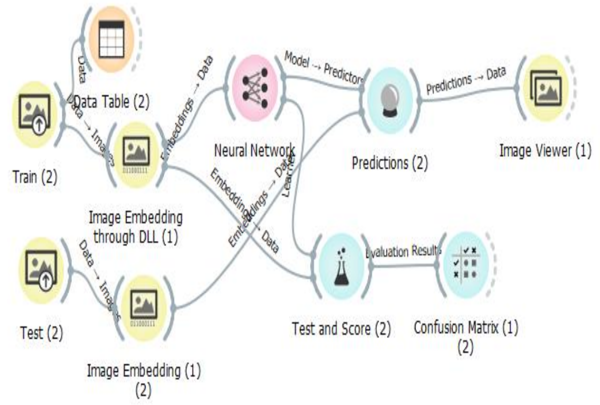

Figure 5: Neural network for classification of disease in Mango plant leaf

Flow diagram for the classification of disease in mango plant disease is shown in Fig.5.

Neural networks are a rough approximation of how the human brain learns. An Artificial Neural Network comprises Neurons, which are responsible for layer creation. Tuned parameters are another name for these Neurons. Each layer's output is passed on to the following layer. Each layer has its nonlinear activation function, which aids in the learning process and the production of the layer. Terminal neurons are another name for the output layer. By implementing Neural Network to classify disease in the mango plant, we can attain a classification accuracy of 94.3%. This model performed relatively better than the other models like logistic regression, support vector machine, and Decision Tree.

5.5 Experimental Results by Implementing Stack Concept

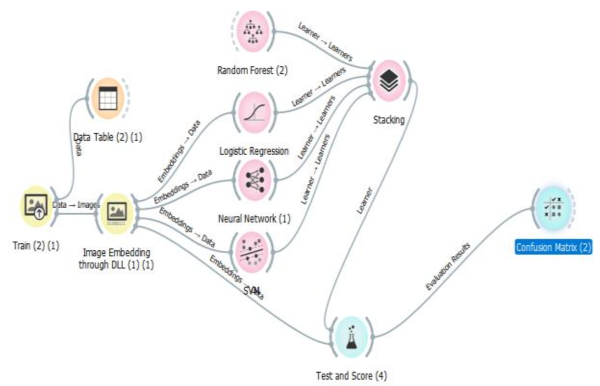

Figure 6: Ensemble technique (Stack) for classification of disease in mango plant leaf

Using stacking of several algorithms, I improved classification accuracy, area of curvature, F1 Score, precision, and Recall. On the whole using stacking of deep learning

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5563
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

improves overall performance in classifying the mango plant leaf image. Flow diagram for the classification of disease in mango plant disease is shown in Fig.6. Their performance metrics are shown in Tab.6. Fig.7 presents a comparison of various classification models.

Table 2: Performance Metrices of algorithms used with their classification accuracy, F1 score, precision and recall

Algorithm used	Area Of Curvature	Classification Accuracy	F1	Precision	Recall
LR	0.988	0.942	0.951	0.966	0.937
DT	0.965	0.923	0.914	0.886	0.943
NN	0.988	0.943	0.952	0.970	0.937
SVM	0.966	0.933	0.928	0.933	0.943
Stack	0.993	0.986	0.986	0.986	0.986

Tab.2. performance metrices of algorithm that were used for comparative study

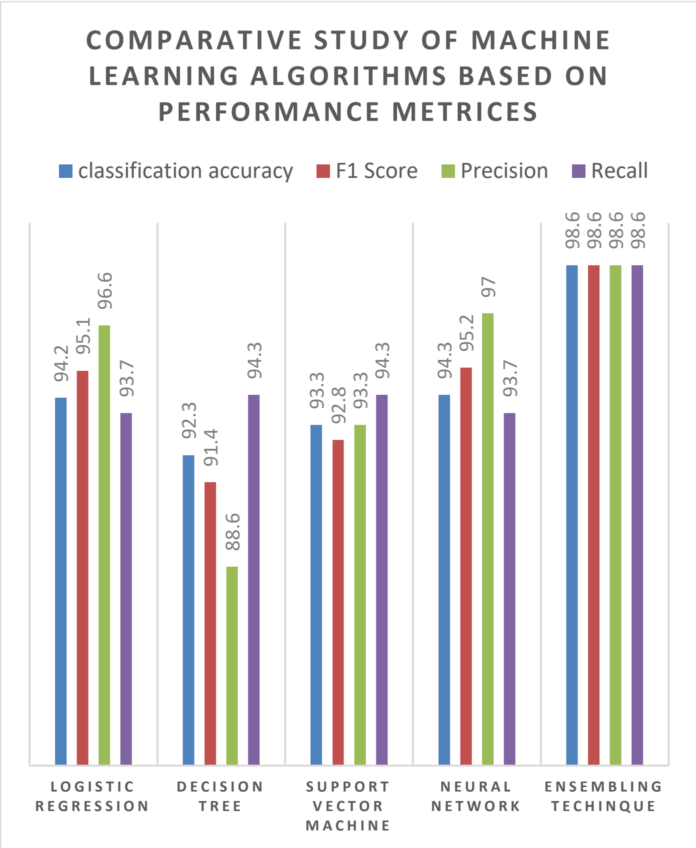

Figure 7: Comparison on machine learning algorithm on the basis of performance metrics

6 Optimization Techniques

Minimizing (or maximizing) any mathematical term is known as optimization. An optimizer is a method for reducing losses by adjusting the parameters of a neural network, such as weights and learning rate. By minimizing a function, optimizers are employed to address optimization problems. Adam (Adaptive Moment Estimation) is the most efficient optimization technique since it is the quickest to reach the minimum and has the lowest training cost. Adam is a momentum-based hybrid of RMSprop and Stochastic Gradient Descent. Adam keeps an exponentially decaying average of past squared gradients and an average of past gradients on his hard drive. The exponential decay rates of these moving averages are governed by hyper-parameters β1, β2 ∈ [0, 1). We compute the decaying averages of past and past squared gradients and respectively as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\, g_t \quad (5)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)\, g_t^2 \quad (6)$$

m_t and v_t are estimates of the first moment (the mean) and the second moment (the uncentered variance) of the gradients respectively.

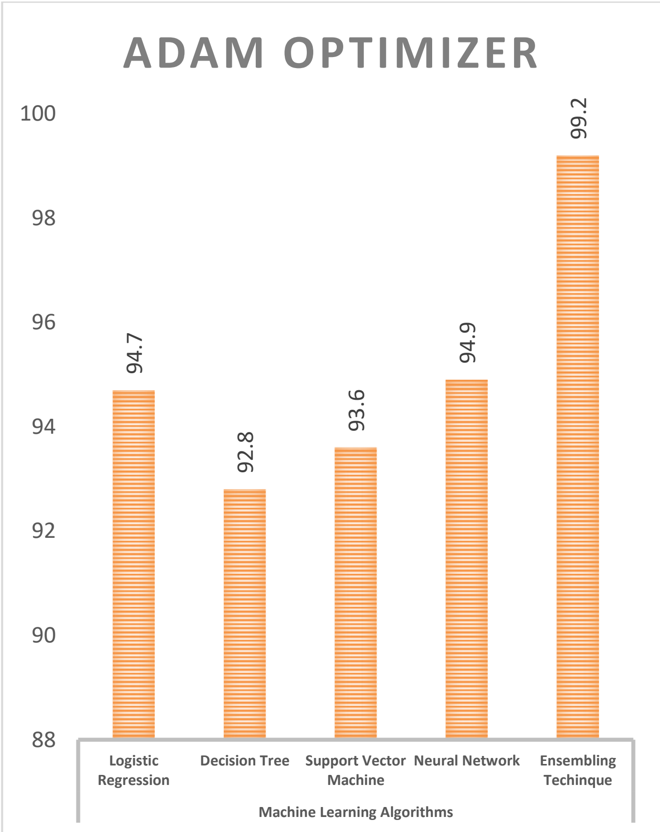

Figure 9: Classification accuracy after applying optimization technique (Adam)

Thus, after applying the stacking concept, we will get a better accuracy in classifying mango plant leaf disease (98.6%), which can be further improved by using the optimization technique (99.2%) as shown in Fig.9. The overall classification accuracy in disease classification can be improved with Adam optimizer. The novelty of this implementation is that we would be able to classify the disease detection in mango plant leaves more accurately, which would help the beneficiary overcome the infection and eradicate it in the preliminary stages. Instead of depending on only the model, we have implemented an ensemble of several algorithms which has more accuracy in classifying our given image. Thereby, we would get a better result in organizing mango plant leaf disease detection. The Ensemble technique has the highest classification accuracy of 98.6%. With Adam optimization, the classification accuracy is further increased to 99.2%. Other than that, different machine learning algorithm classification accuracy can also be increased with Adam Optimizer. Thus, optimization helps to improve the classification accuracy result.

7 Conclusion and Future Work

The outcome of this evaluation is to classify (healthy/diseased) the disease in mango plant leaf images. When implemented, logistic regression, Support vector machine, Neural network, and decision tree came up with excellent classification accuracy because all four algorithms can solve classification problems. But we need an improvement in all aspects to prove that our model is more stable and accurate in classifying the images. So, by using the ensemble technique (Stacking) of all four algorithms, we could come up with better results in classification accuracy, Area of Curvature, F1 Score, precision, and Recall. With this model, the classification accuracy would be improved, which would help the model users to predict and classify the disease in the mango plant leaf (98.6 %). The classification accuracy can be further improved using the optimization technique (Adam optimizer) is 99.2%.

Our subsequent work uses the real-time dataset from the land and develops an outcome as a healthy or unhealthy leaf. This would help the end-users(farmers) classify and solve their problems encountered in the field. If it is unhealthy, the remedy would be given as a solution for the pain they have faced on the farm.

Acknowledgement: The authors wish to thank the management of VIT University, Chennai for their constant motivation and support during the course of this implementation work.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors state that they have no conflicting interests to disclose in relation to this work.

References

[1]. D. Das, M. Singh, S. S. Mohanty, and S. Chakravarty, "Leaf Disease Detection using Support Vector Machine," in *2020 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, Jul. 2020, pp. 1036–1040.

[2]. S. S. Chouhan, A. Kaul, U. P. Singh, and S. Jain, "Bacterial Foraging Optimization Based Radial Basis Function Neural Network (BRBFNN) for Identification and Classification of Plant Leaf Diseases: An Automatic Approach Towards Plant Pathology," IEEE Access, vol. 6, pp. 8852–8863, 2018.

[3]. S. D. Khirade and A. B. Patil, "Plant Disease Detection Using Image Processing," presented at the 2015 *International Conference on Computing, Communication, Control and Automation* (*ICCUBEA),* Pune, Feb. 2015.

[4]. Pawan Kumar Tiwari, Mukesh Kumar Yadav, R. K. G. A. . (2022). Design Simulation and Review of Solar PV Power Forecasting Using Computing Techniques. International Journal on Recent Technologies in Mechanical and Electrical Engineering, 9(5), 18–27. https://doi.org/10.17762/ijrmee.v9i5.370

[5]. G. Madhulatha and O. Ramadevi, "Recognition of Plant Diseases using Convolutional Neural Network," in *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, India, Oct. 2020, pp. 738–743.

[6]. S. S. Hari, M. Sivakumar, P. Renuga, S. karthikeyan, and S. Suriya, "Detection of Plant Disease by Leaf Image Using Convolutional Neural Network," *International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, Vellore, India, Mar. 2019, pp. 1–5.

[7]. U. P. Singh, S. S. Chouhan, S. Jain, and S. Jain, "Multilayer Convolution Neural Network for the Classification of Mango Leaves Infected by Anthracnose Disease," *IEEE Access*, vol. 7, pp. 43721–43729, 2019.

[8]. N. Kumari, A. Kr. Bhatt, R. Kr. Dwivedi, and R. Belwal, "Hybridized approach of image segmentation in classification of fruit mango using BPNN and discriminant analyzer," *Multimed Tools Appl*, vol. 80, no. 4, pp. 4943–4973, Feb. 2021.

[9]. S. Arya and R. Singh, "A Comparative Study of CNN and AlexNet for Detection of Disease in Potato and Mango leaf," in *2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, GHAZIABAD, India, Sep. 2019, pp. 1–6.

[10]. S. Arivazhagan and S. V. Ligi, "Mango Leaf Diseases Identification Using Convolutional Neural Network," *Semantic Scholar* p. 14, 2018.

[11]. T. N. Pham, L. V. Tran, and S. V. T. Dao, "Early Disease Classification of Mango Leaves Using Feed-Forward Neural Network and Hybrid Metaheuristic Feature Selection," *IEEE Access*, vol. 8, pp. 189960–189973, 2020.

[12]. K. Trang, L. TonThat, N. G. Thao, and N. T. T. Thi, "Mango Diseases Identification by a Deep Residual Network with Contrast Enhancement and Transfer Learning," in *2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET)*, Penang, Malaysia, Nov. 2019, pp. 138–142.

[13]. Md. R. Mia, S. Roy, S. K. Das, and Md. A. Rahman, "Mango leaf disease recognition using neural network and support vector machine," *Iran J Comput Sci*, vol. 3, no. 3, pp. 185–

ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5563
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

193, Sep. 2020.

[14]. G. Shrestha, M. Das and N. Dey, "Plant Disease Detection Using CNN," in *2020 IEEE Applied Signal Processing Conference (ASPCON)*, Kolkata, India, Oct. 2020, pp. 109–113.

[15]. M. Kawatra, S. Agarwal, and R. Kapur, "Leaf Disease Detection using Neural Network Hybrid Models," in *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, Greater Noida, India, Oct. 2020, pp. 225–230.