

Automatic Classification of Mango Leaf Disease based on Machine Learning and Deep Learning Techniques

Deepali Joshi¹, Harshali Patil²

¹Computer Engineering

Thakur College of Engineering & Technology
Mumbai, INDIA

deepali.joshi@thakureducation.org

²Computer Engineering

Thakur College of Engineering & Technology
Mumbai, INDIA

harshali.patil@thakureducation.org

Abstract—Fruits are the essential source of nutrition for the human body. The fruit needs to be nurtured and cared for in order to remain healthy. Lack of upkeep, illnesses, blemishes, and fungi result in a considerable loss of produce and profit. One of the important and popular fruit that is consumed worldwide is Mango. It is a fragile fruit and is vulnerable to diseases that affects its quality and quantity. Manual inspection for diseases or infection is a tedious process and requires abundant resources such as time and labour. Manual inspection is inefficient and inaccurate. Automatic inspection on the other hand has numerous benefits. The image classification techniques and algorithms can be used to detect infected and healthy mangoes thus reducing the losses to the farmers.

Keywords- Mango, Mango disease, Mango Leaf, Production, CNN, SVM, Machine Learning, Deep Learning

I. INTRODUCTION

India is the leading manufacturer of various fruits, vegetables, flowers, crops, pulses. One of the popular and most loved fruit by all age groups across the globe is Mango also known as the king of fruits. The entire mango tree has lots of benefits such as Mango leaves are microbial-resistant. It releases abundant amount of oxygen. The mango fruit is powerhouse of minerals, vitamins, fibre, it is gut-friendly and filled in antioxidants. In spite of having all advantages the mango tree and the fruit is very delicate and susceptible to diseases[4]. First of all it is difficult to detect the disease and if the mango tree or the fruit is infected by any disease then the yield is affected and due to this the farmers have to face lot of loss of income. Therefore disease detection is very crucial. The disease detection can be done in manual way or automatic. If the disease detection is done by manual method then it is problematic, tedious, time-consuming. Disease detection is not simple and it requires skilled human labour for doing the task[5]. In order to reduce the problems faced by farmer it is very essential to thus, detect the disease in an automated manner.

Disease detection through automatic way is beneficial because there are different plant varieties and with it there are numerous diseases that affect or infect the plant. The symptoms of different diseases are quite similar and may result

in inaccurate diagnosis. The incorrect diagnosis can have adverse effect on the plant. Therefore having an automated system for disease detection and prediction will help the farmer to increase the yield thereby increasing the profit. The automated systems are capable of generating accurate results in the form of diagnosis, detection and prediction of diseases.

II. LITERATURE SURVEY

The literature survey for plant disease detection and prediction consists of surveying various papers. The authors have applied various algorithms to the datasets. Some of the authors have worked on the real time data whereas some authors have utilised data available from online resources.

Sunanya Arya [4] the authors addresses the issues of fungal diseases namely Anthracnose disease. For detection of disease CNN and Alexnet is used. The dataset used is of mango leaves and were collected from GBPUAT field location.

Uday Pratap Singh [5] the author focuses on developing effective technique for disease diagnosis named as Anthracnose symptoms. The MCNN algorithm is proposed for classifying Mango leaves that have the fungus infected them. The dataset considered is of mango leaves and contains 1070

images of mango leaves. The accuracy of proposed model is 97.13%.

U Sanath Rao [7] has worked on three diseases named as Bacterial canker, Scab, Powdery Mildew. The Alexnet model has been trained and tested to identify if the leaves are infected with diseases or they are healthy ones. The dataset considered is of mango leaves. The proposed model provides accuracy of 89%.

Akshay Koushik [8], they have worked on two diseases such as Anthracnose, Scab. Anthracnose, Scab, and Healthy are the three categories into which the input is divided after being gathered from various sources. The dataset is split into 80%-20% as training and testing data. The technique uses 20 epochs. The algorithm used is CNN to differentiate between healthy mangoes and the unhealthy mangoes. The accuracy of proposed model for 64 batch size is 91.8%, for 32 batch size is 95.6%, for 16 batch size is 94.3%.

Tan Nhat Pham [9] they have worked on diseases such as Anthracnose, Powdery Mildew, Gall Mildew. The algorithms used are VGG, Alexnet, ResNet, ResNet 50. The dataset used for the operation is of mango leaves. The approach was developed using an Adaptive Particle-Grey Wolf metaheuristic (APGWO). The suggested approach outperformed transfer learning-enhanced deep learning models like VGG, AlexNet, and ResNet-50 with respective values 89.41% vs 78.64%, 79.92%, and 84.88%.

Shih-Lun Wu [10], the authors address the issues of grading of mango fruit with CNN algorithm with different models such as Mask R-CNN, AlexNet, VGGs, and ResNets. The dataset consists of mango fruit images.

Pankaj Kumar [12], the authors focus on disease named Anthracnose. The algorithm used is CNN. The dataset considered is of mango leaves and actual dataset was acquired from the farms located in Karnataka, Maharashtra and New Delhi. The algorithm proposed by the authors provides classification accuracy of about 96.16%.

Suwit Wongsila [13], the authors focus on constructing a model that will detect the Anthracnose disease and the main core of the system is CNN algorithm. The dataset used is Mango fruit. The main aim is to isolate the diseased mangoes. The proposed system has accuracy greater than 70%.

Amisha Sharma [14] has focused on diseases such as Anthracnose, Red Rust, Powdery Mildew. The authors have proposed a model using Convolutional Neural Networks for the early identification and categorization of mango leaf

diseases. The dataset considered is of mango leaves. On the supplemented data, a convolutional neural network model has been trained to identify and categorize mango leaf illnesses. Accuracy of the proposed CNN-based model is 90.36%.

Sarder Iftekhhar Ahmed [15] the authors have specified various diseases that infect mango tree with their description in detail. Various datasets are considered such as PlantVillage, DigiPathos, Indian leaves, Indonesian mango leaves. The symptoms for each disease is mentioned along with an image, making it easy to understand. The algorithms used are CNN, CNN-SVM, Resnet 50 and the CNN algorithm provides highest accuracy. The dataset considered is of mango leaves.

Demba Faye [16] has worked on a review paper and the author focuses on CNN algorithm. The dataset consists of Mango leaf, fruit. The solutions are divided into two categories: classical ML based solutions and DL based solutions. The diagnosis are divided in two categories, automatic and manual. The author is providing detailed description about the working and difference between the two methods.

The detailed description and findings of literature review is mentioned in the table given below:

TABLE 1: SUMMARY OF LITERATURE REVIEW

Paper Title and Year	Algorithms	Mango Disease	Findings
"Mango Leaf Diseases Detection using Deep Learning", IJKBCS, June 2022	CNN	Anthracnose, Red Rust, Powdery Mildew.	The proposed CNN-based model achieves an accuracy of 90.36%.
"MangoLeafBD: A Comprehensive Image Dataset to Classify Diseased and Healthy Mango Leaves", Research Gate, Aug 2022, Elsevier	CNN, CNN-SVM, Resnet 50	Anthracnose, Bacterial Canker, Cutting weevil, Die Back, Sooty Mould, Powdery Mildew.	CNN algorithm provides highest accuracy
"Classification of Mango Leaves Infected by Fungal Disease Anthracnose Using Deep Learning", ICCMC 2021, IEEE Xplore	CNN	Anthracnose	The algorithm proposed by the authors provides classification accuracy of about 96.16%.

“Deep Learning Precision Farming: Grapes and Mangoes Leaf Disease Detection by Transfer Learning”, Global Transitions Proceedings 2(2021) KeAi, Chinese Roots Global Impact	Alexnet	Bacterial canker, Scab, Powdery Mildew	The proposed model provides accuracy of 89%.	“Deep Learning for Automatic Quality Grading of Mangoes: Methods and Insights”, 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)	CNN	Grading of mango Fruit	The authors address the issues of grading of mango fruit
“Detection and Classification of Diseased Mangoes”, Research Gate, November 2020,	CNN	Anthracoise , Scab	The accuracy of proposed model for 64 batch size is 91.8%, for 32 batch size is 95.6%, for 16 batch size is 94.3%.	“A Comparative study of CNN and Alexnet for Detection of Disease in Potato and Mango Leaf”, 2019 2 nd International Conference on ICICT, IEEE	CNN, Alexnet	Anthracoise	the authors address the issues of anthracnose disease.
“Multilayer Convolutional Neural Network for the Classification of Mango Leaves Infected by Anthracnose Disease”, IEEE Access, March 2019	CNN, MCNN	Anthracoise	The accuracy of proposed model is 97.13% .	“Hybrid approach for Anthracnose detection using intensity and size features”, 2016 IEEE ICT for Agriculture and Rural Development,	Segmentation and Thresholding technique	Anthracoise	Detection of affected area using proposed hybrid approach is 56.22 for Apple, 62.11 for Mango and 58.67 for Tomato.
“Early Disease Classification of Mango Leaves using Feed-Forward Neural Network and Hybrid Metaheuristic Feature Selection”, IEEE Access, October 12,2020	ANN, Alexnet	Anthracoise	transfer learning with respective values 89.41% vs 78.64%, 79.92%, and 84.88%.	<h3>III. PROPOSED METHODOLOGY</h3> <h4>A. Steps</h4> <ol style="list-style-type: none"> 1) Collection of real-time data consisting of leaves of two classes that is mango leaves collected, both unhealthy and infected from Kaggle[18]. 2) Pre-processing of images from the dataset which consists of image cropping and contrast boosting using the Histogram equalization technique. 3) Labelling the images. 4) Splitting dataset in training, testing dataset. 5) Train algorithms with training dataset and testing the algorithms with testing dataset. 6) In the evaluation step the results will be generated. 			
“Mango Diseases Classification Solutions Using Machine Learning or Deep Learning: A Review”, Journal of Computer and Communications, 2022,10,16-28, Scientific Research Publishing	CNN	Describes about Mango diseases.	The author is providing detailed description about the algorithm				
“Machine Learning Algorithm Development for detection of Mango infected by Anthracnose Disease”, 978-1-6654-1569-9/21/\$31.00 ©2021 IEEE, The 6th International Conference on Digital Arts, Media and Technology (DAMT)	CNN	Anthracoise	The proposed system has accuracy greater than 70%.				

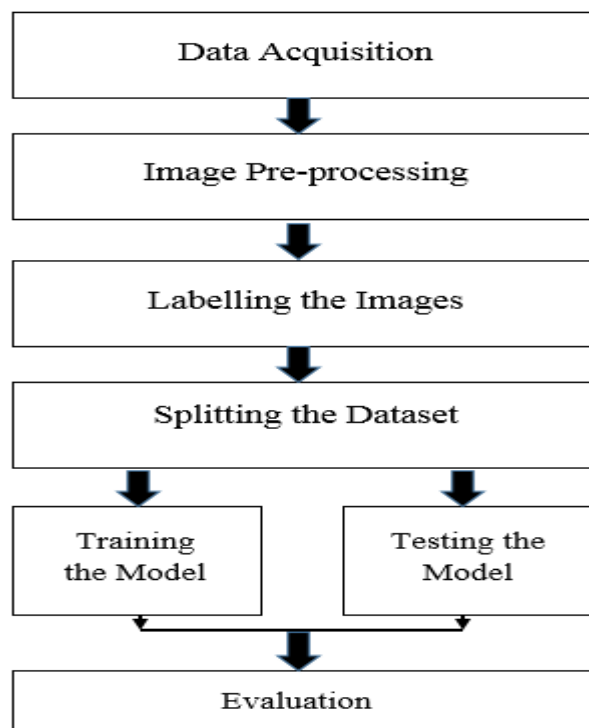


Figure 1. Steps

B. Proposed Model

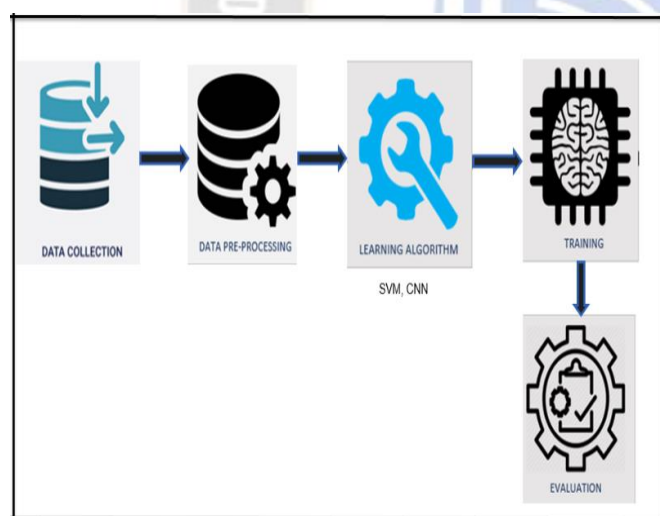


Figure 2. Diagram of Proposed Model

1) Data Collection

The dataset is acquired from kaggle and contains data in the form of image. There are two types of images mainly healthy leaf images and non-healthy leaf images. There are eight classes named as Anthracnose, Cutting Weevil, Bacterial Canker, Die Back, Gall Midge, Powdery Mildew, Sooty Mould and healthy. Each of the category contains 500 images[18].

2)Data Pre-processing

It indicates a process of preparing raw data and transform it into a information that is suitable for learning model. It is a crucial step in creation of learning model.

3)Learning Algorithms

The learning algorithms used for the proposed method is SVM and CNN for detection of mango leaf disease.

4)Training

Using a dataset of 3500 images, 80% were used for training and 20% were used for validation. The following Fig. 2 depicts samples of images from the dataset.



Figure 3. Anthracnose



Figure 4. Healthy Leaf

5) Evaluation

The evaluation step will include testing the data with test dataset and applying the algorithms for generating the results.

C. Convolution Neural Network

Convolutional Neural Network is one of the popular and widely used algorithms for image recognition. The areas where CNN are commonly sued are Scene labelling, objects detections, and face recognition. The input provided to CNN is an image which will be classified and processsed under certain category. If images of leaves are considered then the category can be healthy or infected leaf. The CNN algorithm is made up of various layers and each input image will pass through a sequence of layers. Once the image passes through the layers then classification of the object with probabilistic values 0 and 1 will be done.

1) CNN layers

A CNN algorithm consists various layers which are given below:

1. Convolutional Layer
2. Pooling layer
3. Fully connected (FC) layer.

1. Convolutional layer is the first layer of CNN, and fully connected layer is the final layer. The complexity of the CNN increases from first layer to last layer. The increasing complexity of CNN allows it to identify larger portions of image along with complex features until the object is identified entirely.

1. Convolutional layer. The major number of computations take place in the convolutional layer. The core building block of a CNN algorithm is the convolutional layer. The convolutional layer consists of filter or kernel which is inside it and which moves across the receptive fields of the image in order to check if the feature is present in the given image or not.

The kernel runs through the entire given image over the course of many iterations. Every iteration ends with the calculation of a dot product. Dot product calculations are made between the filter and the input pixels. The result of the dots is a feature map, also referred to as a convolved feature. The CNN algorithm can interpret and extract important patterns from the image thanks to this layer's conversion of the image into numerical values.

2. Pooling layer. A kernel or filter is also swept across the input image by the second layer, which is referred to as the pooling layer. The pooling layer will decrease the number of input parameters and could lead to information loss. On the other hand, the pooling layer will simplify the CNN algorithm and contribute to its increased effectiveness.

3. Fully connected layer. The next layer is known as Fully Connected Layer and this is where the image classification occurs in the CNN based on the features that are extracted from the previous layers. The interpretation of this layer indicates that the inputs from one layer will be connected to every activation unit of the layer which is next.

Since a dense network would result, which is not what is needed, the CNN layers are not fully connected. The dense network would increase losses, expenses and may affect the quality of the output.

2) Advantages of CNN

1. It provides efficient image processing
2. It has higher accuracy rates
3. It is Robust to noise
4. Transfer learning
5. It has Automated feature extraction

3) Disadvantages of CNN

1. It requires High computations
2. Not efficient for small datasets
3. To provide higher accuracy CNN requires large datasets.
4. It has limited ability to generalize

D. SVM

Support Vector Machines belong to the category of supervised machine learning algorithm. SVM perform classification as well as regression tasks.

An image is perceived as two-dimensional array of pixels when a computer processes it. The array size will correspond to the resolution of image.

Let us assume that the array will have the following dimensions if the image is 100 pixels wide and 100 pixels tall: 100 x 100 x 3. In this illustration, the image's width and height will be represented by the first two dimensions while the RGB color channels will be represented by the third dimension. The array's values will run the gamut from 0 to 255. The intensity of each pixel at each point is shown by this range.

For image classification using SVM there are various steps to be followed such as

1. Extract features from the image. The features can be color values of the pixels, edge detection, or the textures present in the image.
2. The extracted features are used as input for the SVM algorithm.

1) Types of SVM

SVM can be of two types:

1. Linear SVM
2. Non-linear SVM

1. When data can be separated linearly, the linear SVM is used. A dataset is said to be linearly separable if it can be

divided into two classes using only a single straight line, and the classifier used for this type of data is known as a Linear SVM classifier.

2. It is used for data that is non-linearly separated, which means that if a dataset cannot be classified using a straight line, it is considered to be non-linear data.

2) Advantages of Support Vector Machine:

1. It works well when there is proper separation between classes.
2. It works well in situations where there are more dimensions than samples.
3. It can handle high-dimensional data for example images.
4. They are less prone to overfitting.

3) Disadvantages of Support Vector Machine:

1. SVM is not suitable for large data.
2. It cannot handle noise

IV. RESULTS

For Implementation of the proposed model dataset from kaggle is used named as “Mango Leaf Disease Dataset”. It contains around 4000 images of mango leaf. The number of classes are eight, consisting of healthy leaves along with seven diseased leaves such as Anthracnose, Bacterial Canker, Cutting Weevil, Die Back, Gall Midge, Powdery Mildew, Sooty Mould. The size of dataset is around 108MB. Jupyter notebook is used for implementation of algorithms such as SVM, CNN. The results for algorithms are given below:

1). SVM

The results for SVM algorithm is indicated by the figure given below. Various parameters are considered such as precision, recall, f1-score and support. The SVM algorithm provides accuracy of 82% by applying it on the leaf dataset.

Classification Report

	precision	recall	f1-score	support
0	0.73	0.76	0.75	102
1	0.99	0.98	0.99	117
2	0.95	0.96	0.96	101
3	0.74	0.90	0.81	87
4	0.62	0.61	0.61	99
5	0.56	0.49	0.53	91
6	0.91	0.84	0.88	96
7	0.97	0.93	0.95	107
accuracy			0.82	800
macro avg	0.81	0.81	0.81	800
weighted avg	0.82	0.82	0.82	800

Figure 5. Classification Report of SVM

The graphs for macro average and weighted average accuracy are given below. It can be observed that the macro avg value for precision, recall, f1-score is 0.81 and the weighted avg value for precision, recall, f1-score is 0.82.

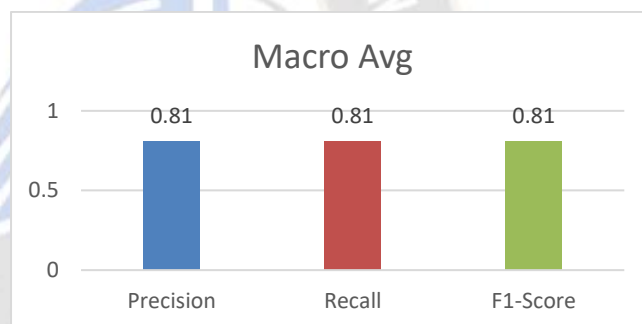


Figure 6. Macro Avg value for SVM

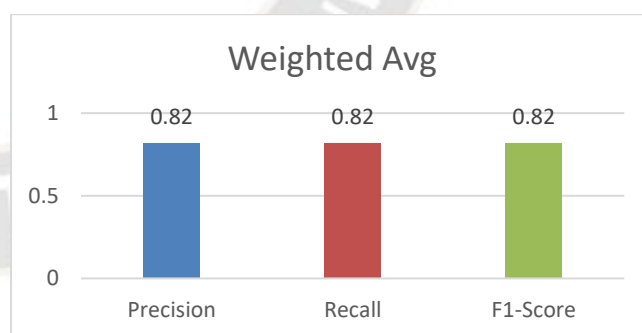


Figure 7. Weighted Avg value for SVM

2) CNN

The results for CNN algorithm is indicated by the figure given below. Various parameters are considered such as precision, recall, f1-score and support. The accuracy of the CNN algorithm is 95% by applying it on the leaf dataset.

Classification Report

	precision	recall	f1-score	support
0	0.94	0.92	0.93	102
1	1.00	1.00	1.00	117
2	1.00	0.95	0.97	101
3	0.99	0.86	0.92	87
4	0.89	0.93	0.91	99
5	0.84	0.99	0.91	91
6	0.99	0.96	0.97	96
7	0.98	0.99	0.99	107
accuracy			0.95	800
macro avg	0.95	0.95	0.95	800
weighted avg	0.96	0.95	0.95	800

Figure 8. Classification Report of CNN

The graphs for macro average and weighted average accuracy are given below. It can be observed that the macro avg value for precision, recall, f1-score is 0.95 and the weighted avg value for precision is 0.96. For recall and f1-score the weighted avg value is 0.95.

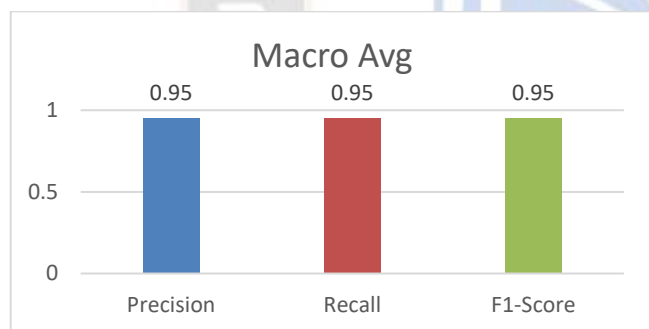


Figure 9. Macro Avg value for CNN

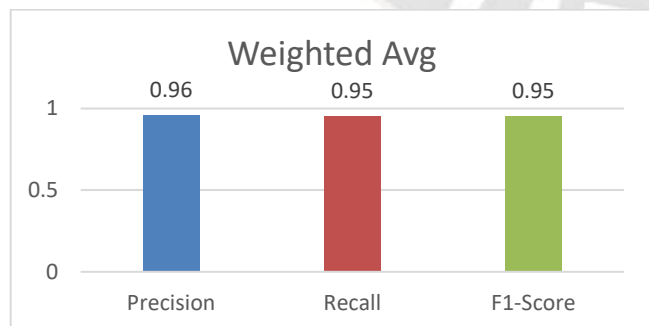


Figure 10. Weighted Avg value for CNN

The table given below depicts the accuracy for SVM, CNN algorithms.

TABLE 2: RESULT OF ALGORITHM

Sr.No	Algorithm	Accuracy %
1	SVM	82%
2	CNN	95%

The results obtained from SVM and CNN algorithms can be visualized from the graph given below:

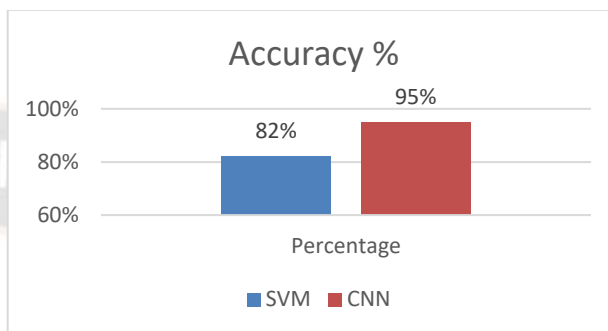


Figure 11. Accuracy % of Algorithms

V. CONCLUSION

In this paper dataset consisting of mango leaves is considered. There are eight classes are consisting of healthy leaves along with seven diseased leaves such as Anthracnose, Bacterial Canker, Cutting Weevil, Die Back, Gall Midge, Powdery Mildew, and Sooty Mould. The algorithms such as SVM, CNN are implemented to detect the leaves infected by diseases. The percentage of training and testing dataset considered is 80%-20%. The SVM provides 82% accuracy, CNN provides 95% accuracy. As seen through the results that the accuracy provided by the CNN algorithm is the highest that is 95%.

REFERENCES

- [1] Peter Flach, "Machine Learning: The Art and Science of Algorithms that Make Sense of Data", Cambridge University Press, 20-Sept-2012
- [2] K.Swetha, Veena V, Sadhana.G.D, R Priyatharshini, "Hybrid approach for Anthracnose detection using intensity and size features", 2016 IEEE International Conference on Technological Innovations in ICT for Agriculture and Rural Development, 978-1-5090-0615-1/16
- [3] Andreas C. Müller, Sarah Guido, "Introduction to Machine Learning with Python", O'Reilly Media, Inc., September 2016, ISBN: 9781449369897
- [4] Sunanya Arya, Rajeev Singh, "A Comparative study of CNN and Alexnet for Detection of Disease in Potato and Mango Leaf", 2019 2nd International Conference on ICICT, 978-1-7281-1772-0 2019 IEEE
- [5] Uday Pratap Singh, Siddharth Singh Chouhan, Sukirty Jain, Sanjeev Jain, "Multilayer Convolutional Neural Network for the Classification of Mango Leaves Infected by Anthracnose

- Disease", IEEE Access, Volume 07, DOI:10.1109/ACCESS.2019.2907383, March 2019
- [6] P Rama Koteswara Rao, K Swathi, "MANGO PLANT DISEASE DETECTION USING MODIFIED MULTI SUPPORT VECTOR MACHINE ALGORITHM", PJAEE, 17 (7) 2020
- [7] Sanath Rao, U.,Swathi, R.,Sanjana, V.,Arpitha, L.,Chandrasekhar, K., Chinmayi,Naik, Pramod Kumar, "Deep Learning Precision Farming: Grapes and Mangoes Leaf Disease Detection by Transfer Learning", Global Transitions Proceedings 2(2021) 535-544, KeAi, Chinese Roots Global Impact
- [8] Akshay Koushik, Ritanya B Bharadwaj, Ram Prasad, Sana Habeeb,"Detection and Classification of Diseased Mangoes", ICOSICA, Research Gate, November 2020, DOI: 10.1109/ICOSICA49951.2020.9243277
- [9] Tan Nhat Pham , 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, October 12,2020, DOI:10.1109/ACCESS.2020.3031914
- [10] Shih-Lun Wu, Hsiao-Yen Tung, and Yu-Lun Hsu," Deep Learning for Automatic Quality Grading of Mangoes: Methods and Insights", 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), 978-1-7281-8470-8/20/\$31.00 ©2020 IEEE DOI 10.1109/ICMLA51294.2020.00076
- [11] BALASUBRAMANYAM MEDA, S KALYAN, "EARLY DISEASE CLASSIFICATION OF MANGO LEAVES USINGFEED- FORWARD NEURAL NETWORK MODEL", IJTE, Volume XIII, Issue III, 2021
- [12] Pankaj Kumar,Sunidhi Ashtekar,S. S Jayakrishna, K P Bharath, P. T Vanathi, M Rajesh Kumar,"Classification of Mango Leaves Infected by Fungal Disease Anthracnose Using Deep Learning:, ICCMC 2021, IEEE Xplore Part No: CFP21K25-ART, DOI:10.1109/ICCMC51019.2021.9418383
- [13] Suwit Wongsila, Parinya Chantrasri , Pradom Sureephong,"Machine Learning Algorithm Development for detection of Mango infected by Anthracnose Disease", The 6th International Conference on Digital Arts, Media and Technology (DAMT) and 4th ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (NCON), 978-1-6654-1569-9/21/\$31.00 ©2021 IEEE
- [14] Amisha Sharma, Rajneet Kaur Bijral , Jatinder Manhas, Vinod Sharma, "Mango Leaf Diseases Detection using Deep Learning", International Journal of Knowledge Based Computer Systems, IJKBCS, Volume 10, Issue 1, June 2022, 40-44
- [15] Sarder Iftexhar Ahmed ,Muhammad Ibrahim ,Md Nadim, Md Mizanur Rahman, Maria Mehjabin Shejunti , Taskeed Jabid , Md Sawkat Ali , "MangoLeafBD: A Comprehensive Image Dataset to Classify Diseased and Healthy Mango Leaves", Research Gate, Aug 2022, Elsevier
- [16] Dembe Faye, Idy Diop, Doudou Dione,"Mango Diseases Classification Solutions Using Machine Learning or Deep Learning: A Review", Journal of Computer and Communications, 2022,10,16-28, Scientific Research Publishing, ISSN Online: 2327-5227, ISSN Print: 2327-5219
- [17] Israa Mohammed Hassoon, "Classification and Diseases Identification of Mango Based on Artificial Intelligence: A Review", Journal of Al-Qadisiyah for Computer Science and Mathematics Vol. 14(4) 2022 , pp Comp. 39–52
- [18] Kaggle Dataset
- [19] Asha Rani K P, Gowrishankar S "ConvNeXt-based Mango Leaf Disease Detection: Differentiating Pathogens and Pests for Improved Accuracy", International Journal of Advanced Computer Science and Applications(IJACSA), Volume 14 Issue 6, 2023.
- [20] Sapna B Kulkarni, J Keerthi, "Mango Leaf Disease Detection Using Deep Learning", IJRASET, ISSN: 2321-9653, Volume 11 Issue VIII Aug 2023