International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Classification of Healthy and Rot Leaves of Apple Using Gradient Boosting and Support Vector Classifier

K.R.N.V.V.D.Aravind, S.Prayla Shyry, Yovan Felix

Abstract: Conventional Techniques Such As Convolutional Neural Network (Cnn), Deep Neural Network Have Shown Its Own Footprints In The Field Of Image Classification With Promising Results. In The Past Decades, Classification Of Images Has Been Done With Varying Features Like Shape, Texture Etc. In This Paper, A Novel Approach Is Used To Classify The Leaf Images And Determine The Health And The Diseased Leaf. The Image Is Preprocessed By Extracting The Shape Feature And Classified The Leaves Of Apple As Healthy And Diseased (Rot Leaves) Using Two Novel Effective Approaches Gradient Boosting And Support Vector Classifier. We Have Collected 1813 Images Of Apple Leaves As Dataset And Out Of These, 70% Of The Data Is Used To Train And Remaining 30% Is Used To Test The Data. Our Algorithm Has Outperformed Other Traditional Techniques With Good Scale Of Accuracy(Gradient Boosting-87%, Support Vector Classifier-91%). Strong Comparison Of Both Gradient Boosting And Support Vector Is Made And There Is Dominant Show Off Of The Confusion Matrix. Classification Of Healthy And Diseased Leaf Well In Advance Gives Nice Warning To The Producer Thereby Decreasing The Rate Of Diseased.

Keywords: Gradient Boosting, Support Vector Machine, accuracy, classification, confusion matrix

I. INTRODUCTION

Apple Scab, Fire Blight, Cork Spot, Powdery Mildew, Black Rot, Frog Eye Leaf Spot, Crown Rot etc are the most common apple leaf diseases. Early detection and diagnosis of apple leaf diseases can control the spread of infection and helps in the healthy development of the apple industry. The existing research uses complex image pre-processing techniques like cnn didn't gave good results in identification of apple leaf diseases. This paper proposes an accurate identification of apple leaf diseases based on gradient boosting and support vector classifier. It includes pre-processing of image after identification of global features like color, texture and geometrical features like shape. Using a dataset of 1813 images of both diseased and healthy apple leaves, in that 70% of the data is used for training and 30% is used for testing the sample data for identification of leaf diseases in apple leaves.

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author
K.R.N.V.V.D.Aravind, Student, Dept of CSE, Sathyabama Institute of Science and technology, Chennai, India
S.Prayla Shyry, Professor, Dept of CSE, Sathyabama Institute of Science and technology, Chennai, India
Yovan Felix, Asst. Professor, Dept of CSE, Sathyabama Institute of Science and technology, Chennai, India

The proposed model is trained to identify the accuracy scores of apple leaves and draw the confusion matrix for both gradient boosting and support vector classifier.
A strong comparison is made between both gradient boosting and support vector classifier to know the best algorithm which can relay efficient results. The proposed disease identification approach achieves an overall accuracy of 87% for gradient boosting, 91% for support vector classifier.

II. STATE OF THE ART

BinLiu et.al (2017) proposed a novel methodology to identify the apple leaf diseases accurately. They have generated pathological images of apple leaf disease and they have found the results for four types of leaf diseases. They have evaluated with 13,689 leaf images and identified the disease at the early stage in order to avoid the spread of infection. After the pre-processing of images, author has proposed an AlexNet model based novel architecture of deep cnn to remove partial and full connected layers. They have achieved the 97.62% of accuracy recognition rate and this result is increased by 10.83% of accuracy compared to the convolutional models. In future we can implement FasterRCNN (Regions with CNN), YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector) algorithms.
Konstantinos et.al. (2018) used CNN to classify the plant leaf disease using both diseased and healthy plant. They have tested with AlexNet, AlexNetOwtBn, GoogleNet, Overfeat and VGG using Torch7. They have compared with original image and pre-processed image for 5 different CNN architectures. They have found that the accuracy rate is 99.44% for AlexNetOwtBn and it is much better when compared to all the 4 architectures. They have also proved that CNN is very much suitable for the early stage detection of plant diseases.
Serawork Wallelign et.al. (2018) describes the feasibility of CNN for plant disease classification of leaf images with 99.32% accuracy.
They have designed a model based on LeNet architecture to perform Soybean plant disease classification. They have taken 12,673 sample images. In which 70% is used for training, 10% for validation and 20% for testing. They have proposed three convolutional layers of which each is followed by max pooling layer and then ReLu activation function is applied to get output of every convolutional layer and fully connected layer. It is used to filter size, kernel size, Learning parameter were selected by trial and error method.

They have compared and analysed the true positive, false negative, true negative, false positive and classified the healthy and 4 types of diseased leaves.

Manuel Cortes et.al. (2017) have taken 84,147 images of diseased and healthy plants to classify crop species and disease status of 57 different classes using deep convolutional network and semi supervised methods. They have used GLCM functions, Lacunarity and shen features to characterize texture of image by calculating specific values and spatial relationship that occur in an image for pairs of pixel. Using HSV features and SVM, They have also proposed a novel neural network architecture to classify the infected leaf and healthy leaf. After classification they have implemented genetic algorithm to optimize SVM loss and identify the disease type in the infected leaves. Their combination of this algorithm never works in this research field However for unstructured data the performance is imported to 12%.

Srdjan Sladojevic et.al. (2016) have proposed a novel methodology to facilitate a quick and easy implementation of plant disease recognition to recognize 13 different types of plant diseases. They have performed Deep CNN training using a Deep Learning framework named Caffe and achieved an accuracy of 96.3%. After fine tuning of parameters in 100th training. The images taken from the dataset has undergone 4 layer filtering technique using CNN. The proposed model can automatically classify and detect 13 different plant diseases using 3,000 original images later extended up to 30,000 images using appropriate transformation.

Vipinadas.M.J et.al. (2016) have proposed a novel method for image pattern classification in banana leaf using SVM and ANFIS classifier. They have graded banana leaves using ANFIS and compared the performance of SVM and ANFIS classifiers. Using multiple levels SVM they have classified the diseases as blacksigatoka and panamawait in banana leaves. In this classification process first RGB color image is converted into Ycbcr color space and grey scale image.

Adaptive contrast map is applied to detect leaf diseased portions. To obtain white pixels in the diseased area and black pixels in normal portions they have applied dilation and filing holes operations by setting the threshold value to 0.18 after converting to binary image.

For classifying diseased and healthy leaves 50 video samples of both healthy and diseased leaves are taken from Mohandas college banana farm and generated image frames using video reader function in Mat lab 2013a. They have achieved 100% accuracy using ANFIS and 92% accuracy using SVM.

III. MATERIALS AND METHODS

We have taken plant village dataset that consist of 1813 images of healthy and rot leaves of apple. We have used sklearn to load images for our project. For simplicity of calculation, we have resized the image size to 190x190 pixels and stored in the form of numpy array.

Out of these, we have taken 1269 Images (70%) for training the data and 544 images (30%) for testing respectively. Initially the leaf images are pre-processed to remove the unwanted noise by notch filter and then both the general visual features and domain related features are extracted.

The extracted features are then fed as the input to the classifiers namely gradient boosting and support vector classifier.

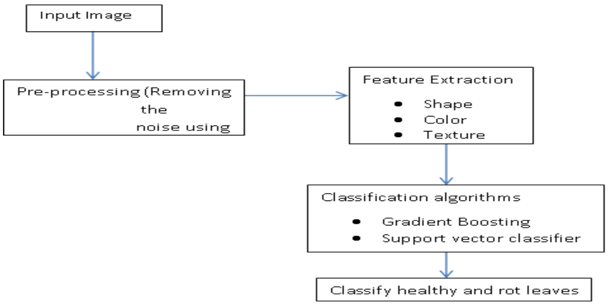

Figure 1.General Block Diagram of Leaf Classification

1. Preprocessing:

The visual quality of the image is increased by reducing or removing the noise that creates unwanted patterns in the image using notch filter. Periodic noises are unwished and spurious signals that create repetitive pattern on images and decrease the visual quality three circular shape notch filters are used to reduce frequent domain filtering with an appropriate radius to enclose complete noise spikes in the Fourier domain. Notch filter rejects frequencies in predefined neighbourhoods around a center frequency because number of notch filters and shape of notch areas is arbitrary.

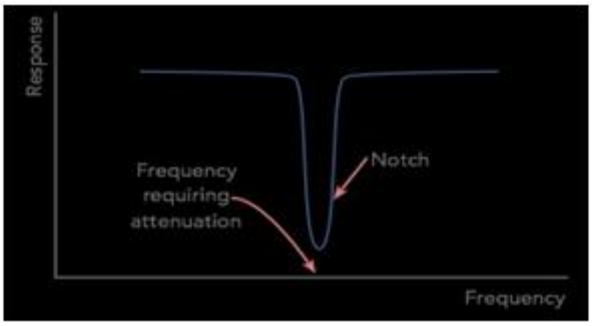

Figure: 2 Noise removal using Notch Filter

The noise in the image is attenuated and the image is retrieved without any distortion. In order to smooth the image 0-1 normalization is performed with the pixels. This normalization process scales up the individual samples as single framework. Also it removes amplitude variation of data that leads to poor ratio of accuracy. The correlation between the pixels is normalised using 0-1 normalization. Formulae used to normalize z [th]value,

$$z_i = (z-\min(z))/(\max(z)-\min(z)) \qquad \text{- Equation (1)}$$

Where z_i is the normalized data, z is the real value, min (z) is the minimum value of z, max of z is the maximum value in that row.

2. Feature Extraction

In the proposed system, global features like color, texture and geometrical features like shape are extracted to identify the healthy and diseased leaves.

The length, width and leaf area is calculated. The length and width of the leaf is calculated using the Euclidean distance and their formulae is, the **distance** between two points in any with coordinates (c, d) and (a, b) is given by,

dist ((c, d), (a, b)) = $\sqrt{(c - a)^2 + (c - b)^2}$ - Equation (2)

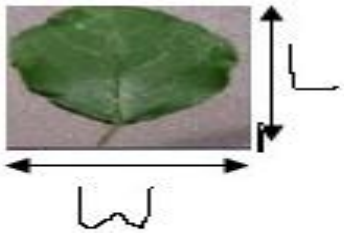

Figure 3: Length, Width calculation of single image (L-Length, W-Width)

Leaf area estimation plays vital role in understanding the process of flowering, fruit set, crop growth, yield, and quality and supply demand chain in global market. The product of Length and Width dimensions is efficient to estimate the Leaf Area and it is revealed as the best of Malus domestica based on the single dimension (L, W, L^2 or W^2). Therefore, the linear regression "Leaf Area, LA=0.412+0.573WL" provided the most accurate estimate of the leaf area.

Grey Level Co-occurrence Matrice (GLCM) measures the combinations of pixel brightness values (grey levels) occur in an image and it is for a series of "second order" texture calculations. The matrices are designed to measure the spatial relationships between pixels. The GLCM method is very sensitive for the any changes in the images such rotation, scale and etc there is a change in extracted features in different neighbourhood degrees. The computation time for GLCM method is less and recognition rate of GLCM is very fast.

3. Classification

In the proposed system we have classified healthy and rot leaves of apple dataset taken from plant village dataset using gradient boosting and support vector classifier by taking global features like color and shape. After classification we have calculated and compared the accuracy scores of both gradient boosting and support vector classifier and drawn accuracy graphs for each of the algorithm with confusion matrix.

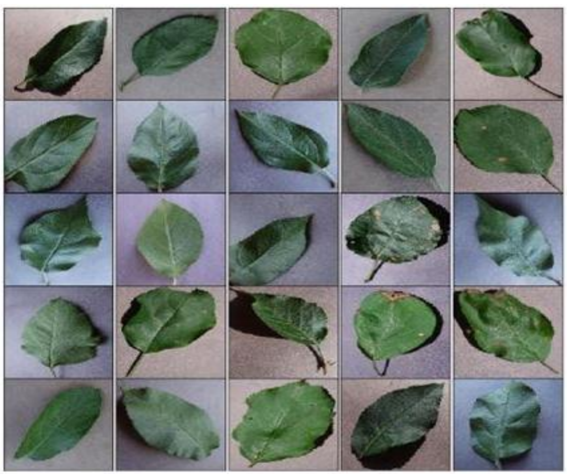

Figure 4: Sample of Input images

4. Accuracy recognition

In this proposed system, Our system is able to classify the leaves with around 85 percent accuracy of the apple leaf dataset and algorithms have performed well and gave the best results of accuracy scores (gradient boosing-87%, support vector classifier-91%) based on global features like shape, texture for classification of healthy and rot leaves in apple leaf dataset.

The leaf area is calculated using the linear regression model and the Grey Level Co-occurrence Matrice is used to extract the texture features. The accuracy scores of both gradient boosting and support vector classifier are calculated and compared each other and it is revealed that support vector machine identifies the rotten leaves much better when compared to gradient boosting method.

IV.RESULTS AND DISCUSSION

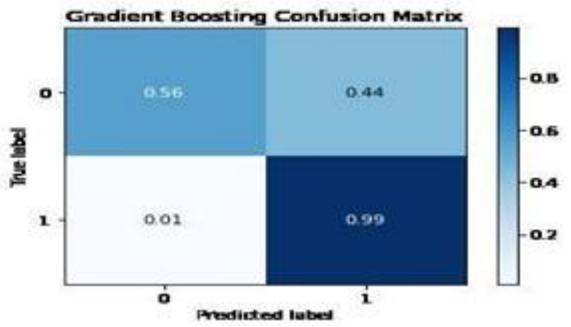

Figure 5: Confusion matrix of gradient boosting algorithm

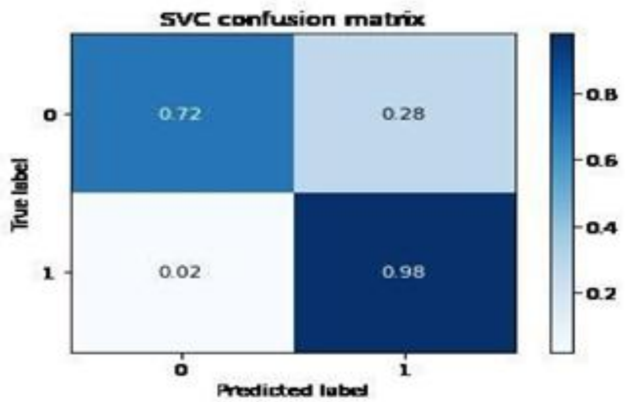

Figure 6: Confusion matrix of support vector classifier

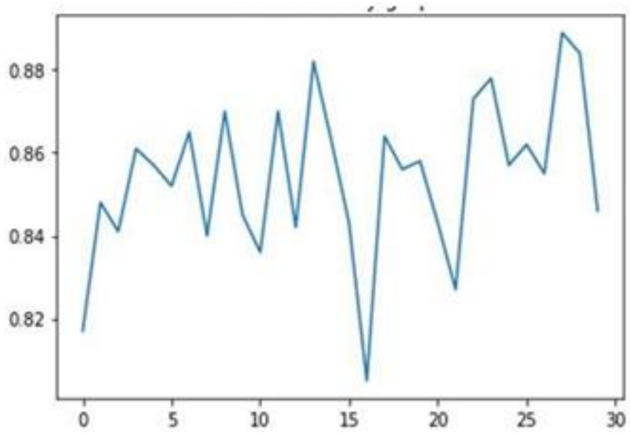

Figure 7: Accuracy rate of gradient boosting

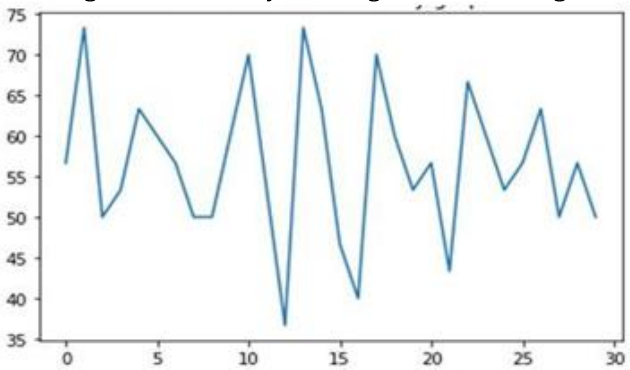

Figure 8: Accuracy rate of support vector classifier

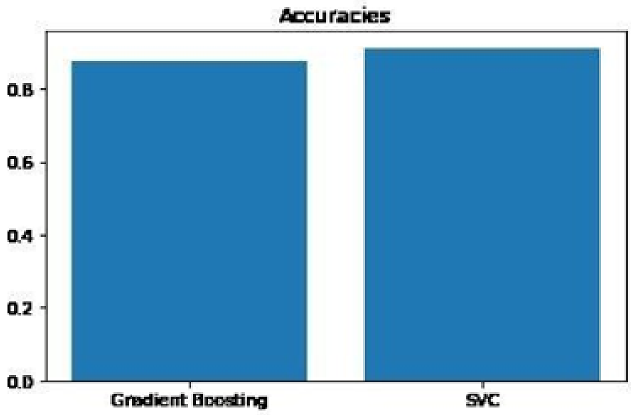

Figure 9: Accuracy rate of gradient boosting and support vector classifier

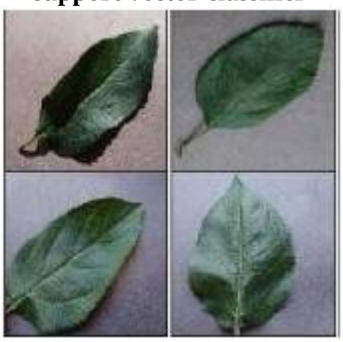

Figure 10: Sample healthy leaves

Figure 11: Sample Rot leaves

V. CONCLUSION:

We have developed a novel approach to classify the health and the rot leaves of apple using gradient boosting and support vector. Our system is able to classify the leaves with around 85 percent accuracy. This is quite reasonable for the amount of data and hardware we have. Our dataset consist of 1813 images of apple healthy and apple rot leaves. After importing the images , the images are resized to 190x190 pixels and in the form of numpy array. Then we split the images into training and testing sets each of which consist of 1269 and 544 images respectively. As we observe the above accuracies the SVC is performing better than gradient boosting. It is observed that the as the size of the dataset increases, the performance of the gradient boosting is also increased. Our approach has outperformed traditional techniques with good scale of accuracy(Gradient Boosting-87%, Support Vector Classifier-91%).The work can be extended with different algorithms like CNN's, combining multiple machine learning models or using ensemble techniques to classify the leaves. This will most likely improve the performance of our model and help's in achieving better accuracy.

REFERENCES

1. KonstantinosP.Ferentinos, Deep learning models for plant disease detection and diagnosis, Computers and Electronics in Agriculture, https://www.sciencedirect.com Volume 145, February 2018, Pages 311-318.
2. Serawork Wallelign, Mihai Polceanu, Cedric Buche, Soybean Plant Disease Identification Using Convolutional Neural Network, The Thirty-First International Florida Artificial Intelligence Research Society Conference (FLAIRS-31).
3. Emanuel Cortes, Plant Disease Classification Using Convolutional Networks and Generative Adverserial Networks, http://cs231n.stanford.edu/reports/2017/pdfs/325.
4. Bin Liu 1,2,* ,† ID , Yun Zhang 1,†, DongJian He 2,3 and Yuxiang Li 4, Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks, Symmetry 2018, 10, 11; doi:10.3390/sym10010011.
5. Srdjan Sladojevic,1 Marko Arsenovic,1 Andras Anderla,1 Dubravko Culibrk,2 and Darko Stefanovic1, Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification, Hindawi Publishing Corporation Computational Intelligence and Neuroscience Volume 2016, Article ID 3289801, 11 pages http://dx.doi.org/10.1155/2016/3289801.

6. Vipinadas.M.J, A.Thamizharasi, Detection and Grading of diseases in Banana leaves using Machine Learning, International Journal of Scientific & Engineering Research, Volume 7, Issue 7, July-2016 916 ISSN 2229-5518.
7. Divya Srivastava1 , Rajesh Wadhvani2 and Manasi Gyanchandani3, A Review: Color Feature Extraction Methods for Content Based Image Retrieval, IJCEM International Journal of Computational Engineering & Management, Vol. 18 Issue 3, May 2015 ISSN (Online): 2230-7893 www.IJCEM.org.
8. Gaurav Mandloi, A Survey on Feature Extraction Techniques for Color Images, Gaurav Mandloi / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 4615-4620.
9. Jyotismita Chaki, Ranjan Parekh, Plant Leaf Recognition using Shape based Features and Neural Network classifiers, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 10, 2011.
10. Anandhakrishnan MG Joel Hanson1 , Annette Joy2 , Jerin Francis3, Plant Leaf Disease Detection using Deep Learning and Convolutional Neural Network, IJESC Volume 7 Issue No.3.

AUTHORS FROFILE

Dr S.Prayla Shyry is currently working as Professor in the department of Computer Science and Engineering. She acquired her M.E from Annamalai University and Ph.D from Sathyabama University in the year 2014. She has actively participated as chair person in many workshop, conferences. She has also been a reviewer in reputed journals. She has also published more than 45 national, International journals and conferences. Her area of specialization includes cyber security, network security and overlay networks, Artificial Intelligence, Machine Learning. She has also published patents and modeled manyproducts.