ISSN 2347 - 3983
Volume 8. No. 10, October 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter1148102020.pdf
https://doi.org/10.30534/ijeter/2020/1148102020

Machine Learning based Application to detect Pepper Leaf Diseases using HistGradientBoosting Classifier with fused HOG and LBP Features

Matta Bharathi Devi [1], Dr Amarendra K[2]
[1] Ph.D. Scholar, Dept of CSE, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India
mattabharathi@gmail.com
[2] Professor, Dept of CSE, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India
amarendra@kluniversity.in

ABSTRACT

Pepper leaf disease detection is one of the interesting challenges in the field of machine learning. In this article we propose a machine learning based approach to extract texture features and use dimensionality reduction technique called Principal Component Analysis (PCA) and create composite feature descriptor. We use two different texture-based feature representations extracted by using HOG and LBP feature engineering techniques, from pepper leaf images and apply PCA to get reduced representations. These representations are fused and passed to Machine Learning models like Logistic Regression, Naïve Bayes, decision tree, Support Vector Machine and HistGradientBoosting Classifier for classification. HistGradientBoosting Classifier achieved highest accuracy of 89.11% and outperformed other models.

Key words: Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP), Principal Component Analysis (PCA), HistGradientBoosting Classifier (HGB), Machine Learning.

1. INTRODUCTION

Detecting plant leaf diseases is one of the major challenges faced by farmers in agriculture. It is very important to identify the type of leaf diseases accurately for appropriate use of pesticides. Any mistakes in identifying diseases of plants leads to reduced yield. Plant diseases can be either biotic [1][30] or abiotic. Primary cause behind the biotic diseases are various living organisms like bacteria, virus, and fungi. Biotic diseases are affected by viruses unlike abiotic diseases which are affected by inorganic conditions like weather changes, chemicals etc. Identifying leaf diseases accurately by observing with naked eye is a difficult task. Hence, there is a requirement of an application that can detect leaf diseases accurately. There are various automated applications to identify plant leaf diseases. Most of them used texture representations extracted from leaf images with conventional Machine Learning models [2][19][20].

Most of the recent works in literature used feature extraction techniques like Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), Gray Level Co-occurrence Matrix (GLCM) are used in literature to extract texture-based features from plant leaf images [3][21]. These features were fed to popular classifiers like Support Vector Machine (SVM) to categorize different types of diseased leaves [4][22]. However, using these features directly with ML models results in reduced performance. So, in this work we investigate to reduce the dimensions of texture features and blend them to get composite representation with pepper leaf dataset [23][24][25].

Initially, we performed necessary pre-processing to remove background noise obtained during image acquisition. Later, we extracted two types of texture-based features from pepper leaf images using HOG and LBP feature engineering methods and applied Principal Component Analysis (PCA) dimensionality reduction technique to get reduced representations of HOG and LBP features. In our experiments we observed that, HOG features are better than LBP. Using reduced representations lead to improve performance. When LBP features are fused with HOG, composite representations are obtained. These representations contain more discriminant information which help classification models to identify pepper leaf diseases accurately. Our proposed fused representation achieved a highest accuracy of 89.11% with HGB Classifier.

2. RELATED WORK

This part of the article provides an overview of various methodologies employed for detecting plant leaf diseases in past. First part of this section describes various pre-processing techniques used in literature followed by Feature Engineering methods algorithms that are used for classification.

In recent past, several pre-processing techniques have been applied on plant leaf images to correctly identify the type of plant diseases. Most of the previous works used image processing techniques and applied smoothing, sharpening filters the enhance the image and used several filters to remove additive noise from the images [5][26]. ROI segmentation is major task employed to detect and segment diseased portions from images to improve the performance of automated plant leaf disease diagnosis systems [6][27][31].

Texture based features obtained from images play a vital role and effects the performance of image classification systems. Histogram of Oriented Gradients (HOG), GIST, Scale Invariant Feature Transform (SIFT), Local Binary Patterns (LBP) are majorly employed feature engineering algorithms to obtain intensity and texture-based features [7][8][28]. These feature engineering methods are employed in various tasks like medical image classification, scene classification, object recognition and leaf disease identification [9][10][15][16]. Most of machine learning algorithms like K-Nearest Neighbour Classifier (KNN), Decision Tree Classifier, Random Forest, Support Vector Machine (SVM), Naïve Bayes Classifier are trained on these texture-based features for classification purpose [11].

A K-Nearest Neighbour (K-NN) Classifier with Gray-Level Co-occurrence Matrix (GLCM) texture features of plant leaf images were used to identify plant leaf diseases [12][29]. Another machine learning based system is was proposed for grapes plant leaf disease detection by Harshal Waghmare at al. First, background of all images was removed and segmentation is performed as a pre-processing step. A high-pass filter is applied on segmented images to analyse disease part of the leaf. Local Binary Patterns (LBP) based texture features are extracted from pre-processed images and these features were used to identify different types of grape plant diseases using Support Vector Machine (SVM) Classifier [9][17]. A cotton leaf disease detection and classification technique based on machine learning and image processing tools is proposed by V Pooja at al. Initially, Region of Interest (ROI) is segmented from plant leaf images using image processing tools and features are extracted. These features were passed to SVM Classifier to identify the type of disease [13][18].

In this work we use HOG and LBP feature extracted from pepper leaf images and fuse them to create composite representation. These representations are then projected to lower dimension using PCA. Then we apply various popular classification algorithms like Logistic Regression, Naïve Bayes Classifier, Decision Tree Classifier, Support Vector Machine (SVM) with linear and Radial Basis Function (RBF) kernel and HistGradientBoosting Classifier for classification purpose.

3. PROPOSED METHODOLOGY

This part of the article provides an illustration of various stages of proposed method for pepper leaf disease detection. Our proposed work consists of four phases, followed one after other. They are Data Pre-processing, Feature Extraction, Dimensionality Reduction and Classification.

3.1 Data Pre-processing:

Data acquired from real world consist of random noise in background. So, background subtraction is performed on pepper leaf images to remove random background noise. This is done by creating suitable mask for every image present in dataset and then background removal operation is performed by using corresponding masks. Figure 1(a) represents images from original dataset and Figure 1(b) represents background removed images. These processed images are passed to feature extraction phase.

(a) Pepper leaf images with background noise

(b) Pepper leaf images after background removal

Figure 1. ***Pepper leaf images before and after pre-processing***

3.2 Feature Extraction:

Feature extraction is an important phase in any machine learning task. In our work, we use two different feature extraction techniques, Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) which extracts texture-based features from pepper leaf images.

3.2.1 Feature Extraction from pepper leaf images using HOG: The HOG feature descriptor counts the occurrences of gradient orientation in localized portions of an image. Initially, all processed images of dimension (256 X 256) are reshaped to (64 X 128) dimensions. Next, changes in X and Y directions of images (gradients) are computed by dividing the entire image into (8 X 8) patches. Next, magnitude and orientations are computed by using gradients. Then, Histogram of Gradients are calculated for each (8 X 8) cells and these cells are combined to create (16 X 16) cells. The gradients of these cells are normalized to get a vector of (1 X 36) dimension for each cell. Finally, for every image of dimension (64X128) we get a feature vector of 3780 dimensions. This feature descriptor is normalized using minmax normalization method. Figure 2 represents Histogram of Oriented Gradients computer for a given pepper leaf image.

Figure 2. ***Histogram of Gradients for a given input image***

3.2.2 Feature Extraction from pepper leaf images using LBP: Local Binary Patterns (LBP) computes texture features from local regions instead of computing global texture features as in the case of Gray Level Cooccurrence Matrix (GLCM). Initially, all processed images of dimension (256 X 256) are reshaped to (128 X 128) dimensions. Next, all these images are converted to gray-scale. LBP histogram is obtained from those images by appropriately selecting p and r values, where p represents the number of points in neighbourhood of a pixel and r is the radios. Finally, for every image of dimension (128 X 128) we get a feature vector of 26 dimensions. This feature descriptor is normalized using minmax normalization method. We used OpenCV module of python to extract LBP features.

3.3 Dimensionality Reduction:

In this phase, All the features of dimension 3780 obtained from HOG feature extraction technique and 26 dimensions obtained from LBP are projected into lower dimensional space with 512 and 13 dimensions for HOG and LBP respectively. For this, we used Principal Component Analysis. In the case of limited data, high dimensional features may lead to curse of dimensionality. To resolve this problem, we included this module in our work. Figure 3, represents the architecture of proposed method for pepper leaf disease detection.

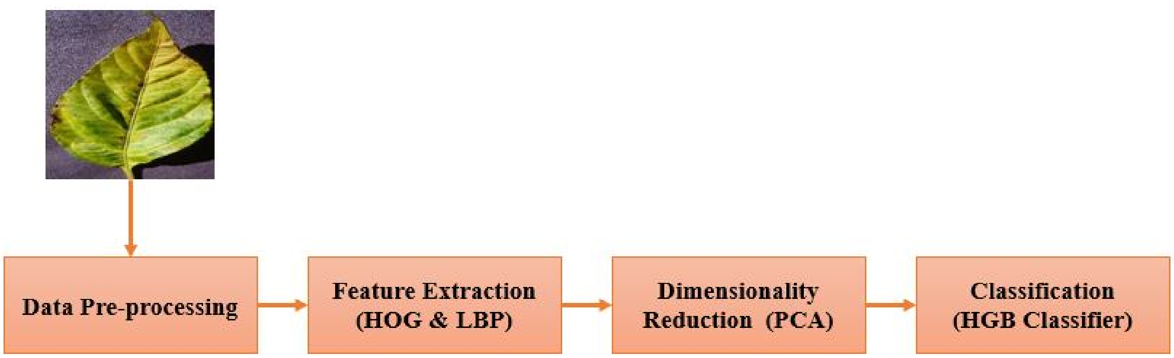

Figure 3. ***Architecture of Proposed System for Pepper leaf disease detection***

3.4 Classification:

In this work we used popular classification models like Logistic Regression, Naïve Bayes Classifier, Decision Tree Classifier, Support Vector Machine (SVM) with linear and Radial Basis Function (RBF) kernel and HistGradientBoosting Classifier. We test the performance extracted features before and after applying Dimensionality Reduction. We observed that SVM with RBF kernel and HistGradientBoosting kernels perform better than other classifiers for both HOG and LBP features in both the cases of dimensionality reduction. Finally, we fused HOG and LBP features to form a composite feature representation of dimension 3806. With these features HistGradientBoosting classifiers achieved a highest accuracy of 89.11% and outperformed all other models.

4. EXPERIMENTAL RESULTS

This section provides clear picture of experiments conducted and results obtained using proposed method. First overview of dataset used for experiments is described followed by evaluation metrics used to measure the performance of proposed method. Finally, a summary of experiments and their results is provided.

4.1 Pepper Leaf Disease Dataset:

We used Pepper Leaf Disease dataset, part of Plant Village dataset which contains 54,306 samples of 26 type of diseased leaf images belonging to 14 types of plant species. This dataset contains 2475 samples of pepper plant representing both healthy and diseased leaves. 997 samples belong to healthy category and 1478 samples belong to bacterial spot category. Totally 1980 samples are considered for training and 495 samples are used for testing model performance.

4.2 Performance Evaluation Measures:

Different classification model performance evaluation measures like accuracy, precision, recall, f1 score are calculated to prove the efficiency of proposed method on test data. These measures can be computed using Confusion Matrix.

4.3 Result Analysis:

We conducted our experiments in three different ways to check the performance of classification models with HOG, LBP and fused features before and after applying PCA for pepper leaf disease detection.

4.3.1 Experiments without Dimensionality Reduction: These experiments are conducted to check how well HOG and LBP features can detect pepper leaf diseases before dimensionality reduction.

Model	Accuracy	Precision	Recall	F1
LR	75.4	75	75	75
Naïve Bayes	70.97	78	71	66
Decision Tree	67.74	67	68	67
SVM - linear	77.82	78	78	78
SVM - RBF	83.06	84	83	83
HGB	84.81	85	84	85

Table 1. ***Performance of ML algorithms with HOG features***

Model	Accuracy	Precision	Recall	F1
LR	80.24	80	80	80
Naïve Bayes	67.74	69	68	68
Decision Tree	72.58	73	73	73
SVM - linear	81.05	81	81	81
SVM - RBF	81.85	82	82	82
HGB	83.87	84	84	84

Table 2. ***Performance of ML algorithms with LBP features***

From table 1, it is clear that HistGradientBoosting Classifier outperformed other ML by achieving accuracy of 84.81%. Comparatively, Decision Tree classifier could not perform well. SVM with RBF kernel also obtained an accuracy of 83.06% which is second highest measure. From table 2, we can observe that HGB classifier achieved 83.87% accuracy with LBP features and Naïve Bayes classifier obtained lower accuracy. From both experiments we can conclude that, HOG features perform better than LBP features for the task of pepper leaf disease detection and HGB classifier out performed all other models used, with both types of features.

4.3.2 Experiments after applying Dimensionality Reduction: These experiments are conducted to check how well HOG and LBP features can detect pepper leaf diseases after dimensionality reduction.

Model	Accuracy	Precision	Recall	F1
LR	73.39	73	73	73
Naïve Bayes	76.21	76	76	76
Decision Tree	62.5	62	62	62
SVM - linear	77.98	79	78	79
SVM - RBF	84.27	84	84	84
HGB	85.47	85	85	85

Table 3. ***Performance of ML algorithms with HOG features***

Model	Accuracy	Precision	Recall	F1
LR	79.44	79	79	79
Naïve Bayes	82.26	82	82	82
Decision Tree	77.02	77	77	77
SVM - linear	80.24	80	80	80
SVM - RBF	83.47	83	83	83
HGB	84.27	84	84	84

Table 4. ***Performance of ML algorithms with LBP features***

From previous experiments, it is clear that after applying PCA, there is significant improvement in the performance of classification models with both, HOG and LBP features. HGB classifier followed same trend and outperformed other classification models with both HOG and LBP features after applying PCA. Even after reduced dimension, there is significant improvement in all measures used to test efficiency of models.

4.3.3 Experiments with fused features of HOG and LBP: This experiment is conducted to check the performance of ML models with composite representation obtained after blending LBP features with HOG features.

Model	Accuracy	Precision	Recall	F1
LR	80.24	80	80	80
Naïve Bayes	75.81	76	76	76
Decision Tree	69.35	70	69	69
SVM - linear	81.05	81	81	81
SVM - RBF	88.71	89	89	89
HGB	89.11	89	89	89

Table 5. ***Performance of ML algorithms with fused HOG & LBP features***

From table 5, it is clear that HGB classifier trained on fused feature descriptor obtained 89.11% accuracy which is highest when compared with the performance of same classifier trained on HOG and LBP features before and after applying

PCA. So, we conclude that, fused texture representations of pepper leaf images help to identify diseases accurately rather than using conventional usage of LBP and HOG features.

5. CONCLUSION

Pepper is most used ingredient in dishes. Identifying pepper leaf diseases is a challenge for formers. There is a high requirement to automate the process of detecting pepper leaf diseased for correct usage of pesticides and reduce loss of yield. In this paper we investigated the performance of various classification models with two different types texture-based features. During our experiments we observed that models can perform well with reduced representations of HOG and LBP features rather than using them directly. We also observed that fused representation of HOG and LBP features helped the models to perform well, and there is 4% improvement in accuracy with fused features. In our experiments we also observed that HGB classifier outperforms other ML algorithm in every case.

REFERENCES

[1]. Husin ZB, Aziz AHBA, Shakaff AYBM, Farook RBSM (2012) Feasibility study on plant chili disease detection using image processing techniques. In: IEEE 3rd international conference on intelligent system modeling and simulation ISMS., Kota Kinabalu, pp 291–296

[2]. Barbedo JGA (2016) A review on the main challenges in automatic plant disease identification based on visible range images. Biosyst Eng 144:52–60

[3]. Mahapatra, S., Kannoth, S., Chiliveri, R. and Dhannawat, R., 2020. Plant Leaf Classification and Disease Recognition using SVM, a Machine Learning Approach. Sustainable Humanosphere, 16(1), pp.1817-1825.

[4]. Bhagat, M., Kumar, D., Haque, I., Munda, H.S. and Bhagat, R., 2020, February. Plant Leaf Disease Classification Using Grid Search Based SVM. In 2nd International Conference on Data, Engineering and Applications (IDEA) (pp. 1-6). IEEE.

[5]. Asfarian A, Herdiyeni Y, Rauf A, Mutaqin KM (2013) Paddy diseases identification with texture analysis using fractal descriptors based on Fourier spectrum. In: IEEE international conference on computer, control, informatics and its applications IC3INA, Jakarta, pp 77–81

[6]. Khirade SD, Patil AB (2015) Plant disease detection using image processing. In: IEEE international conference on computing communication control and automation (ICCUBEA), pp 768–771

[7]. Patil, R. and Kumar Dr, S., 2020. A Bibliometric Survey on the Diagnosis of Plant Leaf Diseases using Artificial Intelligence.

[8]. Mahapatra, S., Kannoth, S., Chiliveri, R. and Dhannawat, R., 2020. Plant Leaf Classification and Disease Recognition using SVM, a Machine Learning Approach. Sustainable Humanosphere, 16(1), pp.1817-1825.

[9]. Bodapati, J.D., Veeranjaneyulu, N., Shareef, S.N., Hakak, S., Bilal, M., Maddikunta, P.K.R. and Jo, O., 2020. Blended Multi-Modal Deep ConvNet Features for Diabetic Retinopathy Severity Prediction. Electronics, 9(6), p.914.

[10]. Dondeti, V., Bodapati, J.D., Shareef, S.N. and Naralasetti, V., Deep Convolution Features in Non-linear Embedding Space for Fundus Image Classification Deep Convolution Features in Non-linear Embedding Space for Fundus Image Classification.

[11]. Bhagat, M., Kumar, D., Haque, I., Munda, H.S. and Bhagat, R., 2020, February. Plant Leaf Disease Classification Using Grid Search Based SVM. In 2nd International Conference on Data, Engineering and Applications (IDEA) (pp. 1-6). IEEE.

[12]. Trivedi, J., Shamnani, Y. and Gajjar, R., 2020, February. Plant Leaf Disease Detection Using Machine Learning. In *International Conference on Emerging Technology Trends in Electronics Communication and Networking* (pp. 267-276). Springer, Singapore.

[13]. Waghmare, H., Kokare, R. and Dandawate, Y., 2016, February. Detection and classification of diseases of grape plant using opposite colour local binary pattern feature and machine learning for automated decision support system. In 2016 3rd international conference on signal processing and integrated networks (SPIN) (pp. 513-518). IEEE.

[14]. Pooja, V., Das, R. and Kanchana, V., 2017, April. Identification of plant leaf diseases using image processing techniques. In 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (pp. 130-133). IEEE.

[15]. Bodapati, J.D., Shaik, N.S., Naralasetti, V. and Mundukur, N.B., 2020. Joint training of two-channel deep neural network for brain tumor classification. Signal, Image and Video Processing, pp.1-8.

[16]. Bodapati, J.D., Veeranjaneyulu, N. and Shaik, S., 2019. Sentiment Analysis from Movie Reviews Using LSTMs. Ingénierie des Systèmes d Inf., 24(1), pp.125-129.

[17]. Anila, M., & Pradeepini, G. " Study of prediction algorithms for selecting appropriate classifier in machine learning" Journal of Advanced Research in Dynamical and Control Systems, 9(Special Issue 18), 257-268, 2017.

[18]. Shariff, M. N., Saisambasivarao, B., Vishvak, T., & Rajesh Kumar T., "Biometric user identity verification using speech recognition based on ANN/HMM", Journal of Advanced Research in Dynamical and Control Systems, 9(12 Special issue),pp.1739-1748. 2017.

[19]. Mannepalli, K., Sastry, P. N., & Suman, M. "Accent recognition system using deep belief networks for Telugu speech signals, International Journal of Speech Technology, 19(1), pp.87-93, 2017.
[20]. T. Rajesh Kumar, G. R. Suresh, S. Kanaga Suba Raja, "Conversion of Non Audible Murmur to Normal Speech based on Full-rank Gaussian Mixture Model", Journal of Computational and Theoretical NanoScience, 1546-1955, Iss.1,vol-15, Pp:185-190, 2018.
[21]. Ayushree, & Balaji, G. N. "Comparative analysis of coherent routing using machine learning approach in MANET", Smart Computing and Informatics (pp.731-741), 2018.
[22]. Puri, G. D., Haritha, D. (2018). Framework to avoid similarity attack in big streaming data. International Journal of Electrical and Computer Engineering, 8(5), 2920-2925, 2018.
[23]. Anjali Devi, S., Siva Kumar.S," Comprehensive survey on sentiment analysis based on workflow foundation", Journal of Advanced Research in Dynamical and Control Systems, 10(9 Special Issue), 1189-120,2018.
[24]. Rajesh Kumar T., Vamsidhar T., Harika B., Madan Kumar T., Nissy R., "Students Performance Prediction Using Data Mining Techniques", IEEE Explorer (ICISS-2019), 978-1-5386-7798-8, 2019.
[25]. Talasila, V., Rajesh Kumar, T., Sai, C.P., Satya Sai, S., Ayyappa, "Predicting the Risk of Heart Failure with EHR sequential Data Modelling", International Journal of Recent Technology and Engineering(IJRTE), 2277-3878, Iss.7,Vol-6,Pp:458-461, 2019.
[26]. Bommadevara, H. S. A., Sowmya, Y., Pradeepini, G. (2019). Heart disease prediction using machine learning algorithms. International Journal of Innovative Technology and Exploring Engineering, 8(5), 270-272, 2019.
[27]. Rajesh Kumar T, Suresh GR, Kanaga Subaraja S, Karthikeyan C., "Taylor-AMS features and deep convolutional neural network for converting non-audible murmur to normal speech", Computational Intelligence, Pp.1–24. 2020.
[28]. Dudi, B. and Rajesh, V. 2019. Medicinal plant recognition based on cnn and machine learning. International journal of Advanced Trends in Computer science and Engineering, 8(4),pp.628-631.
[29]. Inthiyaz, S., Prasad, M.V.D., Lakshmi, R.U.S., Sai, N.S., Kumar, P.P. and Ahammad, S.H., 2019. Agriculture based plant leaf health assessment tool: A Deep Learning perspective. International Journal of Emerging Trends in Engineering Research, 7(11), pp.690-694.
[30]. Balakrishna, G. and Moparthi, N.R., 2020. Study report on indian agriculture with IoT. International Journal of Electrical and Computer Engineering, 10(3), p.2322.
[31]. Dudi, B., and Rajesh, V.,2018An efficient algorithm for medicinal plant recognition. International Journal of pharmaceutical Research, 10 (3),pp.87-93.