**Classification on Leaves Based on their Species using Machine Learning Technique**

**Abstract**

Recognition of plants has become a lively space of analysis as most of the plant species are at the danger of extinction. This paper uses an economical machine learning approach for the classification purpose. This projected approach consists of 3 phases: pre-processing, feature extraction and classification. The pre-processing section involves typical image process steps like reworking to grey scale and boundary enhancement. The feature extraction section derives the common DMF from 5 basic features. The biggest contribution of this approach is the Support Vector Machine (SVM) classification for economical leaf recognition. Different leaf options that are extracted and orthogonalized into many principal variables are given as input vector to the SVM. When tested with a [14]flavia-dataset and a true dataset, the projected approach produces very good accuracy while taking significantly less time to execute.

**Introduction**

Plants play a crucial function withinside the environment. There could be no lifestyles of the earth’s ecology with out plant life. However, recently, numerous species of plant life are on the risk of extinction. In order to protect plant life and to catalogue diverse species of plants diversities, a plant database turns into very important. There is massive quantity of plant species worldwide. In order to address such volumes of information, improvement of a fast and equipped type method has end up an lively location of studies [1].

More over, at the side of the conservation characteristic, reputation of plant life has also end up important to take advantage of their medicinal residences and the use of them as re assets of opportunity strength re assets like bio-fuel. There are diverse approaches to apprehend a plant, like flower, root, leaf, fruit etc. Recently, laptop imaginative and prescient and sample reputation strategies had been carried out closer to automated system of plant reputation [2]The type of plant leaves is a crucial mechanism in botany and in tea, cotton and different industries [3], [4]. Additionally, the morphological functions of leaves are hired for plant type or withinside the early analysis of sure plant diseases [5].

Plant reputation is an important and hard mission. Leaf reputation performs an critical function in plant type and its key trouble lies in whether or not the selected functions are consistent and feature right functionality to discriminate diverse types of leaves. The reputation method could be very time consuming. Computer aided plant reputation remains very hard mission in laptop imaginative and prescient due to flawed models and inefficient illustration approaches. The principal intention of plant reputation is to assess the leaf geometrical morphological and Fourier second based functions. This statistics could be very crucial in figuring out the diverse training of plant life. Ji Xiang, Huang and Xiao Feng [6] completed their research on recognizing the recognized plant species through salient functions of the leaf along with physiological length, width, diameter, perimeter, location, clean factor, component ratio and Fourier moments which will be hired to discriminate with every different. The extraction of leaf functions from a plant is a key step withinside the plant reputation system [7, 8].

This characteristic extraction system creates a brand new venture withinside the area of sample reputation [9] [10]. The statistics acquisition from dwelling plant robotically through the laptop has now no longer been implemented. This paper implements a leaf reputation set of rules the use of smooth to extract functions and excessive green reputation set of rules. The principal levels worried on this studies are characteristic extraction and the type. All functions are extracted from virtual leaf image. Many functions are orthogonalized through Principal Components Analysis (PCA) and are given to the classifier. The classifier used on this method is Support Vector Machine for its rapid velocity and easy structure

**Literature Survey**

[1]In this paper they used Zernike Moments and Histogram of Oriented Gradients for category of plant leaves, After applying Zernike Moments and Histogram of Oriented Gradients they used Support Vector Machine Classifier is used for leaf recognition .[2] They used Preferential image segmentation is proposed for automatic recognition of leafs and flowers. This method encodes the prior information for preferential segmentation as trees of shapes.

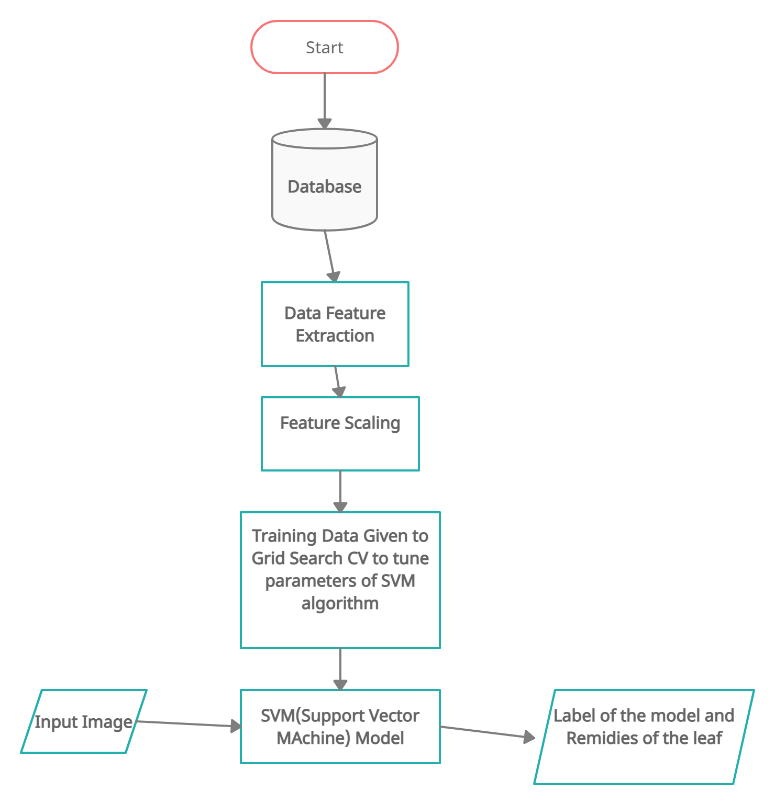
[3]In this paper, an efficient computer-aided plant species identity (CAPSI) method is proposed, which is primarily based totally on plant leaf snap shots the usage of a form matching technique. Firstly, a Douglas - Peucker approximation set of rules is followed to the unique leaf shapes and a brand-new form illustration is used to shape the collection of invariant attributes. Then a changed dynamic programming (MDP) set of rules for form matching is proposed for the plant leaf popularity. Finally, the superiority of our proposed technique over conventional techniques to plant species identity is confirmed through experiment. The experimental end result showed that our proposed set of rules for leaf form matching may be very appropriate for the popularity of now no longer most effective intact however additionally partial, distorted and overlapped plant leaves because of its robustness.

[18] They utilised two algorithms in their paper. These two algorithms, Support Vector Machine and Artificial Neural Network, were used to predict whether the leaf was fresh or dry. The shading histogram, edge histogram, and sobel edge bearing are all used to extract features. They used 50 photos to compare the two models. They found that SVM classifies better than ANN after testing.

[19]In this paper, they used Convolutional Neural Network for classifying Indian leaf species using Smartphones. They used images with white background to train the CNN model. They used separate edge detection called Prewitt. It is a discrete differentiation operator that computes picture intensities' gradient approximation. To put it another way, the Prewitt operator estimates image intensity point-by-point to capture the smooth variation of leaf image changes in any direction. The horizontal and vertical intensities are determined, and the direction with the highest potential intensity is checked.Leaf edge and vein segmentation is also done using the Sobel and Laplacian edge detectors. With prewitt edge detection, the results of the sobel and laplacian operators are averaged, and the skeleton of the leaf is acquired for further classification.

[20] They compared both standard Machine Learning Algorithms and Deep Learning in their paper, using a TensorFlow plantvillage leaf dataset. Nearly 54,306 photos of 14 different plant species are included in the plantvillage dataset, which is divided into 38 different classes based on species/disease pairs. Popular machine learning algorithms include KNN, SVM, and Fully Connected Neural Network (FCNN) and a deep learning method CNN(Convolutional Neural Network). They came to the conclusion that CNN outperforms the previously described popular machine techniques.

[21] They concluded that k-nearest-neighbour approach is possibly the handiest of all algorithms for predicting the magnificence of a check example. An obvious drawback of the k-NN approach is the time complexity of creating predictions. Additionally, neural networks are tolerant to noisy inputs. But in neural network it’s tough to recognize structure of algorithm. In identifying high-dimensional data sets, SVM was found to be competitive with the best current machine learning methods. The computational complexity of SVM is reduced to a quadratic optimization problem, making the complexity of the decision rule and the frequency of error easy to regulate. When training data is not linearly separable, it is challenging to establish optimal parameters with SVM.



**Working:**

To obtain the desired results, we must take different steps in order to achieve the project's goals. The first of these steps is dataset compilation. It's a crucial component of every Machine Learning project. In general, we may obtain datasets from an online database or generate datasets on our own in response to a request. Following the collection of the results, Under the pre-processing stage, we removed the unwanted data. After pre-processing, the data must be fed into an algorithm for training purposes, but first we must extract the leaf's features, which is known as the feature extraction method. Split the collected data into training and testing after obtaining the featured vector.We give the training data to the SVM algorithm for computation, and once we achieve a reasonable level of accuracy, we predict the results.

**Dataset Gathering:**

Dataset which we used in project is flavia-dataset which contains nearly 33 categories[pubescent bamboo ,Chinese horse chestnut ,Anhui Barberry ,Chinese redbud, true indigo, Japanese maple, Nanmu, castor aralia, Chinese cinnamon, goldenrain tree, Big-fruited Holly, Japanese cheesewood, wintersweet, camphortree, Japan Arrowwood, sweet osmanthus, deodar, ginkgo, Crepe myrtle, oleander, yew plum pine, Japanese Flowering Cherry, Glossy Privet, Chinese Toon, peach, Ford Woodlotus, trident maple, Beale's barberry, southern magnolia, Canadian poplar, Chinese tulip tree and tangerine]. File names of all images are 4-digit numbers , followed by a “.jpg” suffix.

**Pre-processing:**

* Converting each image in the dataset from RGB to BGR color
* After Converting the Image in BGR which takes input and then convert into gray scale image

**Feature Extraction:**

For computation, a machine can only recognise binary values, but here image data (.jpg) is used, which the computer is unaware of. It's important to extract data that can be used in calculations. Feature extraction starts with a set of measured data and generates derived values (features) that are meant to be informative and non-redundant, making understanding and generalisation easier and, in some cases, resulting in better human interpretations. Feature extraction is related to dimensionality reduction. The features that are extracted can differ depending on the requirements.

**Shape Feature:**

Here we are using DMF features where DMF Stands for (Digital Morphological Features).These are the basic Features

1. Area(A):Which will calculate area of the leaf
2. Perimeter(P):Which will calculate the perimeter of the leaf.
3. Physiological Length(Lp):Which will calculate the physiological length
4. Physiological width(Wp):Which will calculate the physiological width.

The derived features are

1. Aspect ratio: the ratio between physiological length and physiological width
2. Rectangularity: Which will calculate the similarity between leaf. which is
3. Circularity: Which is perimeter square divide with area

**Colour feature:**

Every image is combination of Red, Green and Blue

In colour feature we will extract mean of Red, mean of Green, mean of Blue, Standard Deviation of Red, Standard Deviation of Green and Standard Deviation of Blue.

**Texture Feature:**

It will partition image into regions of interest and to classify those regions

It provides arrangement of colours and intensities in an Image

The texture feature using here are

1. Contrast
2. Correlation
3. Inverse difference moments
4. Entropy

**Feature Scaling**

Feature Scaling is a process to standardize the independent features present in the data in fixed range. It is performed during the pre-processing phase to handle highly varying values. If feature scaling is not done, then a Machine Learning algorithm tends to weigh greater values .The 2 most important feature scaling methods are Min-Max Normalization and standardization. In this project standardization is used. It is effective technique which re-scales a feature values.It has the distribution with 0 mean value and variance equals to 1.

**Algorithm Implementation:**

This is classification problem where there are many classification algorithms . Here we are using SVM ( Support Vector Machine) . Along with SVM , Grid Search CV which will help to tine parameters in the SVM Algorithm.

****Support Vector Machine**

The Support Vector Machine is a classification algorithm that categorises data based on its features. The sklearn package's default parameters for the SVM algorithm are (C=1.0, kernel='rbf, gamma='scale'). However, there are numerous parameters that influence the parameters of SVM, which must be manually checked for each parameter. It will take more time to obtain parameters with higher accuracy. As a result, GridSearchCV is used in along with SVM.

**GridSearchCV:**

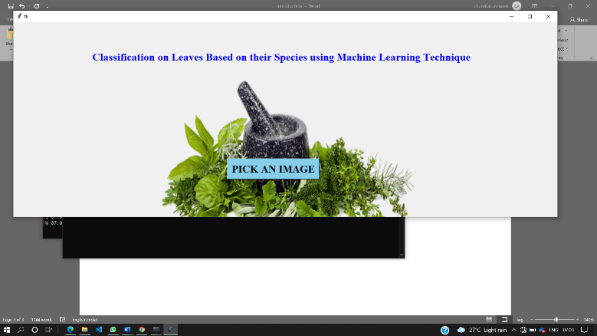
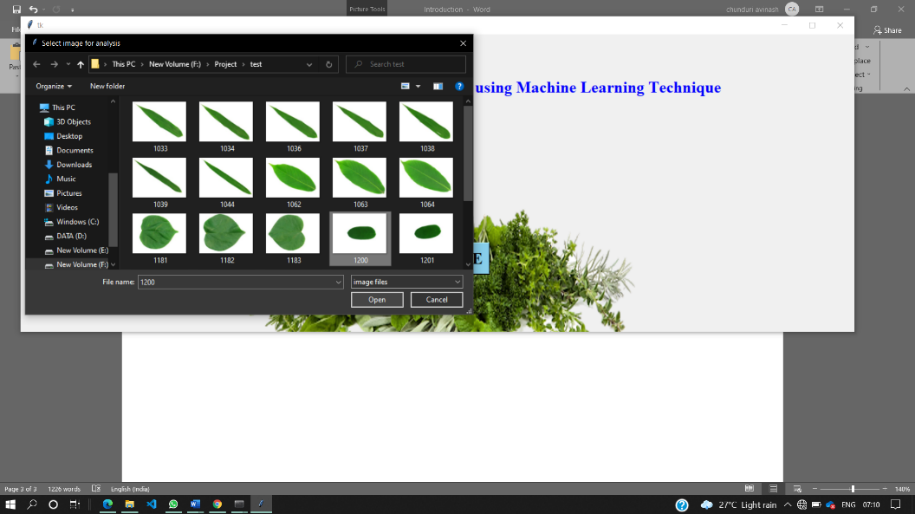
We can select Hyperparameter which is used to increase accuracy of modal. We can use GridSearchCV for any classification algorithm

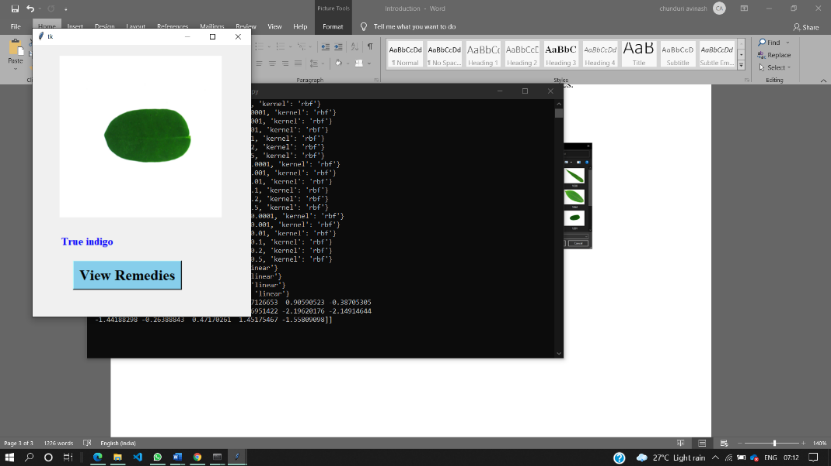
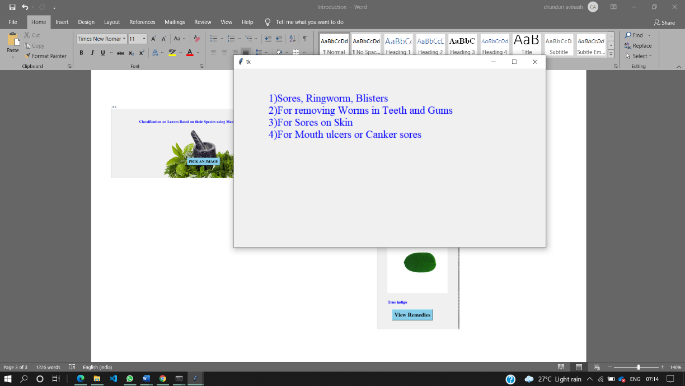
The parameters of GridSearchCV are

* Estimator: Actual Algorithm
* Parm-grid: Hyperparameters
* CV: Cross Validation

The GridSearchCV iterate through all the parameters and gives the best modal which is giving the best accuracy.

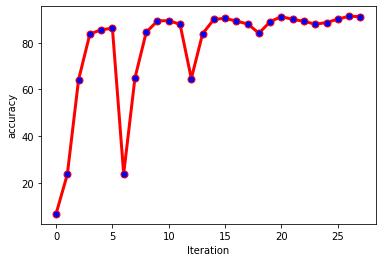
**Predictions:**

In this step, Take leaf image as an input through GUI interface as shown in the Figure 2. Then after selecting an image for predictions. The image will go to various steps internally, Removing the background of an image after removing background extracting the features from that leaf and storing in a feature vector before giving the feature vector to the pretrained model. The features have to be scaled between 0 to 1. Modified feature vector will be given to enhanced SVM model as an input which will predicts the label of an image and along with their remedies.



Predicting the result of the classifier

**Conclusion:**

In this project, an approach to plant classification based on leaf recognition is proposed. The method was divided into three stages: pre-processing, feature extraction, and classification. The dataset reveal 32 different types of plants. The classification method is based on the SVM classifier, which has a higher accuracy, faster training speed, and a simple structure. PCA extracts and processes 12 features to construct the SVM input vector. The accuracy and execution time of the proposed solution are used to test its performance. The efficiency of the classifier can be enhanced in future ****research by integrating efficient kernel functions.

**Results**

The Performance of the model is high when the parameters of SVM are C=100 and kernel=” linear”. By, using this value we trained the SVM Model and predict the leaf label and their Medicinal Remedies to the user via GUI interface, as show in figure above.

**References**:

[1]Plant Leaf Recognition Using Zernike Moments and Histogram of Oriented Gradients by Dimitris G. TsolakidisDimitrios I. KosmopoulosGeorge Papadourakis

[2] J. Pan, and Y. He, ”Recognition of plants by leaves digital image and neural network,” International Conference on Computer Science and Software Engineering, 2008, vol. 4, pp. 906 – 910.

[3] Cotton Incorporated USA, The classification of Cotton, 2005,http://www.cottoninc.com/ClassificationofCotton.

[4] National Institute for Agricultural Botany, Chrysanthemum Leaf Classification. Cambridge, 2005.

[5] N. Kumar, S. Pandey, A. Bhattacharya, and P.S. Ahuja, “Do leaf surface characteristics affect agro bacterium infection in tea,” J. Biosci., vol. 29, no. 3, 2004, pp. 309–317.

[6] Ji-Xiang Du,De-Shuang Huang, Xiao-Feng Wang, and Xiao Gu, “Computer-aided plant species identification (capsi) based on leaf shape matching technique,” Transactions of the Institute of Measurement and Control, vol. 28, 2006, pp. 275-284.

[7] Y. Li, Q. Zhu, Y. Cao, and C. Wang, “A Leaf Vein Extraction Method based on Snakes Technique,” Proceedings of IEEE International Conference on Neural Networks and Brain, 2005.

[8] H. Fu, and Z. Chi, “Combined thresholding and Neural Network Approach for Vein Pattern Extraction from Leaf Images,” IEEE Proceedings-Vision, Image and Signal Processing, 2006, vol. 153, no. 6.

[9] J.-X. Du, X.-F. Wang, and G.-J. Zhang, “Leaf shape based plant species recognition,” Applied Mathematics and Computation, 2007, vol. 185.

[10] F. Gouveia, V. Filipe, M. Reis, C. Couto, and J. Bulas-Cruz, “Biometry: the characterization of chestnut-tree leaves using computer vision,” Proceedings of IEEE International Symposium on Industrial Electronics,

1997.

[11] J. Shlens, “A tutorial on principal component analysis”, 2005. [Online]. Available: http://www.cs.cmu.edu/\_elaw/papers/pca.pdf.

[12] Xiao Gu, Ji-Xiang Du, and Xiao-Feng Wang, “Leaf Recognition Based on the Combination of Wavelet Transform and Gaussian Interpolation,” Advances in Intelligent Computing, vol. 3644, 2005, pp. 253-262.

[13] Xiao-Feng Wang, Ji-Xiang Du, and Guo-Jun Zhang, “Recognition of Leaf Images Based on Shape Features Using a Hypersphere Classifier,” Advances in Intelligent Computing, vol. 3644, 2005, pp. 87-96.

[14] http://flavia.sf.net.

[15] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang, and Qiao-Liang Xiang, “A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network,” IEEE

International Symposium on Signal Processing and Information

Technology, 2007.

[16]Plant Classification using Machine Learning Algorithms, K.B.Shobana and P.Perumal, IEEE publications. (2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)

[17] CNN Approaches for Classification of Indian Leaf Species Using Smartphones , M. Vilasini and P. Ramamoorthy , tech Science Press(2020)

[18] Image-Based Plant Disease Detection: A Comparison of Deep Learning and Classical Machine Learning Algorithms, Draško Radovanović 1 , Slobodan Đukanović (18-22 February 2020)

[19] Detection and Classification of Plant Leaf Diseases Using Image processing Techniques: A Review, Savita N. Ghaiwat, Parul Arora(2014)