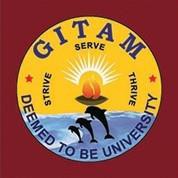
**GANDHI INSTITUTE OF TECHNOLOGY AND MANAGEMENT**

**(Deemed to be university)**

**Bengaluru-561203**

**Project Report on**

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

Submitted in partial fulfilment of the requirement for the degree of

**Bachelor of Technology in Computer Science and Engineering**

Submitted by:

**Chunduri Avinash 321710306006**

**Akash Sridhar 321710306001**

**K Sai Samarth 321710306019**

**M Gopi Chand 321710306026**

**Under the guidance of:**

**Dr. Dayanand Lal N**

**Assistant Professor**

**Department of Computer Science and Engineering**

**GITAM School of Technology, Bengaluru Campus,**

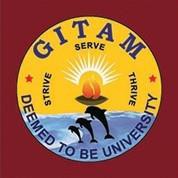
**Nagadenahalli, Doddaballapur Taluk,**

**Bengaluru Rural District, Karnataka-562163**

**2020-2021**

**GANDHI INSTITUTE OF TECHNOLOGY AND MANAGEMENT**

**(Deemed to be University) Bengaluru- 561203**



Department of Computer Science & Engineering

**Certificate**

This is to certify that the Mini-Project titled **“Classification on Leaves Based on their Species using Machine Learning Technique”** is the bonafide work carried out by ‘ChunduriAvinash (321710306006), AkashSridhar (321710306001), KSaiSamarth (321710306019), MGopiChand (321710306026)’with a student of B-Tech (CSE) of GITAM Deemed to be University, Bengaluru campus during the academic year 2020-21, in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology (Computer Science and Engineering) and that the project has not formed the basis for the award previously of any other degree, diploma, fellowship or any other similar title. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the Report deposited in the departmental library.

.

Signature of the Guide Signature of HOD

**Dr. Dayanand Lal N Dr. Vamshidhar Y**

Assistant Professor Head of the Department

Department of CSE, GST Department of CSE, GST

**DECLARATION**

We**,** **ChunduriAvinash , AkashSridhar , KSaiSamarth ,MGopiChand** students of 8th-semester B.Tech in Computer Science & Engineering from GITAM (Deemed to be University), Bangalore, hereby declare that the dissertation work entitled **“Classification on Leaves Based on their Species using Machine Learning Technique”** has been carried out under the guidance of **, Dr. Dayanand Lal N** Assistant Professor Department of Computer Science Engineering, GITAM (Deemed to be University), Bangalore, in the partial fulfilment of the requirement of the degree of the **BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE ENGINEERING OF GANDHI INSTITUTE OF TECHNOLOGY AND MANAGEMENT (GITAM).** We declare that we have not submitted this dissertation either in part or in full to any other University for the award of any degree.

**Signature of the Students**

**Place:** Bengaluru

**Date:**

**ACKNOWLEDGEMENT**

We had been able to complete our project successfully. However, it would not have been possible without the kind support and help of many individuals. We would like to extend our sincere thanks to all of them.

We are highly indebted to GITAM (Deemed to be University), Bangalore for their guidance and constant supervision as well as for providing necessary information regarding the project and also for their support in completing the project.

Would like to express our gratitude towards Prof. Dr Dinesh S (Director of GST), Prof. Dr Brahmananda S.H (HOD of CSE, GST) and also to all the other supporting faculty and staff for their kind co-operation and encouragement which helped us in the completion of this project.

Chunduri Avinash 321710306006

Akash Sridhar 321710306001

K Sai Samarth 321710306019

M Gopi Chand 321710306026

**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 appreciate pre-processing, feature extraction and classification. The pre-processing section involves a typical image process steps appreciate reworking to grey scale and boundary enhancement. The feature extraction section derives the common DMF from 5 basic features. the most contribution of this approach is that 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. Classifier tested with flavia-dataset and a true dataset the projected approach produces very good accuracy and takes terribly less execution time.

**TABLE OF CONTENTS**

[INTRODUCTION 1](#_Toc72668968)

[LITURATURE SUVERY 3](#_Toc72668969)

[SYSTEM REQUIREMENT SPECIFICATION 5](#_Toc72668970)

[3.1 System Analysis 5](#_Toc72668971)

[3.2Functional Requirement 5](#_Toc72668972)

[3.3Non-Functional Requirements 5](#_Toc72668973)

[3.4Tools and Technology details 7](#_Toc72668974)

[3.5Packages 8](#_Toc72668975)

[SYSTEM DESIGN 9](#_Toc72668976)

[4.1 Overall system architecture 9](#_Toc72668977)

[4.2 Activity Diagram 10](#_Toc72668978)

[4.3 State Transition Diagram 11](#_Toc72668979)

[4.4 Data Flow Diagram(DFD) 12](#_Toc72668980)

[IMPLEMENTATION 13](#_Toc72668981)

[5.1 Dataset collection: 13](#_Toc72668982)

[5.2 Data Feature Extraction 14](#_Toc72668983)

[5.3 Feature Scaling 16](#_Toc72668984)

[5.4 Machine learning algorithm apply 17](#_Toc72668985)

[5.5 Grid Search CV: 19](#_Toc72668986)

[5.6 Prediction: 20](#_Toc72668987)

[TESTING 21](#_Toc72668988)

[6.1 Model evaluation in machine learning testing 21](#_Toc72668989)

[6.2 Model tests: 21](#_Toc72668990)

[RESULT 23](#_Toc72668991)

[CONCLUSION 26](#_Toc72668992)

[References 27](#_Toc72668993)

**LIST OF FIGURES**

[**Fig 4.1 Architecture for classification of leaf** 9](#_Toc72764514)

[**Fig 4.2 Activity Diagram for classification** 10](#_Toc72764515)

[**Fig 4.3 State Transition Diagram for classification** 11](#_Toc72764516)

[**Fig 4.4 Data Flow Diagram for classification** 12](#_Toc72764517)

[**Fig 5.1 Work Flow** 13](#_Toc72764518)

[**Fig 5.2 Identifying the right hyper-plane** 17](#_Toc72764519)

[**Fig 5.3 Finding the best hyperplane** 18](#_Toc72764520)

[**Fig 5.4 Ignoring outlier data** 18](#_Toc72764521)

[**Fig 7.1Starting the project** 23](#_Toc72764522)

[**Fig 7.2 GUI interface for picking an image** 23](#_Toc72764523)

[**Fig 7.3 Picking up the image** 24](#_Toc72764524)

[**Fig 7.4 Tuning the parameters** 24](#_Toc72764525)

[**Fig 7.5 Label of the selected image** 24](#_Toc72764526)

[**Fig 7.6 Remedies of the selected leaf** 25](#_Toc72764527)

**LIST OF TABLES**

[**Table 3.1 Hard Ware Requirements** 6](#_Toc72764772)

[**Table 3.2 Software Requirements** 6](#_Toc72764773)

[**Table 5.1 List of Names of Leaves** 14](#_Toc72764774)

**CHAPTER 1**

# 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].

Moreover, 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.

**CHAPTER 2**

# LITURATURE SUVERY

[1] Dimitris G. Tsolakidis et al. proposed Plant Leaf Recognition Using Zernike Moments and Histogram of Oriented Gradients.This technique considers the usage of Zernike Moments and Histogram of Oriented Gradients for category of plant leaf photos and is proposed on this paper. After pre-processing, we compute the form capabilities of a leaf with the usage of Zernike Moments and texture capabilities and the usage of Histogram of Oriented Gradients after which the Support Vector Machine classifier is used for plant leaf photograph category and recognition. Experimental effects display that the usage of each Zernike Moments and Histogram of Oriented Gradients to categorise and apprehend plant leaf photograph yields accuracy this is similar or higher than the nation of the art. The technique has been confirmed on the Flavia and the Swedish Leaves datasets in addition to on a mixed dataset.

[2] N.Valliammal & S.N.Geethalakshmi proposed Leaf and Flower Recognition Using Preferential Image Segmentation Algorithm**,** This paper shows that  
Automatic plant class structures are important for a huge variety of programs together with surroundings protection, plant aid survey, in addition to for education. With the useful resource of superior records technology, photo processing and gadget gaining knowledge of techniques, automated plant identity and class will beautify such structures with extra functionality, such as automated labelling and bendy searching. Image segmentation and item reputation are components of virtual photo processing which can be being an increasing number of used in lots of programs together with leaf reputation. In this paper, the Preferential Image Segmentation (PIS) approach is used to phase an item of hobby from the unique photo. A probabilistic curve evolution approach with particle filters is used to degree the similarity among shapes throughout matching process. The experimental consequences show that the preferential photo segmentation may be efficaciously implemented in leaf reputation and segmentation from a plant photo.

[3] Arunpriya Antony & Selvadoss Thanamani proposed An Effective Tea Leaf Recognition Algorithm for Plant Classification Using Radial Basis Function Machine**.** They stated thatA leaf is an organ of a vascular plant, as recognized in botanical terms, and particularly in plant morphology. Naturally a leaf is a thin, flattened organ endure above floor and it's far specifically used for photosynthesis. Recognition of flowers has end up an lively region of studies as maximum of the plant species are on the threat of extinction. Most of the leaves can't be identified without difficulty in view that a few aren't flat (e.g. succulent leaves and conifers), a few does now no longer develop above floor (e.g. bulb scales), and a few does now no longer go through photosynthetic function (e.g. cataphylls, spines, and cotyledons). In this paper, we specifically centred on tea leaves to become aware of the leaf kind for enhancing tea leaf class. Tea leaf pics are loaded from digital cameras or scanners withinside the system. This proposed technique includes three stages including pre-processing, characteristic extraction and class to process the loaded image. The tea leaf pics may be recognized appropriately withinside the pre-processing section through fuzzy denoising the usage of Dual Tree Discrete Wavelet Transform (DT-DWT) so that you can get rid of the noisy capabilities and boundary enhancement to attain the form of leaf appropriately. In the characteristic extraction section, Digital Morphological Features (DMFs) are derived to enhance the class accuracy. Radial Basis Function (RBF) is used for efficient class. The RBF is skilled through 60 tea leaves to categorise them into 6 types.

[4] Ji-Xiang Du et al. proposed Computer-Aided Plant Species Identification (CAPSI) Based on Leaf Shape Matching Technique and showed 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.

**CHAPTER 3**

# SYSTEM REQUIREMENT SPECIFICATION

## 3.1 System Analysis

Prediction of terrorism sports is an critical place of challenge for researchers. The big wide variety of activities makes it hard to are expecting terrorist organization responsible for a few terrorist activity.The modern studies is centered on locating out the correlation among terrorism and its causal factors. Existing efforts have now no longer been desirable sufficient for prediction. Machine mastering strategies can advert in predicting the probability of a terrorist attack, given the desired data. The consequences of this paintings can help the safety businesses and coverage makers to remove terrorism with the aid of using taking relevant and powerful measures.

Hence there may be an technique to reading terrorism vicinity and us of a with the machine mastering strategies and terrorism particular understanding to fetch conclusions about terrorist conduct patterns.

## Functional Requirement

The particular necessities are user interfaces. The outside clients are the customers.The specific requirements are user interfaces. The outdoor customers are the clients. Everyone of the clients can make use of this product for ordering and looking.

* Hardware interfaces: The outdoor system interface applied for ordering and looking is PCs of the clients. The PCs is probably transportable PCs with far off LAN as the internet affiliation gave may be far off.
* Software Interfaces: The operating Frameworks may be any rendition of windows.
* Performance Prerequisites: The PCs applied ought to be atleast pentium four system with the intention that they are able to supply best execution of the item.

## Non-Functional Requirements

Non utilitarian necessities are the capacities offered by the framework. It incorporates time imperative and requirement on the advancement procedure and models. The non useful prerequisites are as per the following:

* Speed: The framework ought to prepare the given contribution to yield inside fitting time.
* Ease of utilization: The product tought to be easy to understand. At that point the clients can utilize effortlessly, so it doesn't require much preparing time.
* Reliability: The rate of disappointments ought to be less then just the framework is more solid.
* Portability: It thought to be anything but difficult to actualize in any framework

**H/W System Configuration:**

|  |  |
| --- | --- |
| Processor | Dual Core. |
| Speed | 1.1 G Hz. |
| RAM | 1GB. |
| Hard Disk | 500MB. |

**Table 3.1 Hard Ware Requirements**

**S/W System Configuration:**

|  |  |
| --- | --- |
| Operating System | Windows 10. |
| Technology | Machine Learning. |
| Front End | GUI-tkinter. |
| IDLE | Python 3.7 or higher. |

**Table 3.2 Software Requirements**

**Hardware requirements**

The maximum widely identified association of conditions characterised through any running framework or programming utility is the bodily PC assets, in any other case called system, An system requirements listing is often joined through an system similarity listing, especially if there must be an incidence of running frameworks. A HCL information tried, perfect, and now after which incongruent system devices for a particular running framework or utility. The accompanying sub-segments take a look at the exclusive elements of system conditions.

The strength of the focal making ready unit (CPU) is a significant framework necessity for any product. Most programming strolling on x86 engineering signify making ready strength because the version and the clock pace of the CPU. Numerous different highlights of a CPU that effect its pace and strength, much like transport pace, store, and MIPS are often overlooked. This which means of strength is often wrong, as AMD Intel Pentium CPUs at comparative clock pace often have exceptional throughput speeds.

• 10GB HDD(min)

• 128 MB RAM(min)

• Pentium P4 Processor 2.8Ghz(min)

**Software requirements**

Programming requirements define the characteristics of software resource requirements and the requirements that must be input to the PC to make the application run normally. In most cases, these requirements or requirements are not included in the product creation package and must be entered independently before the product is released.

* Python 3.7 or higher
* Pycharm
* opencv

## Tools and Technology details

**Tool:** IDLE is Python's Integrated Development and Learning Environment. It allows programmers to easily write Python code. Just like Python Shell, IDLE can be used to execute a single statement and create, modify, and execute Python scripts.

**Machine Learning** is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that makes it more similar to humans: The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect.

## Packages

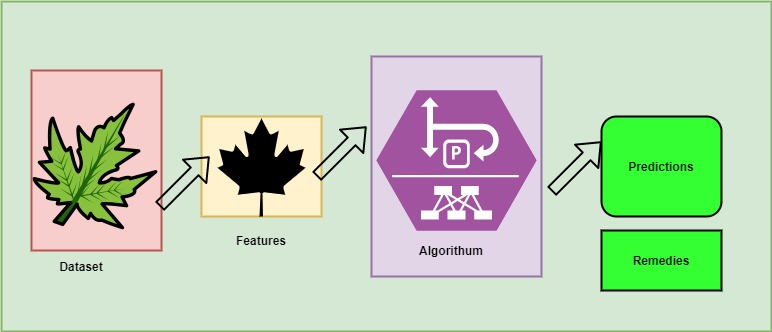
* os : The OS module in Python provides functions for interacting with the operating system. OS comes under Python's standard utility modules.
* cv2: OpenCV (Open Source Computer Vision Library) is a library of programming functions mainly aimed at real-time computer vision.
* Numpy: NumPy is a general-reason array-processing package. It affords a high-overall performance multidimensional array object.
* Pandas: pandas is a software program library written for the Python programming language for facts manipulation and analysis. In particular, it gives facts systems and operations for manipulating numerical tables and time series.
* Mahotas: Mahotas is a computer vision and image processing library for Python. It contains many algorithms implemented in C++ to speed up the work of using a large number of arrays and a very clean Python interface. Mahotas currently has 100 machine vision and computer vision functions, and it continues to grow.
* Matplotlib: This package is used for ploting the result.
* Pip: pip is a package installer .
* Tkinter: It is the standard Python interface to the Tk GUI toolkit.

**CHAPTER 4**

# SYSTEM DESIGN

System design is the process of defining the architecture, components, modules, interfaces and data for a system to satisfy specified requirements. One could see it as the application of systems theory to product development. There is some overlap with the disciplines of systems analysis, systems architecture and systems engineering. If the broader topic of product development "blends the perspective of marketing, design, and manufacturing into a single approach to product development," then design is the act of taking the marketing information and creating the design of the product to be manufactured. Systems design is therefore the process of defining and developing systems to satisfy specified requirements of the user.

## 4.1 Overall system architecture

****

**Fig 4.1 Architecture for classification of leaf**

In system architecture, there are 4 stages involved in system architecture. Collecting dataset, After collecting the data we will extract features from the image . which is given to algorithm by splitting data into training data and testing data. After training phase, the model will go under testing phase. In this testing phase we will calculate the accuracy of the modal. After achieving the expected accuracy predictions are done by giving input image to model which will predict the label of the leaf and with their remedies

## 4.2 Activity Diagram

We use Activity Diagrams to illustrate the go with the drift of manage in a gadget and consult with the stairs concerned in the execution of a use case. We version sequential and concurrent activities the usage of interest diagrams. So, we essentially depict workflows visually the usage of an interest diagram. An interest diagram specializes in situation of go with the drift and the series where in it happens. We describe or depict what reasons a particular occasion the usage of an interest diagram

Dataset

Algorithm

Prediction

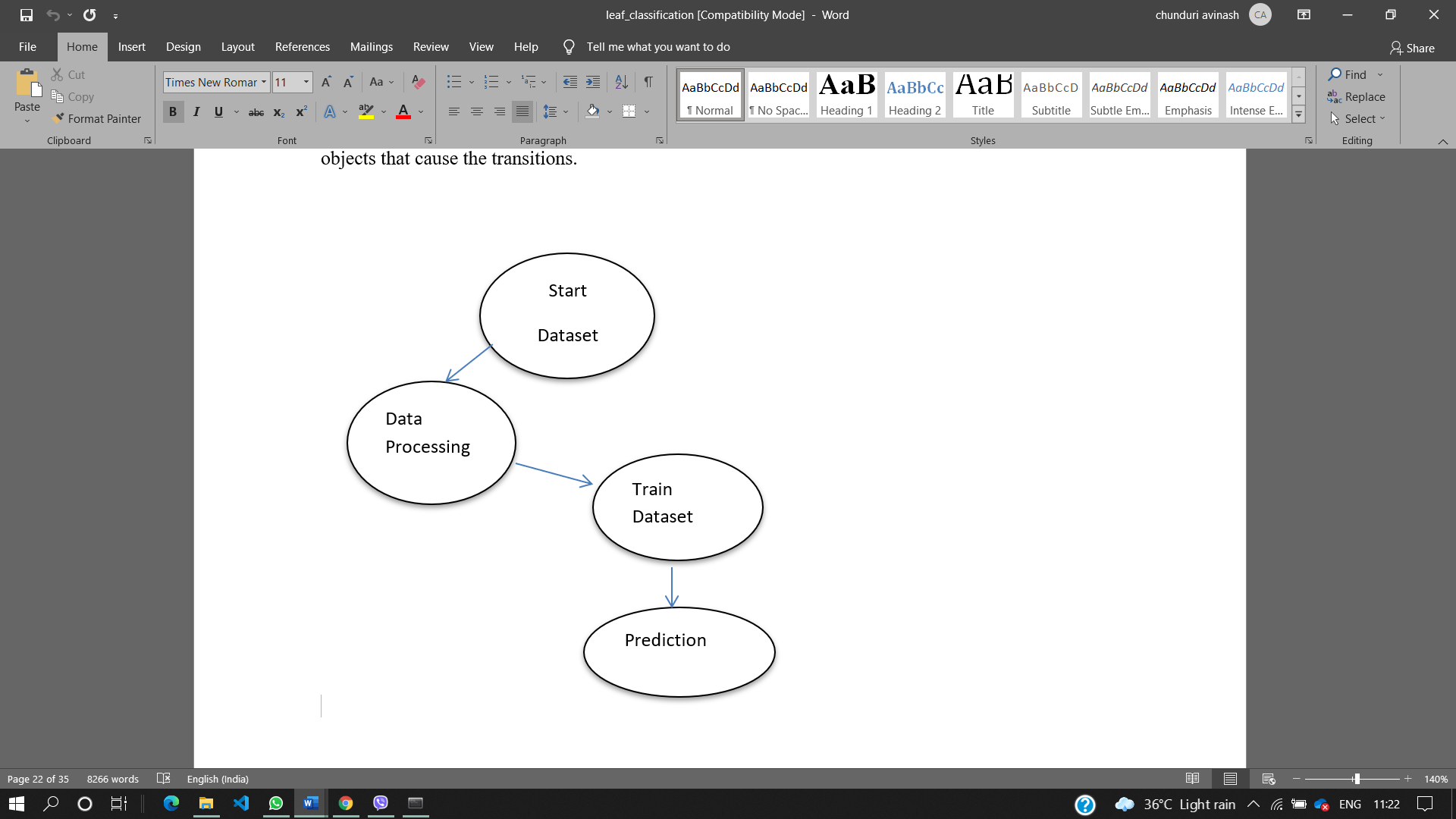
Preprocess

Train data

**Fig 4.2 Activity Diagram for classification**

## 4.3 State Transition Diagram

State-transition diagrams describe all the states that an item can have, the occasions under which an item adjustments state (transitions), the situations that need to be fulfilled earlier than the transition will occur (guards), and the activities undertaken for the duration of the lifestyles of an item (actions). State-transition diagrams are very beneficial for describing the conduct of character items over the full set of use instances that have an effect on the ones items.



**Fig 4.3 State Transition Diagram for classification**

## 4.4 Data Flow Diagram(DFD)

A data flow diagram is a graphical illustration of the "flow" of information through an information system, modelling its method aspects. Often they're a initial step used to create an outline of the system which can later be elaborated. DFDs can also be used for the visualization of information processing (dependent design). The DFD is likewise known as as bubble chart. It is a easy graphical formalism that could be used to symbolize a gadget in phrases of the enter information to the gadget, various processing achieved on those information, and the output information is generated via way of means of the gadget.

Dataset Input

Data Pre-processing

Feature extraction

Apply Algorithm

Predict leaf with remedies

**Fig 4.4 Data Flow Diagram for classification**

**CHAPTER 5**

# IMPLEMENTATION

There are 3 steps for implementation;

1. Dataset collection
2. Data Feature Extraction
3. Machine learning algorithm apply
4. Prediction

**Fig 5.1 Work Flow**

## 5.1 Dataset collection:

We collect the leaf flavia dataset it includes total 32 leaf categories the names of the leaves are

|  |  |
| --- | --- |
| 1 | Castor Aralia |
| 2 | Chinese Horse Chestnut |
| 3 | Anhui Barberry |
| 4 | Chinese Redbud |
| 5 | True Indigo |
| 6 | Japanese Maple |
| 7 | Nanmu |
| 8 | Castor Aralia |
| 9 | Chinese Cinnamon |
| 10 | Goldenrain Tree |
| 11 | Big-fruited Holly |
| 12 | Japanese Cheesewood |
| 13 | Wintersweet |
| 14 | Camphortree |
| 15 | Japan Arrowwood |
| 16 | Sweet Osmanthus |
| 17 | Deodar |
| 18 | Ginkgo, Maidenhair Tree |
| 19 | Crape Myrtle, Crepe Myrtle |
| 20 | Oleander |
| 21 | Yew Plum Pine |
| 22 | Japanese Flowering Cherry |
| 23 | Glossy Privet |
| 24 | Chinese Toon |
| 25 | Peach |
| 26 | Ford Woodlotus |
| 27 | Trident Maple |
| 28 | Beale's Barberry |
| 29 | Southern Magnolia |
| 30 | Canadian Poplar |
| 31 | Chinese Tulip Tree |
| 32 | Tangerine |

**Table 5.1 List of Names of Leaves**

## 5.2 Data Feature Extraction

Feature extraction is part of the dimensionality discount method, in which, an preliminary set of the uncooked statistics is split and decreased to extra plausible groups. So while you need to method it'll be easier. The maximum critical feature of those massive statistics units is that they've a massive variety of variables. These variables require a number of computing assets to method them. So Feature extraction allows to get the great function from the ones huge statistics units with the aid of using choose and integrate variables into functions, thus, efficaciously lowering the quantity of statistics. These functions are clean to method, however nonetheless capable of describe the real statistics set with the accuracy and originality.

The 3 main features are

* Color Feature: Color Histogram is the most widely used technique for extracting the color feature of an image. It represents the image from a different perspective. It represents the frequency distribution of color bins in an image. It counts similar pixels and store it.
* Texture Feature: Texture is a feature used to partition photographs into areas of hobby and to categorize the ones areas. It gives records withinside the spatial association of colors or intensities in an image.  It is characterized with the aid of using the spatial distribution of depth tiers in a neighborhood.
* Shape Feature: Visual features of items are known as the form traits or visual features. For example, round item or triangular items or other shapes, perimeter boundary of the item, the diameter of the border and so on. The visual features confirmed intuitively are all belongs to form features.

Features of image are

**Shape features:**

* Area: The value of leaf area is easy to evaluate, just counting the number of pixels of binary value 1 on smoothed leaf image. It is denoted as A.
* Perimeter: Denoted as p, leaf perimeter is calculated by counting the number of pixels consisting leaf margin.
* Physiological\_length: The simplest human interfered component of our set of rules is which you want to mark the 2 terminals of the main vein of the leaf via mouse click. The distance among the 2 terminals is described because the physiological length. It is denoted Lp
* Physiological\_width: Drawing a line passing through the 2 terminals of the principle vein, you possibly can plot infinite strains orthogonal to that line. The wide variety of intersection pairs among the ones strains and the leaf margin is likewise infinite. The longest distance among factors of these intersection pairs is described on the physiological width .It is denoted by Wp
* Aspect\_ratio: Ratio between Physiological\_length to Physiological\_width ,thus Lp/Wp
* Rectagularity: It describes the similarity between leaf and the rectangle.It is defined as LpWp/A. Where A is Area of a leaf,Lp Physilogical\_length and Wp is the Physiological\_width
* Circularity: It is Calculated as ratio between Square of perimeter and the Area .It is defined as P\*\*2 / A.

**Color Features:**

* Mean\_r : Calculates the mean of the red color in the particular image
* Mean\_g: Calculates the mean of the values of the green color in the particular image.
* Mean\_b: Calculates the mean of the Blue color in the particular image
* Stddev\_r: Standard Deviation of red color. It is calculated using np.std() which is as inbuild method to calculate the standard deviation.
* Stddev\_g: Standard Deviation of green color. It is calculated using np.std() which is as inbuild method to calculate the standard deviation
* Stddev\_b: Standard Deviation of blue color. It is calculated using np.std() which is as inbuild method to calculate the standard deviation

**Texture Feature**

* Contrast : Contrast is the distinction in luminance or coloration that makes an item (or its illustration in an image or display) distinguishable. In visible belief of the actual world, contrast is decided through the distinction withinside the coloration and brightness of the item and different gadgets in the identical discipline of view.
* Correlation: Correlation is the process of transferring a clear out out masks regularly known as kernel over the image and computing the sum of merchandise at every location
* Inverse\_difference\_moments: Inverse difference moment: Inverse difference moment is the measure of local homogeneity.
* Entropy: The entropy or common facts of an image is a degree of the diploma of randomness withinside the image. The entropy is beneficial withinside the context of image coding : it's far a decrease restriction for the common coding duration in bits consistent with pixel which may be found out with the aid of using an most effective coding scheme with none lack of facts .

After calculating all the features, they are stored in vector for future calculations. This feature vector goes with feature scaling in the next step of this execution.

## 5.3 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
* 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.

## 5.4 Machine learning algorithm apply

Support Vector Machine(SVM) is a supervised machine mastering set of rules used for each category and regression. Support Vector Machines can be used in regression problems, though it is better suited for classification alone. The objective of SVM is to use a set of rules to discover a hyperplane in an N-dimensional area that enormously classifies the records points. The size of the hyperplane depends upon the quantity of capabilities. If the quantity of entered capabilities is two, then the hyperplane obtained is a line. If the quantity of entered capabilities is three, then the hyperplane will become a 2-D plane. It will become tedious to assume whilst the number of capabilities exceeds three. Let’s consider two independent variables x1, x2 and one dependent variable which is either a blue circle or a red circle.

****  **Fig 5.2 Identifying the right hyper-plane**

From the figure on top of its very clear that there are multiple lines (our hyperplane here could be a line as a result if we are considering solely 2 input options x1, x2) that segregates our knowledge points or will a classification between red and blue circles. However, we can choose the simplest line or generally the best hyperplane that segregates our knowledge

points. Select One reasonable choice as the best hyperplane is the one that represents the largest separation or margin between the two classes .choose the best hyper-plane:

**Fig 5.3** **Finding the best hyperplane**

So we choose the hyperplane whose distance from it to the nearest data point on each side is maximized. If such a hyperplane exists it is known as the maximum-margin hyperplane/hard margin. So from the above figure, we choose L2.

Here we have one blue ball in the boundary of the red ball. So how does SVM classify the data? It’s simple! The blue ball in the boundary of red ones is an outlier of blue balls. The SVM algorithm has the characteristics to ignore the outlier and finds the best hyperplane that maximizes the margin. SVM is robust to outliers

**Fig 5.4 Ignoring outlier data**

So in this type of data points what SVM does is, it finds maximum margin as done with previous data sets along with that it adds a penalty each time a point crosses the margin. So the margins in these type of cases are called soft margin. When there is a soft margin to the data set, the SVM tries to minimize (1/margin+∧(∑penalty)). Hinge loss is a commonly used penalty. If no violations no hinge loss. If violations hinge loss proportional to the distance of violation.

## 5.5 Grid Search CV:

In machine learning projects , we will train the data with different models and then we will choose the best model which gives the higher accuracy. One important factor in the performance of the models are their hyperparameters, once if we find appropriate values for these hyperparameters, the performance of modal can improve significantly.

GridSearchCV is the process of performing hyperparameter tuning in order to determine the optimal values for a given model .As mentioned above , the performance of the model depends upon values of hyperparameters. There is no way to know in advance the best values for hyperparameters so ideally. we need to try all possible values to know the optimal values. Doing manually it will take lots of time and resources and that is the reason we will use GridSearchCV to automate the tuning of parameters. It is function that comes from Sk-learn model\_selection package.

As mentioned above we will pass values of hyperparameters to the GridSearchCV function. we can do this with dictionaries in python. In this dictionary we will mention hyperparameters along with their values. In this project the tuning parameters are

parameters = [{'kernel': ['rbf'],

'gamma': [1e-4, 1e-3, 0.01, 0.1, 0.2, 0.5],

'C': [1, 10, 100, 1000]},

{'kernel': ['linear'], 'C': [1, 10, 100, 1000]}

]

This parameter dictionary will be given to the GridSearchCV as an parameter along with the standard machine learning modal

In this project, the model is Support Vector Machine. SVM parameters are ‘kernel’ ,’gamma’ ,’C’.

* Kernel:{‘linear’,’rbf’},default =’rbf’
* C: It is s regularization parameter. The strength of regularization proportional to C. Must be positive
* Gamma:{‘scale’ ,’auto’} default=’sacle’

Kernel coefficient for ‘rbf’, ‘poly’ and ‘sigmoid’.

If default is passed then it uses 1 / (n\_features \* X.var()) as value of gamma, if ‘auto’, uses 1 / n\_features.

## 5.6 Prediction:

This is final phase of the project where user gives input to the enhanced SVM modal for the prediction. Before predicting the label and remedies. The image should be pre-processed by removing unwanted data like remove background of the image. After removing background image this image will give to Modal which will gives output initially name of the leaf and remedies according to the leaf.

**CHAPTER 6**

# TESTING

Testing is a critical element which assures quality and effectiveness of the proposed system in (satisfying) meeting its objectives. Testing is done at various stages in the System designing and implementation process with an objective of developing an transparent, flexible and secured system. Testing is an integral part of software development. Testing process, in a way certifies, whether the product, that is developed, complies with the standards, that it was designed to. Testing process involves building of test cases, against which, the product has to be tested

## 6.1 Model evaluation in machine learning testing

Generally, Software testing includes

* Unit tests: The code will break down into blocks and each element will be tested separately
* Regression tests: They cover already tested software to see if it doesn’t suddenly break.
* Integration tests: how multiple components of the program work together.

Having evaluated the performance, we tend to still need to find out wherever and why the errors occur.

ML debugging could be a bit totally different from debugging the other computer code system. Poor quality of predictions created by an machine learning model doesn't essentially mean there's a bug. you have got to research a broader vary of causes than you'd in traditional programming: maybe it's the information that contains errors or hyperparameters aren't adjusted well. This makes debugging machine learning models quite challenging.

## 6.2 Model tests:

There are two kinds of testing

**Pre-train tests:**

This type of test is run early in the process to catch bugs before the model is run. They don't need to be run with any training parameters.

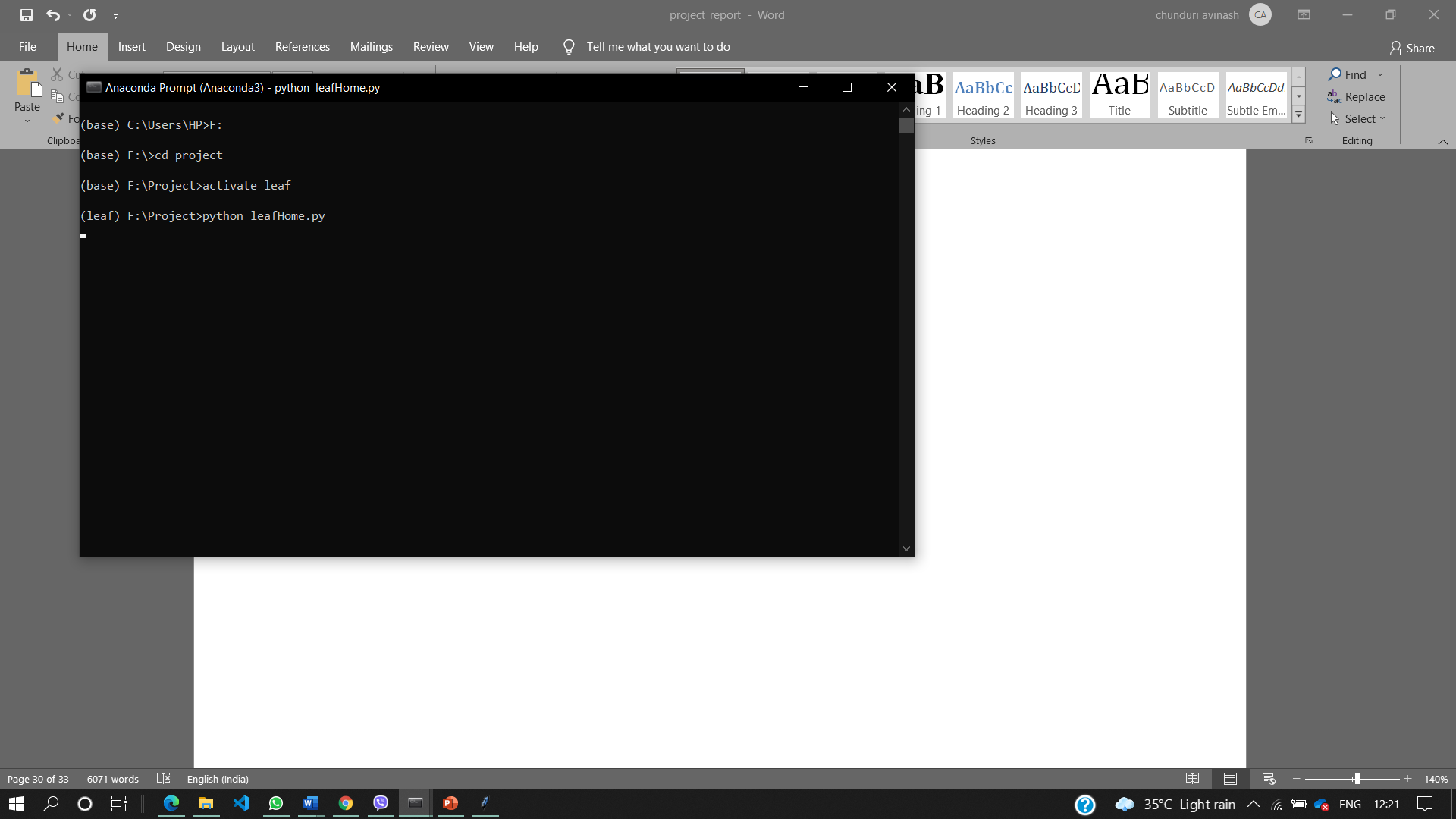
**Post-train tests:**

These tests are run on a trained model to see if it performs as expected. They encourage us to look into the logic behind the algorithm and see if there are any flaws. There are three types of tests that report on the program's behaviour

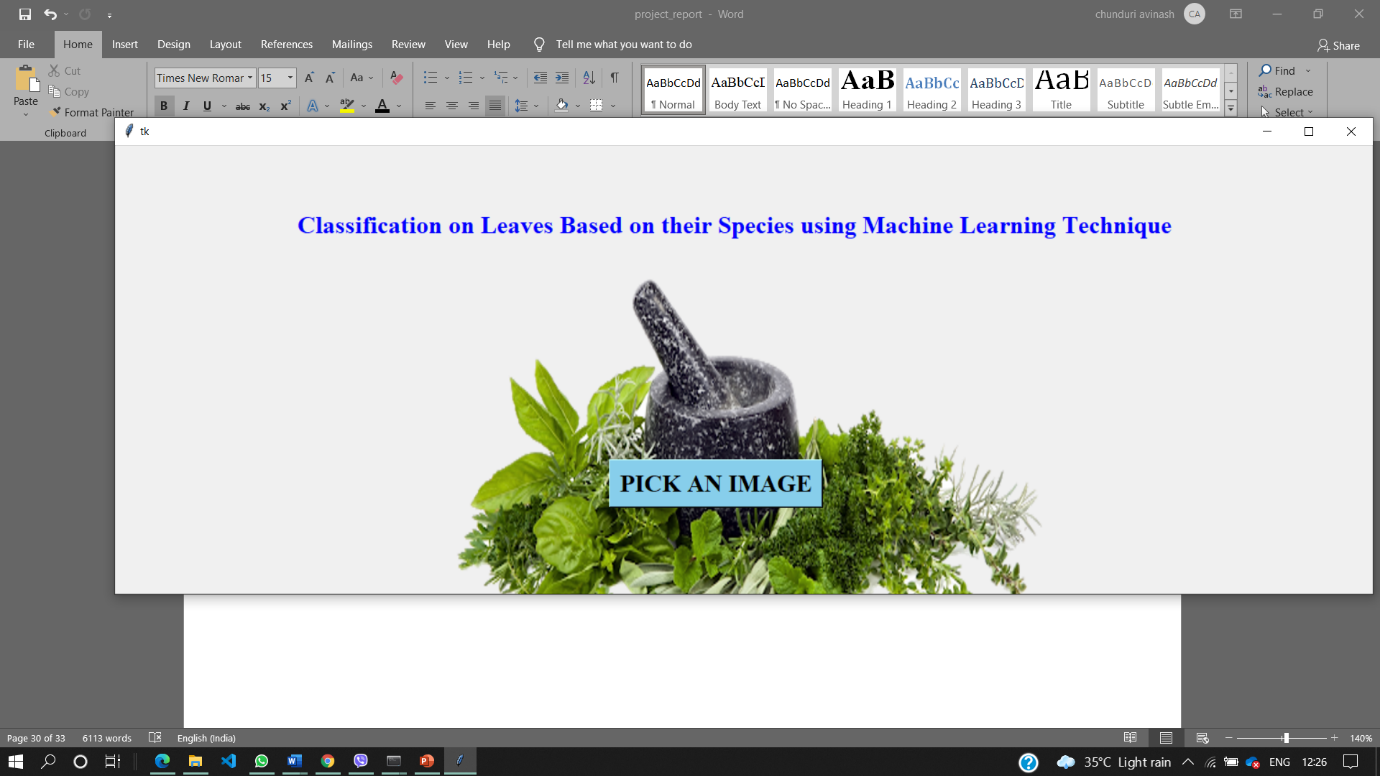
* **Invariance tests**: We can use invariance tests to see how often we can adjust the input without affecting the model's output. We may compare input examples to see if the predictions are consistent.
* **Directional expectation tests**: Unlike invariance tests, directional expectation tests are used to see how input perturbations affect the model's action.
* **Minimum functionality tests**: These tests, like conventional unit tests, enable us to measure each part of the software separately.

**CHAPTER 7**

# RESULT

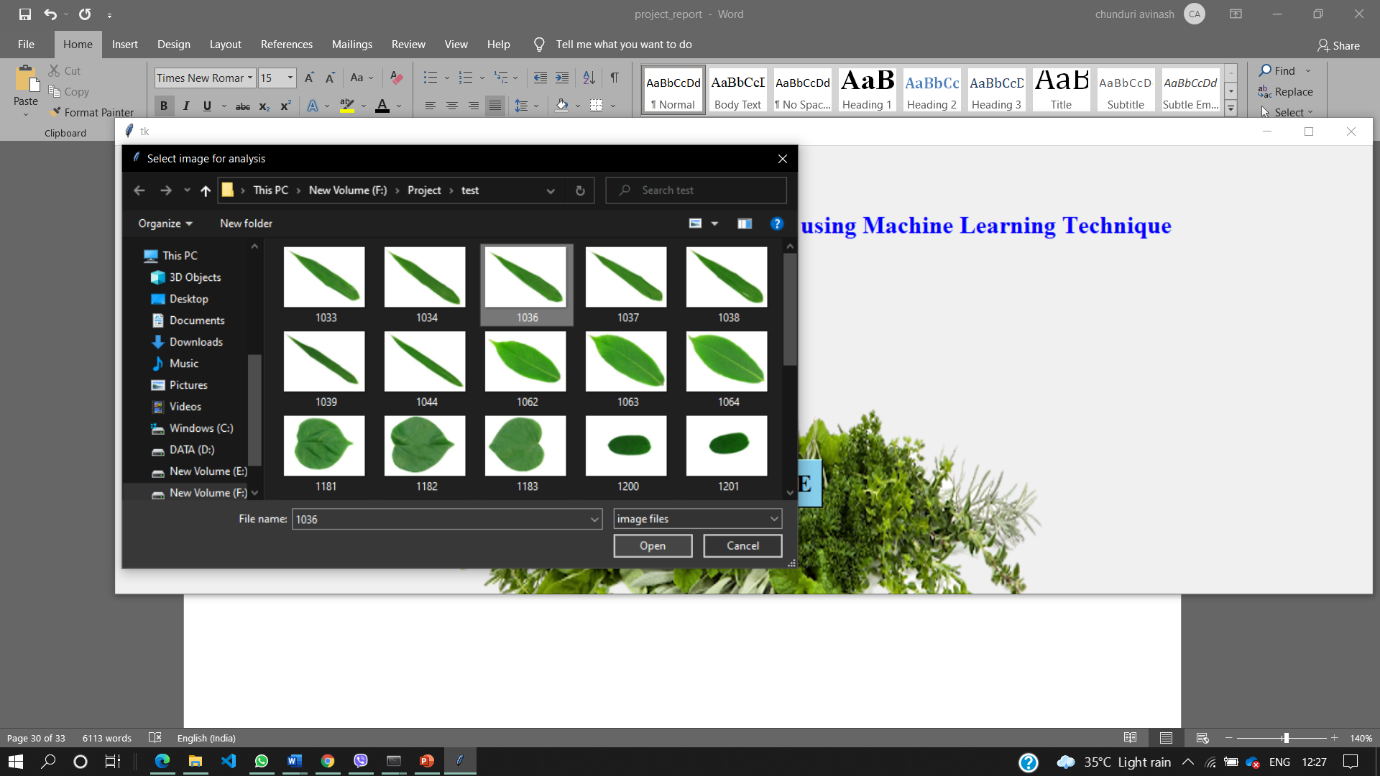
To run the project .Go to anaconda shell redirect to project directory and go to project folder. After redirecting to project folder activate the leaf environment. Until before the environment is base. By executing the main file using the command python leafHome.py

**Fig 7.1Starting the project**

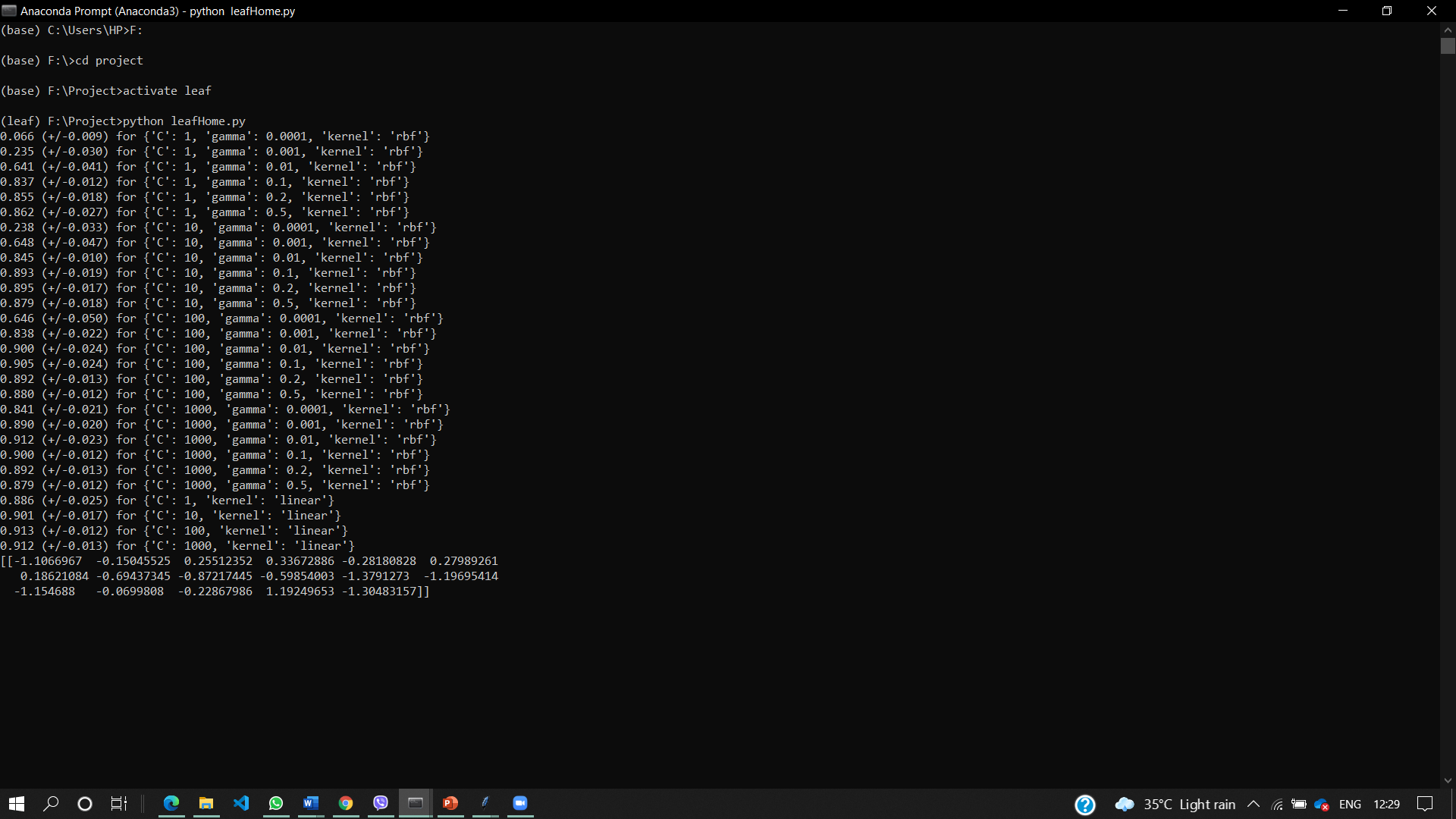
After executing the main file of the project that is leafHome.py the GUI interface will open as shown in the below figure where we have to pick an image for classifying.

**Fig 7.2 GUI interface for picking an image**

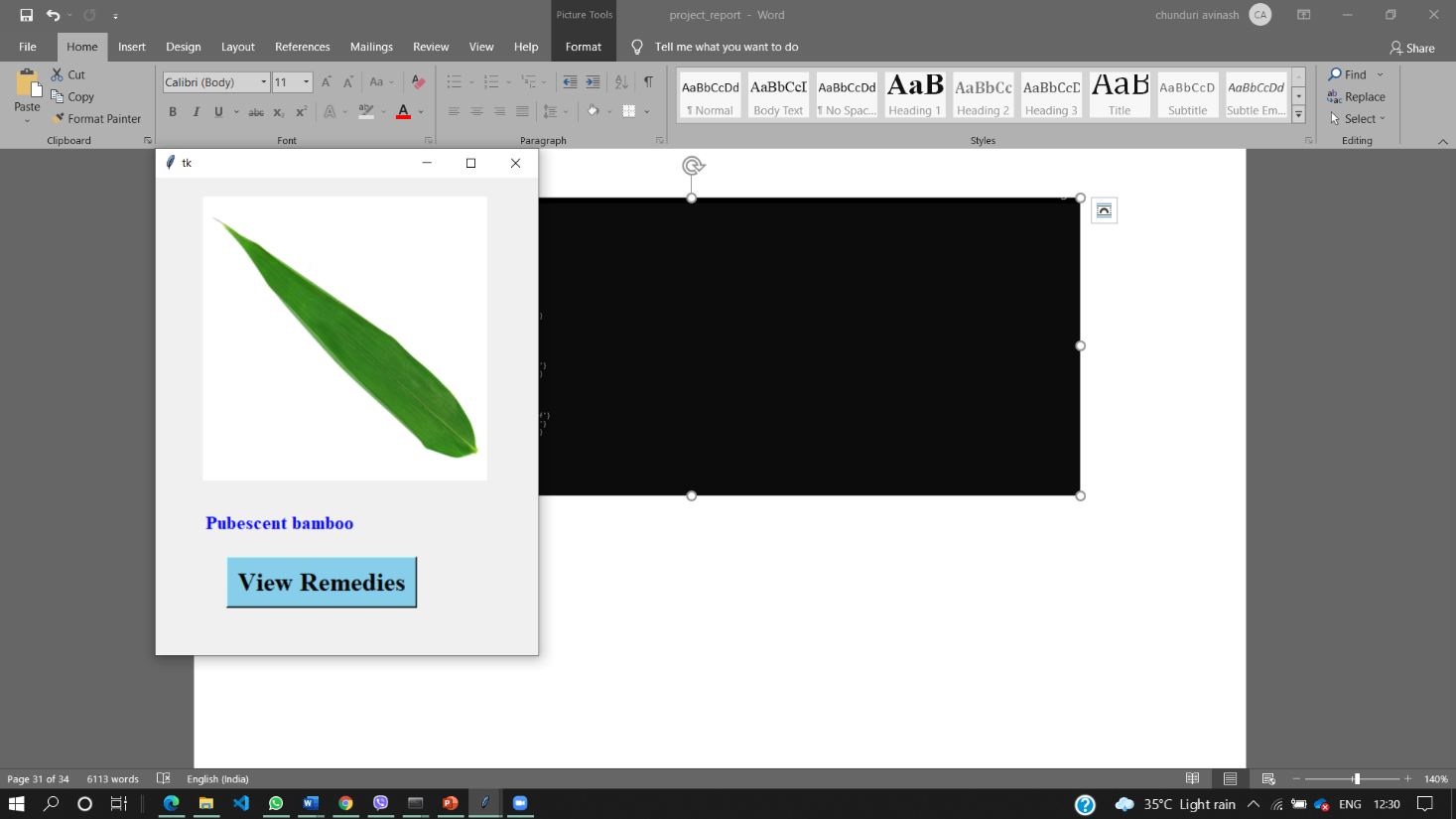
After clicking on the button PICK AN IMAGE it will open the folder of images as shown in the figure

**Fig 7.3 Picking up the image**

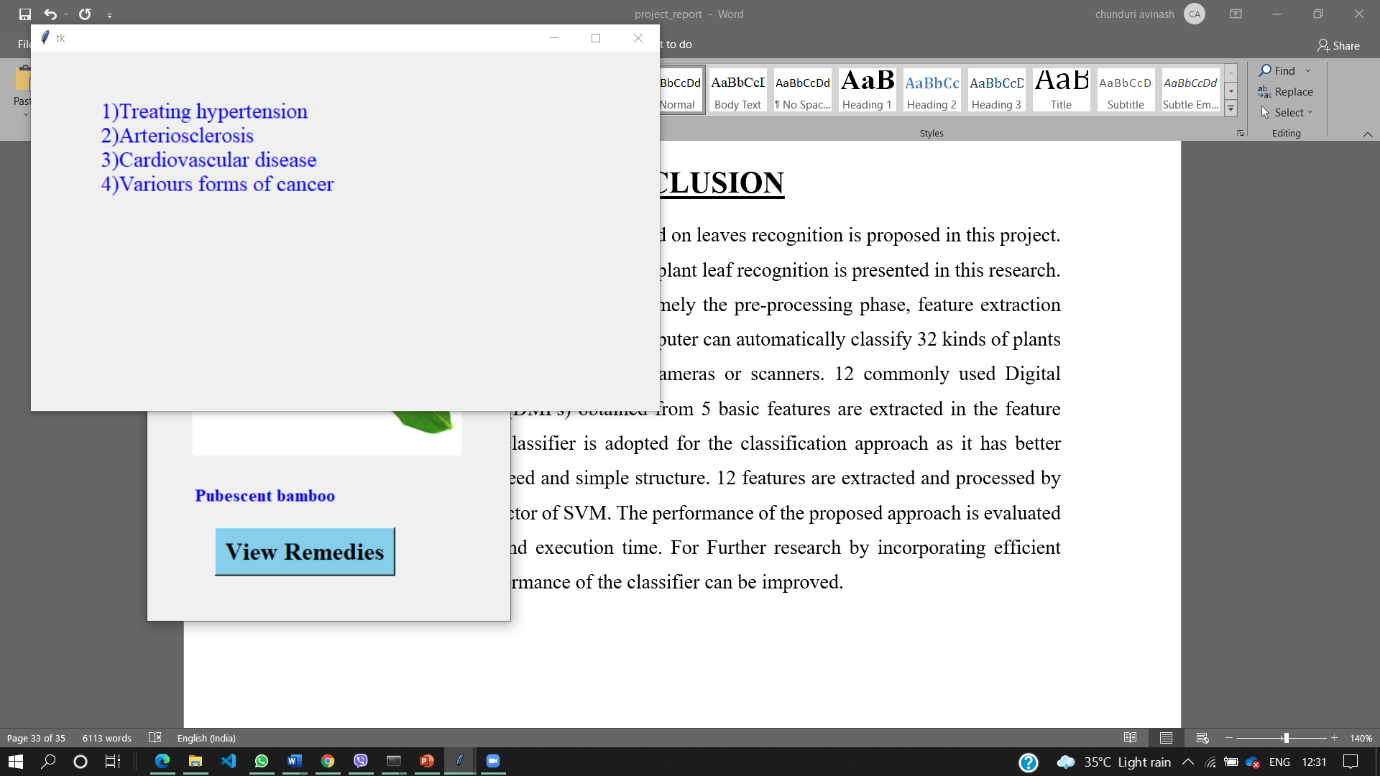
After picking an image the SVM model is trained with hyperparameters as shown in the figure. The modal will take the parameters which is giving higher accuracy with 91.3%



**Fig 7.4 Tuning the parameters**

After successful execution of the model. It will pop the window with the image and label for that image along with the Button with Remedies as shown in the figure

**Fig 7.5 Label of the selected image**

Once Label is found the remedies of the leaf will be displayed as shown in the figure which gives remedies of that leaf

**Fig 7.6 Remedies of the selected leaf**

# CONCLUSION

It is an approach of plant classification based on leaves recognition is proposed in this project. An efficient machine learning approach for plant leaf recognition is presented in this research. The approach consisted of three phases namely the pre-processing phase, feature extraction phase and the classification phase. The computer can automatically classify 32 kinds of plants via the leaf images loaded from digital cameras or scanners. 12 commonly used Digital Morphological Features (DMFs) obtained from 5 basic features are extracted in the feature extraction phase. SVM classifier is adopted for the classification approach as it has better accuracy, fast training speed and simple structure. 12 features are extracted and processed by PCA to form the input vector of SVM. The performance of the proposed approach is evaluated based on the accuracy and execution time. For Further research by incorporating efficient kernel functions the performance of the classifier can be improved.

# References

[1] Jyotismita Chaki, and Ranjan Parekh, “Plant Leaf Recognition using Shape based Features and Neural Network classifiers,” International Journal of Advanced Computer Science and Applications (IJACSA), 2011, vol. 2, no. 10.

[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]Plant classification using machine learning algorithm K.B.Shobana and P.Perumal, IEEE publications.

[10]CNN Approaches for Classification of Indian Leaf Species Using Smartphones : M. Vilasini and P. Ramamoorthy , tech Science Press.

[11]Plant Leaf Recognition Using a Convolution Neural Network: Wang-Su Jeon and Sang-Yong Rhee, ResearchGate.

[12]Review on Techniques for Plant Leaf Classification and Recognition: Muhammad Azfar Firdaus Azlah , Fakhrul Razan Rahmad , Farah Izana Abdullah and Sharifah Rafidah Wan Alwi.