Parametric Model for Flora Detection in the Middle Himalyas

A

Project Report

submitted in partial fulfillment of the
requirements for the award of the degree of

BACHELOR OF TECHNOLOGY
in
COMPUTER SCIENCE
Specialization in
Oil And Gas Informatics

By :

Name	**Roll No**
Ajay Manshani	R970216006
Ayush Bhatt	R970216020
Pranshul Kandwal	R970216081
Soumya Nigam	R970216094

Under the guidance of

Dr. Aviral Sharma
Assistant Professor
Department of Informatics

UNIVERSITY WITH A PURPOSE

Department of Systemics
School of Computer Science
UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
Bidholi, Via Prem Nagar, Dehradun, Uttarakhand

2018-19

CANDIDATES DECLARATION

We hereby certify that the project work entitled **Parametric Model for Flora Detection in the Middle Himalyas** in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science And Engineering with Specialization in Oil and Gas Informatics and submitted to the Department of systemics at School of Computer Science, University of Petroleum And Energy Studies, Dehradun, is an authentic record of our work carried out during a period from **JANUARY, 2019** to **MAY, 2019** under the supervision of **Dr. Aviral Sharma, Assistant Professor, Department Of Informatics**.

The matter presented in this project has not been submitted by us for the award of any other degree of this or any other University.

Name	**Ajay Manshani**	**Ayush Bhatt**	**Pranshul Kandwal**	**Soumya Nigam**
Roll No.	**R970216006**	**R970216020**	**R970216081**	**R970216094**

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(Date: 2nd May 2019)

Dr. Aviral Sharma
Project Guide

Dr.Thipendra Pal Singh

Head
Department of Systemics
School of Computer Science
University of Petroleum And Energy Studies
Dehradun - 248 001 (Uttarakhand)

ACKNOWLEDGEMENT

We wish to express our deep gratitude to our guide **Dr. Aviral Sharma**, for all advice, encouragement and constant support he has given us through out our project work. This work would not have been possible without his support and valuable suggestions.

We sincerely thank to our Head of the Department, **Dr. Thipendra Pal Singh** , for his great support in doing our **Parametric Model for Flora Detection in the Middle Himalyas** at **SoCS**.

We are also grateful to **Dr. Manish Prateek Professor and Director SoCS**, UPES for giving us the necessary facilities to carry out our project work successfully.

We would like to thank all our **friends** for their help and constructive criticism during our project work. Finally we have no words to express our sincere gratitude to our **parents** who have shown us this world and for every support they have given us.

Name	**Ajay Manshani**	**Ayush Bhatt**	**Pranshul Kandwal**	**Soumya Nigam**
Roll No.	**R970216006**	**R970216020**	**R970216081**	**R970216094**

ABSTRACT

With technological advancement and growing computational capabilities people are relying on technology to possess the information. This information can be related to any stream ranging from tech material such as cloud computing, Block chain, Big Data etc. to the non-tech material such as health sciences, biodiversity, economics, Human Resource etc. With the help of technology people can easily access the information related to their field of interest. In this project we are building a parametric model to detect which family a plant belongs to thus narrowing down the biodiversity. It will distinguish between different families of plants on the basis of various parameters like characteristic of leaves, height of plant, color of barks, inflorescence and many more. The machine will be trained to tell a plants family on the basis of the parameters that it will record from the user input. The user will be responsible to give reliable input which can be processed by the machine. This parametric modelling will be very helpful for forest guards, students and professors involved in the research related to the flora, and also for the common people.

Keywords: Parametric modeling, Machine Learning, Biodiversity, flora

TABLE OF CONTENTS

Contents

LIST OF FIGURES

List of Figures

1 Introduction

The present era is been led by image processing and the machine vision which has enabled us to foresee things and attempts to integrate existing technologies in new ways and apply them to solve real world problems. The development of image and video processing algorithms and approaches have received significant attention from the scientific community with the increasing hardware sensing capabilities and the increased quality of the digitally recorded materials. The increased computing capabilities have allowed the development of more complex and powerful approaches. A large increase has been observed in the potential applications that could use the various machine learning algorithm where the machine can be trained to perform various tasks and give unexpected results.The number of industries that could use the advancement of image processing and machine learning for the purpose of object recognition or segmentation from images is quite large. Some of the applications range from medical applications,agriculture, quality inspection to military industry.The ultimate goal of building these applications is to utilize the capabilities of machine learning and vision for process automation in order to increase productivity and reduce human efforts to facilitate everyday tasks. The multiple tasks that could utilize the capabilities of machine vision include object recognition in images, image segmentation to extract region of interests, image database research etc .
Plant specie recognition is one of the complex tasks to be performed in everyday life which involves a number of challenges given the diverse biodiversity present. The complexity increments exponentially if the given job is to be performed for the Himalayan region endowed with a rich variety of gene pools and species,Stretched over an area of four million square kilometer. This biodiversity is huge and more diverse than certain countries in the world. The artificial intelligence and machine vision thus finds an potential application in this area. The recent times have observed a significant expansion and increment of various approaches that could be useful for play species recognition, detection of weeds in plants etc.The research in these topics have the final goal of fully automating the agricultural process beginning from the detecting the specie itself.While the final goal is still to be achieved, significant advancements in the area have been made. The proposed algorithm needs to be precise and should increase productivity with least margin of error possible.

2 Literature Review

- Supervised learning we need classes. That means we need a qualitative attribute that can take a finite set values, so we can say that supervised learning is the inference of a function labelled training data.
  Decision TreeFirst decision tree developed by J. Ross Quinlan known as ID3 (Iterative Dichotomize), then Quinlan presented C4.5 as successor of ID3, which became a benchmark to which newer supervised learning algorithms are often com- pared. In 1984, Breiman, Friedman, Olshen, and Stone published the book Classification and Regression Trees (CART).

- Naive BayesNaive Bayes classifiers belong to a family of simple probabilistic classifiers based on applying Bayes theorem with strong independence assumptions between the features. n machine learning, naive Bayes classifiers are a family of simple ”probabilistic classifiers” based on applying Bayes’ theorem with strong independence assumptions between the features. Naive Bayes has been studied extensively since the 1960sIt has been studied extensively since the 1950s. It used for first time in the domain of text retrieval in the early 1960s, and it became a popular method for text categorisation. Because of good results given by this algorithm.

- K Nearest NeighbourK Nearest Neighbour is a lazy algorithm requires less computation time during the training phase than eager-learning algorithms (such as decision trees, neural and Bayes nets) but more computation time during the classification process. It is based on distance calculation, Aha (1997) and De Mantaras and Armengol (1998) presented a review of instance-based learning classifiers. in other words, k-Nearest Neighbour (kNN) is based on the principle that the instances within a dataset will generally exist in close proximity to other instances that have similar properties (Cover and Hart, 1967).

- Support Vector Classifier (SVC) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. We perform classification by finding the hyper-plane that differentiate the two classes very well.

- Multilayer perceptron (MLP) is a class of feedforward artificial neural network. A MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function.

3 Problem Statement

Develop an AI which can be utilized by forest guards and general population interested in tree recognition.

4 Objective

To develop AI for recognition of trees in the Chakrata region of the middle Himalayas.

5 Design Methodology

Working Methodology of system is as follows.....

1. Domain knowledge collection
2. Data collection and dataset collection.
3. Predictive Model building.
4. GUI and middle layer building.

6 Implementation

6.1 Pseudo Code

SUPPORT VECTOR CLASSIFIER

1. Get the input in the following manners training set (x1,y1),(x2,y2).....(xn,yn) Test set S1,S2.....Sn
2. Creat a dataset for all the input parameter say c1,c2....cn
3. Load the dataset and extract the input output in matrix form
4. Use the SVM.CSV as a inbuild function
5. Fit the model
6. Validate the model
7. Make pridiction
8. check for training and Testing accuracy.

K-NEAREST NEIGHBORS

1. Get the input in the following manners training set (x1,y1),(x2,y2).....(xn,yn) Test set S1,S2.....Sn
2. Creat a dataset for all the input parameter say c1,c2....cn
3. Load the dataset and extract the input output in matrix form
4. Implement the nearest Neighbour using ball tree
5. Fit the model
6. Validate the model
7. Make pridiction
8. check for training and Testing accuracy.

NAIVE BAYES

1. TrainMultinomialNB(C,D)
   V¡-ExtractVocabulary(D)
   N¡-countDocs(D)
   for each c in C
   Nc¡-CountDocsIn Classes(A,c)
   prior[c]¡-Nc/Count(C)
   textc¡-TextOfAllDocsInClass(D,c)

2. for each t in V Ftc¡-CountOccurrencesOfTerm(t,textc)
   for each t in V
   condprob[f][c]¡-(Ftc+1)/sum(Ft'c+1)

3. return v,prior,condprob

DECISION TREE

1. S, where S= Set of classified instances
   Output: Decision tree
   Require: S is not equal to null , num attributes¿0
2. procedure BUILDTREE
3. repeat
4. maxGain¡–0
5. splitA¡–null
6. e¡–Entropy(Attributes)
7. for all Attributes a in S do gain ¡—InformationGain(a,e) if gain ¿ maxGain then maxGain ¡–gain splitA ¡–a
8. end if
9. end for
10. Partition(S,splitA)
11. until all partitions processed
12. end procedure

6.2 Output Screen

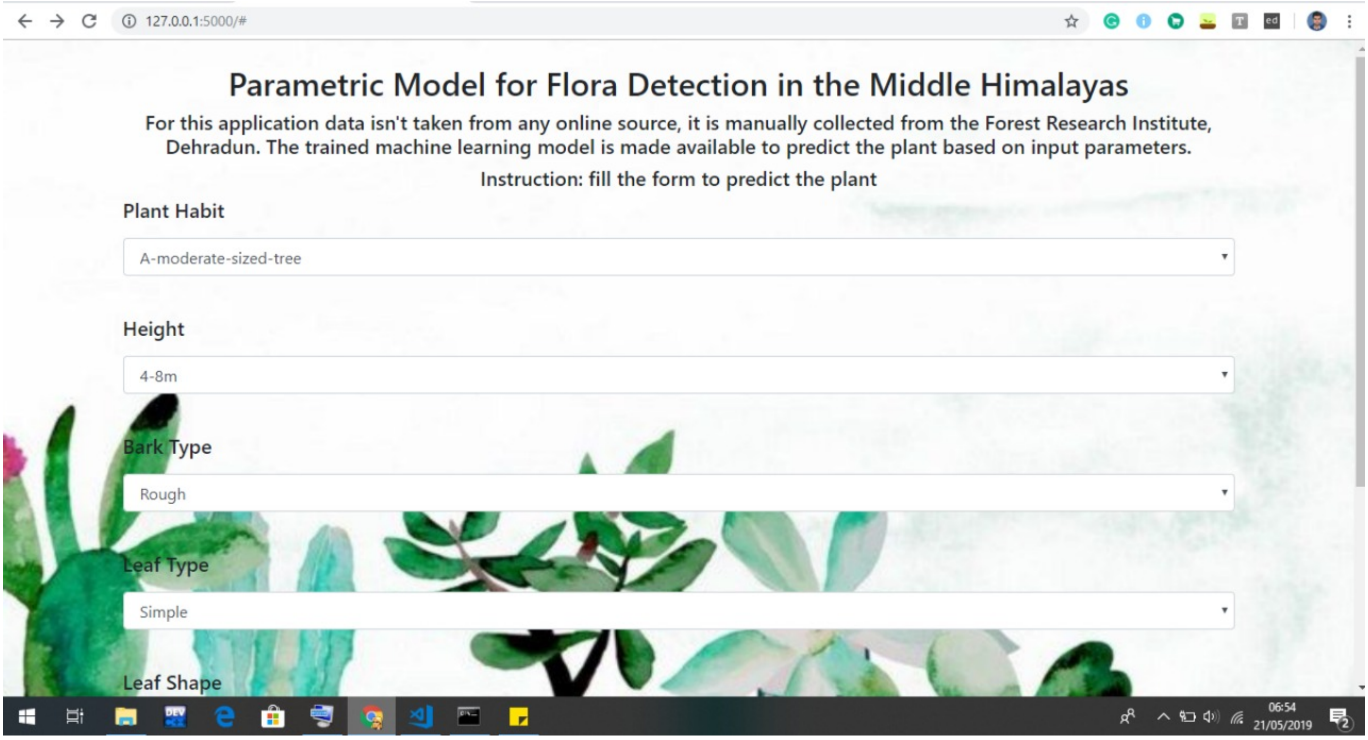

Figure 1: *output 1*

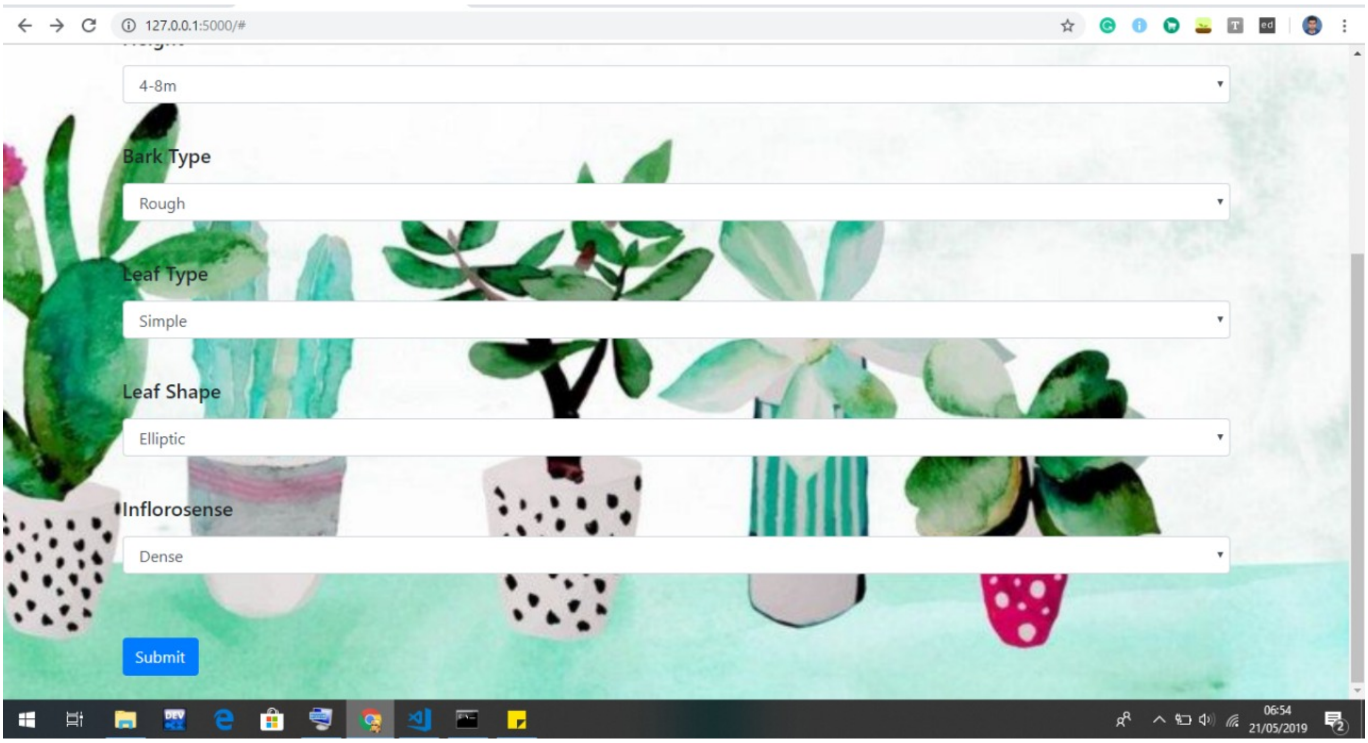

Figure 2: *output 2*

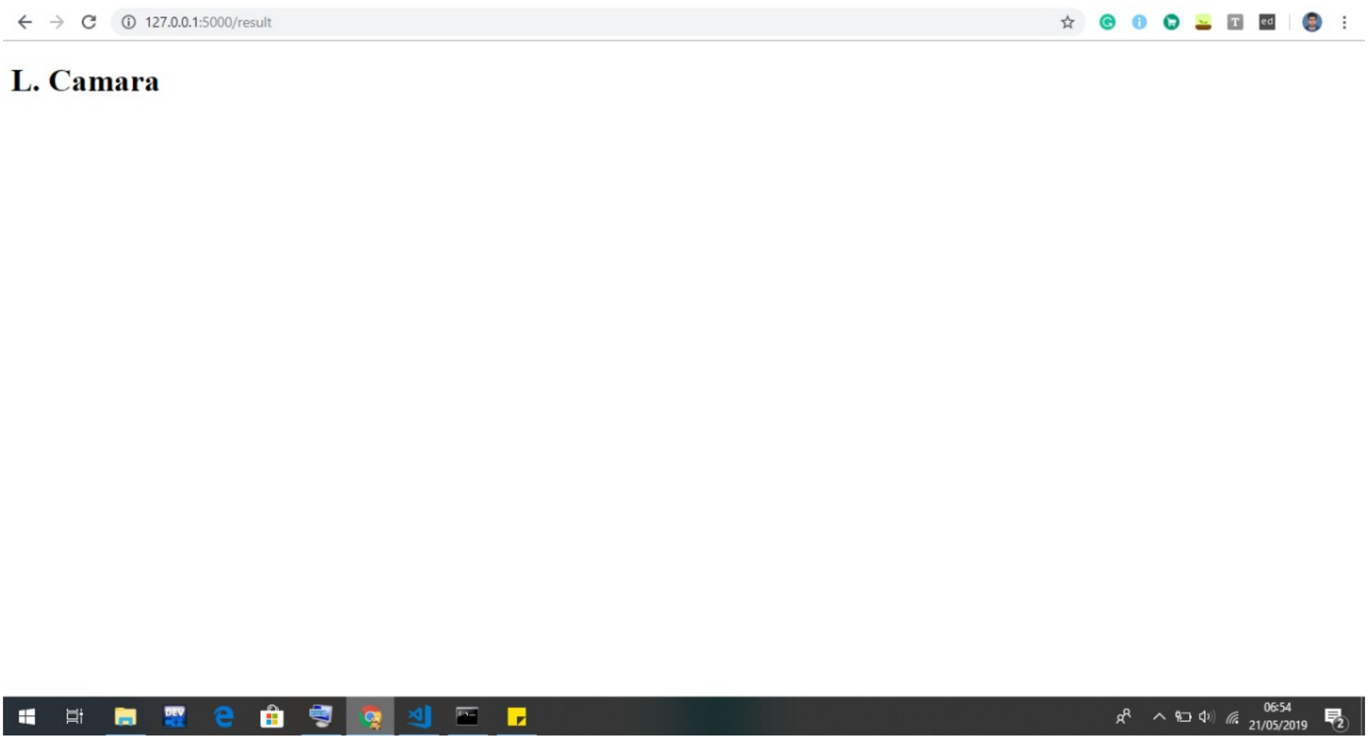

Figure 3: *output3*

6.3 UML Diagrams

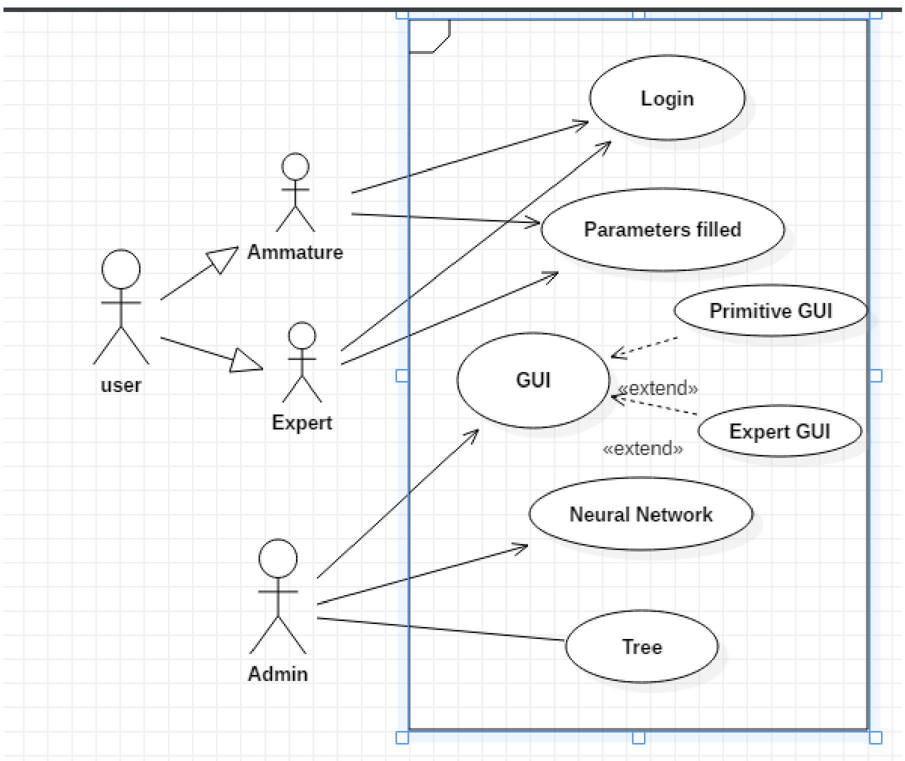

Figure 4: *Use Case Diagram*

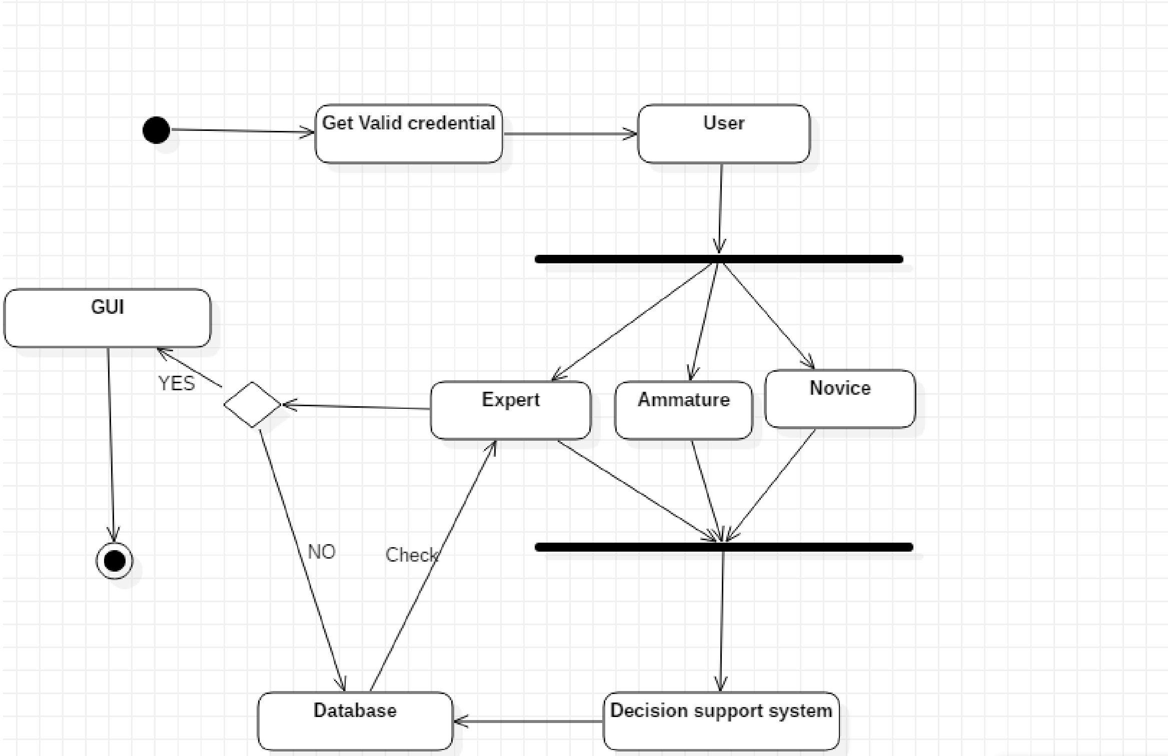

Figure 5: *Activity Diagram*

7 System Requirements (Software/Hardware)

- Hardware Interface:
  - 64 bits processor architecture. .
- Software Interface:
  - Anaconda with python 3.7
  - Visual Studio Code

7.1 Result Analysis

SVC:

Scietific Name	Common Name	Training Sample	No. of Incorrect Recognition
S.Robusta	Sal Tree	10	9
F.Ramontchi	Batoko Plum	10	0
C.Fistula	Golden Shower	10	0
M.Philippinensis	Kumkum Tree	10	0
A.Cordifolia	Baby Son Rose	10	0
C.Infortunatum	Hill Glory	10	0
A.Vasica	Adhatoda	10	0
L.Camara	Tickberry	10	0
F.Religiosa	Peepal Tree	10	0
Murraya	Marraya	10	0

MLP:

Scietific Name	Common Name	Training Sample	No. of Incorrect Recognition
S.Robusta	Sal Tree	10	10
F.Ramontchi	Batoko Plum	10	10
C.Fistula	Golden Shower	10	10
M.Philippinensis	Kumkum Tree	10	10
A.Cordifolia	Baby Son Rose	10	10
C.Infortunatum	Hill Glory	10	10
A.Vasica	Adhatoda	10	0
L.Camara	Tickberry	10	10
F.Religiosa	Peepal Tree	10	10
Murraya	Marraya	10	10

KNN:

Scietific Name	Common Name	Training Sample	No. of Incorrect Recognition
S.Robusta	Sal Tree	10	1
F.Ramontchi	Batoko Plum	10	0
C.Fistula	Golden Shower	10	0
M.Philippinensis	Kumkum Tree	10	0
A.Cordifolia	Baby Son Rose	10	0
C.Infortunatum	Hill Glory	10	0
A.Vasica	Adhatoda	10	0
L.Camara	Tickberry	10	0
F.Religiosa	Peepal Tree	10	0
Murraya	Marraya	10	0

Decision Tree:

Scietific Name	Common Name	Training Sample	No. of Incorrect Recognition
S.Robusta	Sal Tree	10	1
F.Ramontchi	Batoko Plum	10	0
C.Fistula	Golden Shower	10	0
M.Philippinensis	Kumkum Tree	10	0
A.Cordifolia	Baby Son Rose	10	0
C.Infortunatum	Hill Glory	10	0
A.Vasica	Adhatoda	10	0
L.Camara	Tickberry	10	0
F.Religiosa	Peepal Tree	10	0
Murraya	Marraya	10	0

Naive Bayes:

Scietific Name	Common Name	Training Sample	No. of Incorrect Recognition
S.Robusta	Sal Tree	10	10
F.Ramontchi	Batoko Plum	10	0
C.Fistula	Golden Shower	10	0
M.Philippinensis	Kumkum Tree	10	0
A.Cordifolia	Baby Son Rose	10	0
C.Infortunatum	Hill Glory	10	0
A.Vasica	Adhatoda	10	0
L.Camara	Tickberry	10	0
F.Religiosa	Peepal Tree	10	0
Murraya	Marraya	10	0

8 Conclusion and Future Scope

This application finds its usage not only for the people who belong to the communities related to plant sciences but also find its applications for all strata of our society. In future the inclination is towards expanding the data set as much as possible. A major part of the application is related to collaborating with the various institutions related to plant science which will not only help in the collection of domain knowledge but can also act as a major platform where the application can find its academic use as it can be used by the students and researches of these universities.

9 Refrences

1. Abu A, Lim LHS, Sidhu AS, Dhillon SK. 2013a. Biodiversity image retrieval framework for monogeneans. Syst Biodivers. 11(1):1933. doi: 10.1080/14772000.2012.761655[Taylor and Francis Online], [Web of Science], , [Google Scholar]

2. Abu A, Lim LHS, Sidhu AS, Dhillon SK. 2013b. Semantic representation of monogenean haptoral bar image annotation. BMC Bioinf. 14(1): 48. doi: 10.1186/1471-2105-14-48[Crossref], [PubMed], , [Google Scholar]

3. Berg CC, Corner EJH, Jarrett FM. 2006. Moraceae-genera other than Ficus: Flora Malesiana-Series 1. Spermatophyta. 17(1):1146. [Google Scholar]

4. Boser BE, Guyon IM, Vapnik VN. 1992. A training algorithm for optimal margin classifiers. In: Haussler D, editor. 5th annual ACM workshop on COLT. Pittsburgh (PA): ACM Press; p. 144152.[Crossref], , [Google Scholar]

5. Canny J. 1986. A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell. PAMI-8:679698. doi: 10.1109/TPAMI.1986.4767851[Crossref], [Web of Science], , [Google Scholar]

6. Chapelle O, Haffner P, Vapnik VN. 1999. Support vector machines for histogram-based image classification. IEEE Trans Neural Netw. 10:10551064. doi: 10.1109/72.788646[Crossref], [PubMed], [Web of Science], , [Google Scholar]

7. Clark JY, Corney DP, Tang HL. 2012. Automated plant identification using artificial neural networks. CIBCB 2012. Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology; May 912; San Diego, CA: IEEE. doi: 10.1109/CIBCB.2012.6217215[Crossref], , [Google Scholar]

8. Coltelli P, Barsanti L, Evangelista V, Frassanito AM, Gualtieri P. 2014. Water monitoring: automated and real time identification and classification of algae using digital microscopy. Environ Sci Process Impacts. 16:26562665. doi: 10.1039/C4EM00451E[Crossref], [PubMed], [Web of Science], , [Google Scholar]

References