Flower Species Classification using Random Forest Classifier

Ravi Verma[1] Prof. J C Achutha [2]

1 MCA Final Year , 2 Asst. Professor

1,2 Department of MCA , The Oxford Colleg of Engineering

Bommanhalli , Hosur Road, Bangalore – 560068

ABSTRACT - This project introduces an approach of automated classification of flower species using the flower images as input, which is based on the characterization of the rgb (red green blue) color histogram and its texture properties along with the computer vision and machine learning algorithms. It uses the random forest classifier decision trees, python's sci-kit machine learning library, matplotlib for plotting image histograms and python's package, numpy for scientific computing. We randomly take out some block of each flower species image as a training set and another block of the same flower species images as a testing set, using which a decision tree is built which classifies all the species whenever a new query of classification of the images arises.

The Random Forest Classifier decision trees and the python's sci-kit machine learning library give very high accuracy rate of classification which is superior to any other methods of classification. The results of this can give an accuracy of correct recognition of the flower species of up to 98 percent, which is higher than any other methods of classification of flower species. Users of the system can use it for recognition of various flowers or they can pass a compressed folder containing different flower images and get the images classified accordingly to the species they belong, the user can also view the histogram of the image or search for the similar type of images available on the internet.

Keywords: flower species classification, random forest classifier, sci-kit library, numpy.

I. INTRODUCTION

Classification of different objects is one the biggest advantages animals have and out of the all the animals, humans have the best ability to classify between different things in the world. We classify different objects based on our perception of its shape, smell, size, color, uses. This ability of classification has helped us in many ways to understand our environment and build and create new things which have made significant impact on the lives of humans. But classification of things which are very abundant and look very similar with slight changes is complex. To classify such things we need to have vast knowledge of those particular things.

Flowers not only have decorative purpose but they also have many significant uses which are very important for animal survival including humans. They support basic functions of our daily lives by providing food, nutrition in turn, and also many flowers have medicinal use.

Flowers with plants are one the largest consumers of carbon dioxide and producer of oxygen, they also have

significant impact on the bee population, which are very important for the plants themselves since they help them in the process of pollination. They have very important role in our lives and they will continue to have their important role in the future as well. Thus the knowledge of different types of flower species is very important for botanists in order to study different behaviors of different plants and flowers and their role in the field of medicine, cosmetic, food.

Image classification thus has significant role in our society, and flower classification is one of them. Study of flowers also helps botanists to influence the fields of medicine, cosmetics, science among others. There are flowers with varying shapes, colors, size. Generally we recognize flowers by their smell, color, the shape of their petals, and also our knowledge about the different types of flowers available in the environment. The figure 1.1 shows significantly varying types of the flowers species with different color pattern, texture, size, and different shapes of their petals, with significant differences in their overall appearance. But there are also some flowers which have same color, same size, and similar petals shape; which makes it even harder for accurate classification of flowers. Thus with all those factors involved we get to notice that classification of flowers is a very complex process and if do not have sufficient knowledge of various types of flowers available in our environment, the process of classification gets further hard.

Fig 1.1. Various types of flowers species.

Our project makes this classification easier by building a

system which can automatically classify flowers based on the type of the species of the flower. This is done using the various machine learning libraries available in python such as sci-kit library, numpy for arithmetic manipulation on

images. We use numpy to store the images as arithmetic numbers in an array; making it possible for us to do manipulation on the image using the array directly. The random forest classifier available in the sci-kit library helps to build decision trees. We train the random forest classifier by using a sample set of each flower species images as a training set. Another set of sample images of each flower species as the testing set, to make the accuracy of the recognition very high.

A user of the system has the abilities to identify the type of flower by using the image of the flower and passing the image as the query to the system. The system then using its learned behaviors tries to identify the flower and gives back the result i.e. the flower name and its details. The system can also take huge set of images in the form of compressed folder containing different images of flower which have to be organized by their species; the output then is the set of folders which contain flower images based on the type of species of flowers. A user can also get the histogram of an image which is very useful to identify the major components of color intensities the image is composed of. The histogram is plotted using the matplotlib which is used for plotting data. The process of identification of the type of flower is done using the various libraries available in python some of which are opencv2, sci-kit, numpy, matplotlib.

The accuracy of the identification of the image depends on the type of training the random forest classifier has been given. It is done by taking a random set of a block of images of flowers of each species for training and using another set of block of random images of a particular flower species as the test. If the accuracy of the test is high then the system will have a very high accuracy of identification of flower species. After the training and testing is applied to the classifier; to identify the type of flower we just need to pass an image of a flower as query to the system, we can also search for similar types of flower by passing the image as query.

II. Existing and Proposed System

Existing System - Usually the classification of flowers can be done by the two obvious methods, one is the traditional way of the knowledge of flowers and plants using which a person can tell what type of the flower species a particular is, but that is very hard since a person can't knows a certain popular flower species like rose, lily, lotus, sunflower which are very common. But the flowers which are not found abundantly in the environment can be hard to classify. So the person can take the help of specialist (botanist in this case) to recognize the flowers. The other method being the use of technologies which are available today, generally being expensive and may take time before the flower is identified. Below we discuss the two methods.

Available Technologies - By making use of available technologies which extract important features from an image, whose description, interpretation and understanding is provided by complex machine learning algorithms whose precision of recognition is not very close and which are generally expensive and also take time before the flower is identified and a response is sent. Also these types of technologies with high accuracy are generally rare.

Traditional Methods - The other method of classification of the flowers is the manual process of recognition based on our perception, smell, texture which might not be accurate enough to classify the plants correctly. On the other hand we can also use the help of a person who is in the field of botany, which is time consuming, tedious, and might also be expensive.

Proposed System - We propose a system where the flowers are classified automatically based on the rgb properties of an image whose color histogram are created using matplotlib in python and the random forest classifier's decision trees, using the computer vision and machine learning algorithms. Thus the system makes the process of the classification of flower species fast, easy and reduces the cost significantly and can be used in the study of botany by scientists or the students of botany without the hassle of knowing details of the flower species. The system takes an image of flower and with a very high accuracy rate recognizes the species of flower and gives the details of the flower to the user of the system. The major steps in the proposed system are:

- ➢ Capturing an image of a flower as input for the system.
- ➢ Recognizing the image based on the learned methods from previous data, and from the training and testing data.
- ➢ Creating a histogram of the image and using its texture properties.
- ➢ Plotting the histogram using the matplotlib.
- ➢ Training the random forest tree using a set of images of flowers of each species.
- ➢ Testing the random forest tree using another set of images of flowers of each species.
- ➢ The resulting system recognizes each flower with precise accuracy.
- ➢ The details of the flowers species such as its name, species, family, genus is generated based on the type of the flower species.

The system will make it easy for the classification of flowers, organization of different flower species image, which would be very difficult if it is done manually, requiring the help of botanist who has the knowledge of different flower species in order to classify, identify, and organize.

III. Feasibility Study

Feasibility study is the preliminary investigation about the

feasibility of the project and the practicality of the proposed project. It is done before the project is initiated. The feasibility study of the project was done on the following areas:

Technical Feasibility - The project deals with machine learning algorithms and hence it makes it harder for building a system which is completely accurate. The technology required to achieve this has to be able to do complex manipulations on images, complex calculations and a high computing power is required. To do this we make use of free to use libraries of python such as numpy – to do arithmetic computations on images, opencv – to make it easy for accessing and dealing with images, matplotlib – to plot graphs and histograms of images based on the rgb characteristics, scikit learn available in python which handles the machine learning algorithms such as the random forest classifier algorithm is very useful to built this kind of applications. The hardware which is required for the user to use the system is just a camera which can take pictures of good quality, a standalone camera or a mobile phone camera does the job. The rest of the job of processing, analyzing and classification is handled by the algorithms and libraries available in the python programming language and its scikit learn library.

Economic Feasibility - The system is being developed mainly for the applications in the field of botany. A botanical firm which deals with identification of new species or classification of species of flowers needs to have expert botanists who are very good at these. This can be expensive for the organization since the manual classification and identification done by those botanists can be expensive. Our system reduces the cost significantly which is very beneficial for the organization.

Operational Feasibility - The system can solve the problem of identification and classification of different flower species with a very high accuracy using the technology available in python. This can significantly increase the productivity, of a botanical organization which deals with classification and identification of different flower species. Instead of traditional manual way analyzing each and every single image by an expert botanist, the system does it very fast with a great accuracy and precision, as it does the job of identification and classification it also learns and greatly increases it accuracy using the random forest decision trees available in scikit learn.

Target Markets - The system is developed in respect to the needs from the perspective of a botanical organization which includes the tasks of identification of new flower species, classification of flower species. The market for this system can also be the general population who are interested in exploration of new kinds of flowers, for which the system can be transformed and extended as a mobile application also, so that it is very easy for targeting that particular population of explorers, hobbyists, since everyone will generally have access to a smart phone and an app for identifying flowers on the go will be very helpful to them.

Legal Feasibility - The system holds no copyright of the images which are shown at the search results and if it is used by the users elsewhere on the internet then, it has to be used based on the authorization of the owner of the image who holds the copyrights of the image.

IV. Implementation
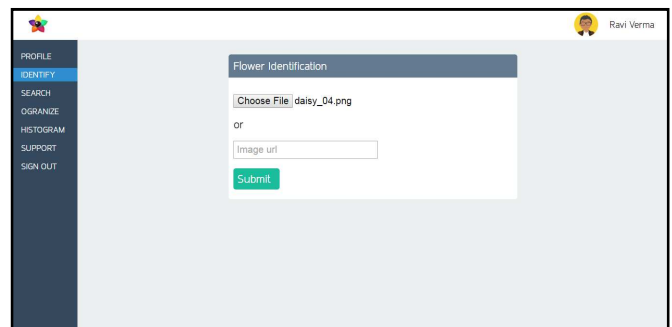
Img 0. Identification page.

The identification page allows the user to submit an image of a flower as the input for its identification of the type of flower species it belongs to. It is done by passing the image from the device itself or from a url of the image location. If the image submitted doesn't qualifies for identification which may be due to poor resolution or if the image is not of flower. Then a message is displayed alerting that identification is not possible on that particular image. If the image is of good resolution and of a flower then the identification process is continued and the image is identified as to belonging to which species, order, family, and genus.
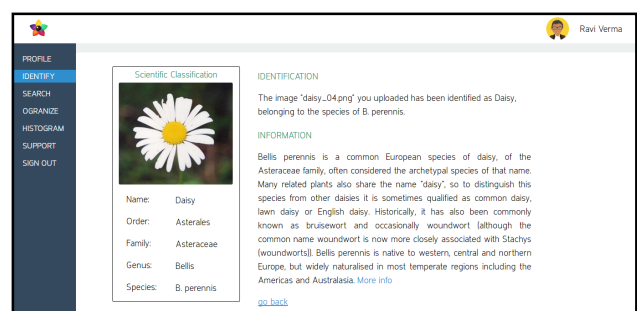

Img 1. Identification result for the flower daisy.

The identification result page shows the result of the identification of the flower daisy. The flower is classified belonging to one species. On the left side of the page, the general details of its scientific classification is displayed which contains the name of the flower species, its order, family, genus, and the species.
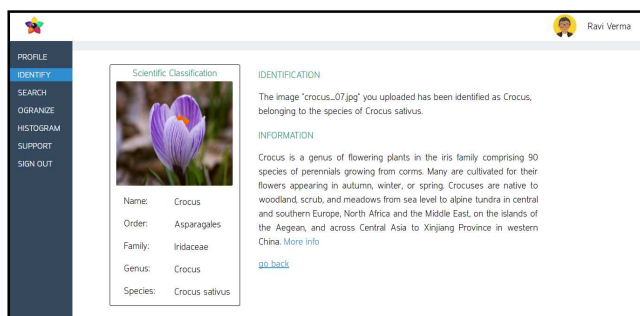
Img 2. Identification result for the flower crocus.

This page shows the result for the identification of the flower crocus. The results are shown similar to the above image but based on the identification of the flower species given by the user.
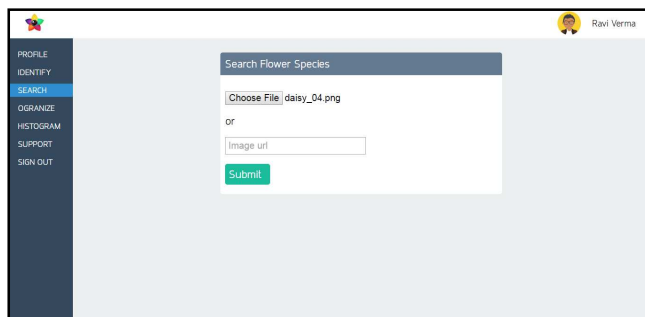

Img 3. Search flower species page.

The search flower species page is used to search for the images similar to a flower image. This is done by passing an image as an input which can be done in two ways. By passing the image from the device or using an image url.
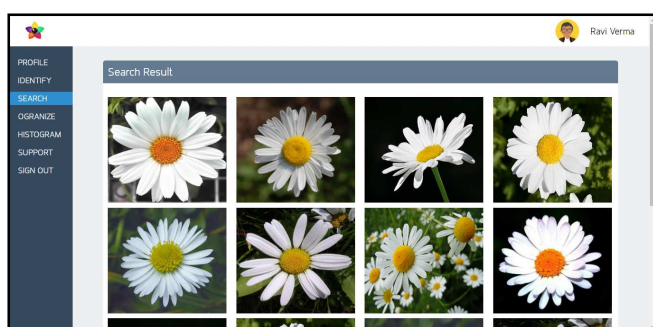

Img 4. Search flower species result page.

The search result page shows the result from the search done on the flower image. The search first classifies the flower i.e. it identifies the flower species, and then shows some similar species images to the user. Here the search

was done on the flower daisy, as a result a list of daisy flowers image are shown on this page.
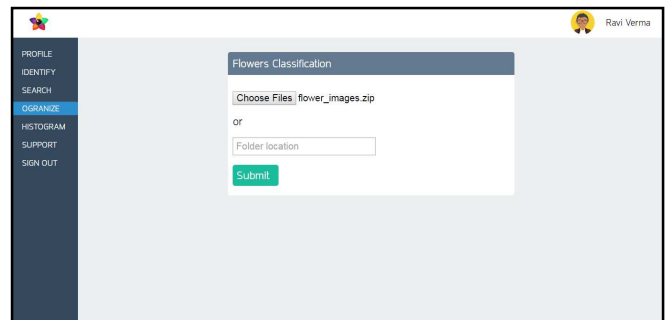
Img 5. Organize flower species page.

The organize page lets the user to organize a folder containing a set of flower images based on the species of the flower. This is done by passing a compressed folder containing flower species images. It can be done in two ways. One is by browsing the folder which is on the device. Another is by passing the url location of the folder.
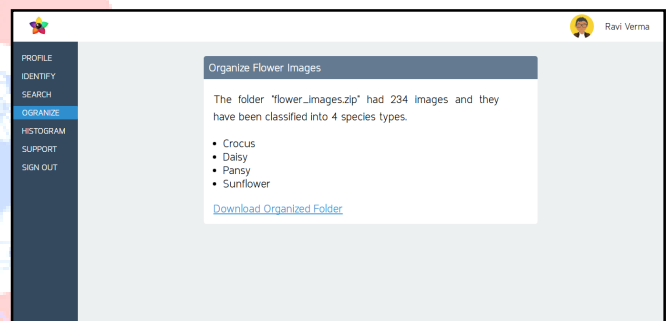
Img 6. Organized result.

The organize result page shows the result for the organization of the folder passed. Here the folder passed contained 234 different images of flower. They have been each identified and classified and a total of four different species have been found in the folder. The organized folder can be downloaded by the user.

V. Conclusion

Classifying images is without a doubt an important part of living things, without which the world as we know it today might not have been possible. Our ability to classify different things has helped us to analyze the importance of certain things which are very crucial for our survival.

One of these is the classification of flower species, which has very important role to play in shaping our society. They provide the nectar which is what the honey bees require for their survival, and in turn the honey bees help in

pollination which is very much crucial for plants to grow. And the plants in turn help in producing oxygen, and absorbing amount of carbon dioxide.

So for this purpose, we have built a system which can be helpful for the botanist's to classify different flower species easily, and the process becomes very fast, helping them in their process of creating medicines, cosmetics, and wide range of products rapidly using our tool.

The system we have built can also be used by the hobbyists to study about different flower species and have knowledge about them.

VI. References

[1] www.cse.chalmers.se/~feldt/courses/reqeng/examples/srs_example_2010 _group2.pdf

[2]   www.udemy.com/data-science-and-machine-learning-with-python-hands-on/

[3] www.tutorialspoint.com

 [4] Mastering OpenCV with Practical Computer Vision Projects - Daniel Lelis Baggio, Khvedchenia Ievgen, Shervin Emami

[5] Head First Data Analysis – Michael Milton

[6] Beginning Programming with Python for Dummies – John Paul Mueller