Project Synopsis
on

Medicinal Plant Leaves Detection

Submitted as a part of course curriculum for

Bachelor of Technology
in
Computer Science

Submitted by
Mohammad Ausaf (2000290120096)
Harsh Agrahari (2000290120069)
Rahul Yadav (2000290120123)

Under the Supervision of
Dr. Sapna Juneja
Assistant Professor

KIET Group of Institutions, Ghaziabad
Department of Computer Science
Dr. A.P.J. Abdul Kalam Technical University

2022-2023

DECLARATION

We hereby declare that this submission is our work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgement has been made in the text.

Signature of Students
Name: Mohammad Ausaf
Roll No.: 2000290120096
Name: Harsh Agrahari
Roll No.:2000290120069
Name: Rahul Yadav
Roll No.: 2000290120123
Date: 11/11/2022

CERTIFICATE

This is to certify that Project Report entitled "**Medicinal Plant Leaves Detection**" which is submitted by **Mohammad Ausaf**, **Harsh Agrahari, Rahul Yadav** in partial fulfilment of the requirement for the award of degree B. Tech. in Department of Computer Science of Dr A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Date: 11/11/2022

Supervisor Signature
Dr.Sapna Juneja
Assistant Professor

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the synopsis of the B. Tech Mini Project undertaken during B.Tech. Third Year. We owe a special debt of gratitude to Ms. Neha Shukla, Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for her constant support and guidance throughout the course of our work. Her sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only her cognizant efforts that our endeavors have seen the light of the day.

We also take the opportunity to acknowledge the contribution of Dr. Ajay Kumar Shrivastava, Head of the Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.
Last but not the least, we acknowledge our friends for their contribution to the completion of the project.

Signature:
Date : 11/11/2022
Name : Mohammad Ausaf
Roll No: 2000290120096
Name: Harsh Agrahari
Roll No.:2000290120069
Name: Rahul Yadav
Roll No.:2000120123

ABSTRACT

Medicinal plants are the source of raw herbal medicines. They are less costly and appear to produce less undesirable side effects than modern medicines. There are approximately 7,200 medicinal plants known in India, of which hundreds of plants have medicinal properties in their leaves.

Identifying such valuable plants often requires an expert or a manual. Therefore, there is a need to automate processes to retrieve information more quickly. Computer vision-based image processing is one of emerging techniques to computerize such tasks by emulating the human visual system.

Although the classification process of plant materials refers to many parameters such as macroscopic and microscopic observation, chemical analysis, and DNA fingerprinting, classification based on macroscopic observation of leaf images based on leaf shape, size, and morphology is the first option for Classification of plant leaves. It is cheap and convenient to select a leaf sample and obtain a leaf sample and its images of her.

The aim of the project is to take raw images of the leaves of plants (provided by the user), upon which it identifies whether the plant is one of the 40 medicinal leaves in the chosen dataset of medicinal herbs, if yes, then provide the user with any relevant information relating to the use case and procedure to use them (if any).

List of Figures

List of Abbreviations

1) ML : Machine Learning
2) CNN : Convolutional Neural Network,
3) LSTM : Long Short-Term Memory
4) BLST : Bidirectional long short-term memory
5) CLSTM : Convolutional long short-term memory
6) K-NN : K Nearest Neighbour
7) SVM : Support Vector Machine
8) DL : Deep Learning
9) RS : Rough Sets,
10) RVM : Relevance vector Machines,
11) RF : Random Forest
12) LR : Logistic Regression
13) OSN : Online Social Network

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Medicinal plants are the source of raw herbal medicines. They are less costly and appear to produce less undesirable side effects than modern medicines. There are approximately 7,200 medicinal plants known in India, of which hundreds of plants have medicinal properties in their leaves.

Human body is complex and fragile, while allopathic medicines contain chemicals, which are inorganic and synthetic, they are not very well suited for continuous consumption by human body. On the contrary, herbal medicines, are admittedly, relatively slow and not an alternative for allopathic medicines, it does help to control and cure diseases with extremely low side effects and cost.

PROBLEM STATEMENT

The main issue with the identification of the medicinal plants are the lack of knowledge and resources, to identify these leaves (by extension, the plants) one either needs an expert or a manual/guide.
Making matters worse, the very availability of the resources is scarce, hence there is a need to automate the identification/detection procedure.

OBJECTIVE

The Objective of the project is to develop a Computer Vision Based Machine Learning Model capable of detecting and classifying any given raw input leaves image provided by the user. Whether the given image is one of the forty species of the medicinal herbs from chosen dataset, if yes, provide the user with all the relevant information about the species, including but not limited to its use case and application/ usage procedure.

SCOPE

The project can help farmers and common people to increase the production of Ayurveda provisions. The automatic classification system will help to identify the medicinal plants without getting any human support in various enterprise sectors such as botanists, taxonomists, Ayurveda manufacturing companies, and Ayurveda practitioners.

The project can be extended to identify and classify more medicinal plants such as flowers, fruit, and seeds in an accurate manner further. The project has the potential to not just identify the plants, but to develop a repository of information for any or every medicinal herb with their application and usage instructions.

LITERATURE REVIEW

1) Real-Time Identification of Medicinal Plants using Machine LearningTechniques

By: Sivaranjani.C, Lekshmi Kalinathan and Amutha.R

- The lighting condition of the environment is uncontrolled, so the segmentation of a leaf from the background is considered as a complex task. Here we propose a system which can identify the plant species based on the input leaf sample. An improved vegetation index, ExG-ExR is used to obtain more vegetative information from the images. The reason here is, it fixes a built-in zero threshold and hence there is no need to use Otsu, or any threshold value selected by the user.Despite the existence of more vegetative information in ExG with Otsu method, our ExG-ExR index works well irrespective of the lighting background. Therefore,the ExG-ExR index identifies a binary plant region of interest. The original color pixel of the binary image serves as the mask which isolates leaves as sub-images. The plant species are classified by the color and texture features on each extracted leaf using Logistic Regression classifier with the accuracy of 93.3%.
- Identification of plants based on the flowers and fruits needs morphological features. Those features are the number of stamens in flowers and number of ovaries in fruits. Distinguishing these plants by using the keys is an extremely tedious process and are done only by the botanists. In addition to this time intensivetask, the few disadvantages in distinguishing plants utilizing these features are the inaccessibility of the needed morphological data and only the experts can understand those botanical terms. The leaves play a vital role in the identification of plant species. Also, the leaves can be easily collected everywhere in all seasons, whereas flowers can only be collected during blooming season.
- In this work, they addressed the problem of identifying the medicinal plant species by the analysis of leaf images obtained directly from their habitat and irrespective of lighting conditions. The fixed zero threshold, ExG-ExR vegetative index is successfully tested for image dataset. The result shows that the algorithm can adequately segment the leaf region. This method worked well in images with reflection. The feature extraction based on the color and texture features is done. The classification of medicinal plant species is done by using Weka and the accuracy of 93.3% is measured. In future we have planned to design and develop a system which automatically identifies plant species through the analysis of not only

2)Machine learning algorithms for teaching AI chat bots

By: Evgeny Tebenkov and Igor Prokhorov

- Machine learning is a method of data analysis, which allows the analytical system to learn while solving many similar problems. Machine learning is based on the idea that analytical systems can learn how to identify patterns and make decisions with minimal human involvement. The history of already completed dialogues between users is used to train chat bots for automated communication with interlocutors.
- Nowadays, chatbots have become widespread in messengers and social networks and due to the proliferation of chatbot platforms have become easy to create and use, while chatbots with artificial intelligence are still rare and not so common. The most common reasons why such chatbots are not popular are the complex and lengthy learning process, as well as imperfect algorithms for processing human requests, which reduces the effectiveness of chatbots and users' trust in them. Today, there are many machine learning algorithms that are used to create chatbotswith artificial intelligence, all of which have their own strengths and weaknesses and are adapted to solve various problems. Within the framework of this article, a comparison of machine learning methods from a theoretical point
- NLP (Natural Language Processing) and machine learning are computer science fields related to AI (Artificial Intelligence). Machine learning can be used in many different fields. NLP takes care of "understanding" the natural language of the person with whom a program (e.g., chatbot) is trying to communicate. This understanding allows a program (e.g., chat-bot) to both interpret input data and produce it in human language. The machine "learns" and uses its algorithms through controlled and uncontrolled learning. Supervised learning means teachinga machine to convert input data into a desired output value. In other words, it assigns the output function to the data so that newer data examples give the same output result for this "learned" interpretation.
- There are various ways that an algorithm can simulate a problem based on its interaction with experience or the environment, or on what we want to call the input data. In machine learning and artificial intelligence tutorials, it is popular to first consider the learning styles that an algorithm can adopt. There are only a few basic learning styles or models that an algorithm can use, and we will look at them here with a few examples of algorithms and the types of problems they suit. This taxonomy, or the way machine learning algorithms are organized, is useful becauseit helps you to define the input roles and, while preparing the model, select the one that best suits your task to get the best result. In the conclusion section of the paper, the author describes the advantages of theproposed method

in comparison to other publications and methods. The author argues that the analysis of the proposed model is well suited for CNN are used todistinguish diseased or non-diseased leaf.

33) Identification of Medicinal Plant Leaves Using Convolutional NeuralNetwork

- **By: Yuanita A. Putri, Esmeralda C. Djamal and Ridwan Ilyas**

- Medicinal plants (herbs) are plants that are known to have certain compounds which are nutritious for health. In Indonesia there are 30,000 types of plants and 7000 of them are classified as medicinal plants (herbs). The human body is complex and organic, while chemical medicines contain chemicals that are inorganic and pure. Therefore, chemical medicine is considered not very suitable for consumption by the human body, which if consumed continuously can even be bad for human health.
- However, some chemical drugs are symptomatic (temporary) so they must be taken for life by patients with certain diseases. Therefore, a system is needed to be able to help the community to recognize medicinal plants better, in this case the medicinal plants are focused on the introduction of medicinal leaves. In this study identification of medicinal plant leaves was carried out using the Convolutional Neural Network method. This research will build a system of identification of medicinal plant leaves by using Convolutional Neural Networks. Using training data that is carried out in a computer set and then implemented in mobile-based software to recognize the types and benefits of medicinal plant leaves identified.
- The human body is complex and organic, while chemical medicines contain chemicals that are inorganic and pure. Therefore, chemical medicine is considered not suitable for consumption by the human body, which if consumed continuously can even be bad for human health. However, some chemical drugs are symptomatic (temporary) so they must be taken for life by patients with certain diseases. The number of medicinal plants is still not balanced with the public's knowledge about the medicinal plants themselves, so many people prefer chemical medicines because they are considered more practical and easier to obtain. Therefore, a system is needed to be able to help the community to recognize medicinal plants better, in this case the medicinal plants are focused on introduction of medicinal leaves.
- This research has successfully implemented the Convolutional Neural Network method to extract features on medicinal plant leaves and identify them into 9 classes

of hypertensive medicinal plant leaves based on the closest value betweenthe training data and test data.

4)Identification of Selected Medicinal Plant LeavesUsing ImageFeatures and ANN

By: R. Janani, A.Gopal

- Medicinal plants constitute a source of raw material for herbal medicines. They have the potential for improving public health at low cost and the occurrence of undesirable side effects seems to be less frequent in comparison withmodern medicines. About 7,200 medicinal plants are knownto occur in India, of which, the leaves of few hundred plants have medicinal properties.

- There is a growing scientific consensus that plant habitats have been altered and species aredisappearing at rates never witnessed before. Correcting this situation requires data about various plant varieties, so that they can be monitored and protected for future generations.

- Identification of such valued plants often needs an expert or a handbook. But there is a dearth of botanists, and the latter process is very cumbersome and time consuming. So, there isa need to automate the process for faster retrieval of information. Computer Vision based image processing is one of the emerging techniques to computerize such tasks by emulating the human visual system.

In this paper, they have implemented a method to identify the correct species of leaf from six different classes. The combination of shape, color, and texture features results in a correct leaf identification rate of 94.4% with a minimum of eight input features. Hence it is seen that the algorithm described in this paper resulted in a reduction of 17.24% in the computation time. The results shown in this method are very promising and thus indicate the aptness of this algorithm for leaf recognition systems. This work can be extended to a larger number of leaf classes with improved efficiency in future.

5) Online Social Network Security: A Comparative Review Using Machine Learning and Deep Learning

By: Chanchal Kumar, Taran Singh Bharati, Shiv Prakash

- This paper essentially deals with the methods to prevent and recover from various classes of cyber security attacks or threats. This paper comprehensively surveys the evolution of online social networks, their associated risks, and solutions.
- The various security models and state-of-the-art algorithms have been discussed along with a comparative meta-analysis using machine learning, deep learning, and statistical testing to recommend a better solution.
- The authors discuss types of security threats on a social network broadly classified into two categories
    - The First is Risk related to the organization: The threat to the software/application that which organization use for personal or official purposes, any severe attack on any such application may put the network of the entire organization at risk.
    - The second is Risk related to the people: Often than not, people intentionally or accidentally reveal or expose their personal information on their social network.
- The author to achieve security of the said OSN(Online Social Networks) recognize the problem, which is the voluminous data, this huge amount of data makes it incredibly hard to make sense of.
- The authors propose to use High-performance computers, essentially supercomputers to process the data in highly specialized parallel computing algorithms.
- The Quality of service (QoS) is measured on the following parameters, Accuracy (A), Precision (P), Recall (R), and F score (F). The authors then go on to compare the results of 9 different algorithms including SVM, KNN, ANN, AIS, AIS(Artificial Immune System), RS(Rough Sets), RVM(Relevance vector Machines), RF (Random Forest), LR(Logistic Regression).
- Metrics used in the comparative analysis were True positives, True Negatives, False Positives, and False Negatives further converted to A, P, R and F scores.
- The author now compares existing algorithms by the means of statistical techniques like t-test, z-test, F-test and Chi-square test.
- After the analysis it is found that the Random Forest with Time complexity of O (M*m*n* log n), where M is the Number of trees, n is the data size, m is the number of features, is the best technique for the classification in the Online Social Network (OSN).

The author gives conclusive remarks about the future direction of this research, which is that Along with the development of new models and techniques, there is a need to compare the existing models to adopt more robust and secure frameworks. Further, the optimization of the implemented framework, the evolutionary, and other approximation algorithms can be used and scaled up.

CHAPTER 3

PROPOSED METHODOLGY

ALGORITHM PROPOSED

The method is divided into two Stages, Stage 1 dealing with edge detection of the medicinal leaf, Stage 2 dealing with classification of the processed image.

Stage 1

The given raw image is processed to determine length, breadth, and dimensions of the leaf[1].

1) The raw input image is taken.

2) The input image is measured for dimensions, that is, for its length and breadth.
3) The image is processed through edge-based detection, giving an edge image outline as an output.
4) The edge outline then undergoes through a Image identification process.

Stage 2

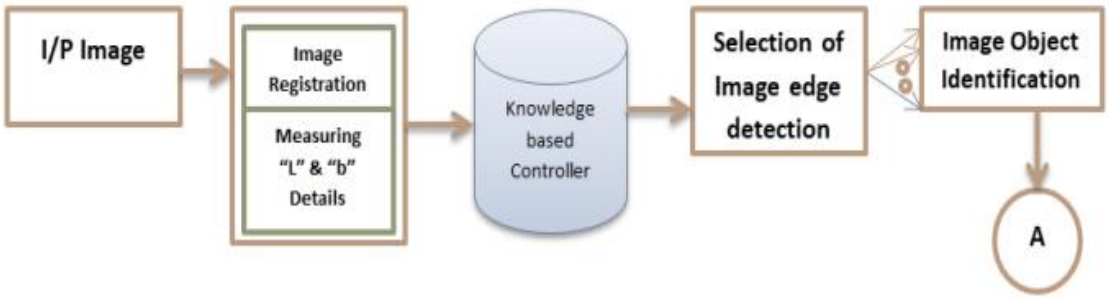
1) The input raw image will be preprocessed to improve the image clarity.
2)Then, the preprocessed image is processed for feature extraction, there are mainly 3 features, namely,

a) Color
b) Shape
c) Measurement

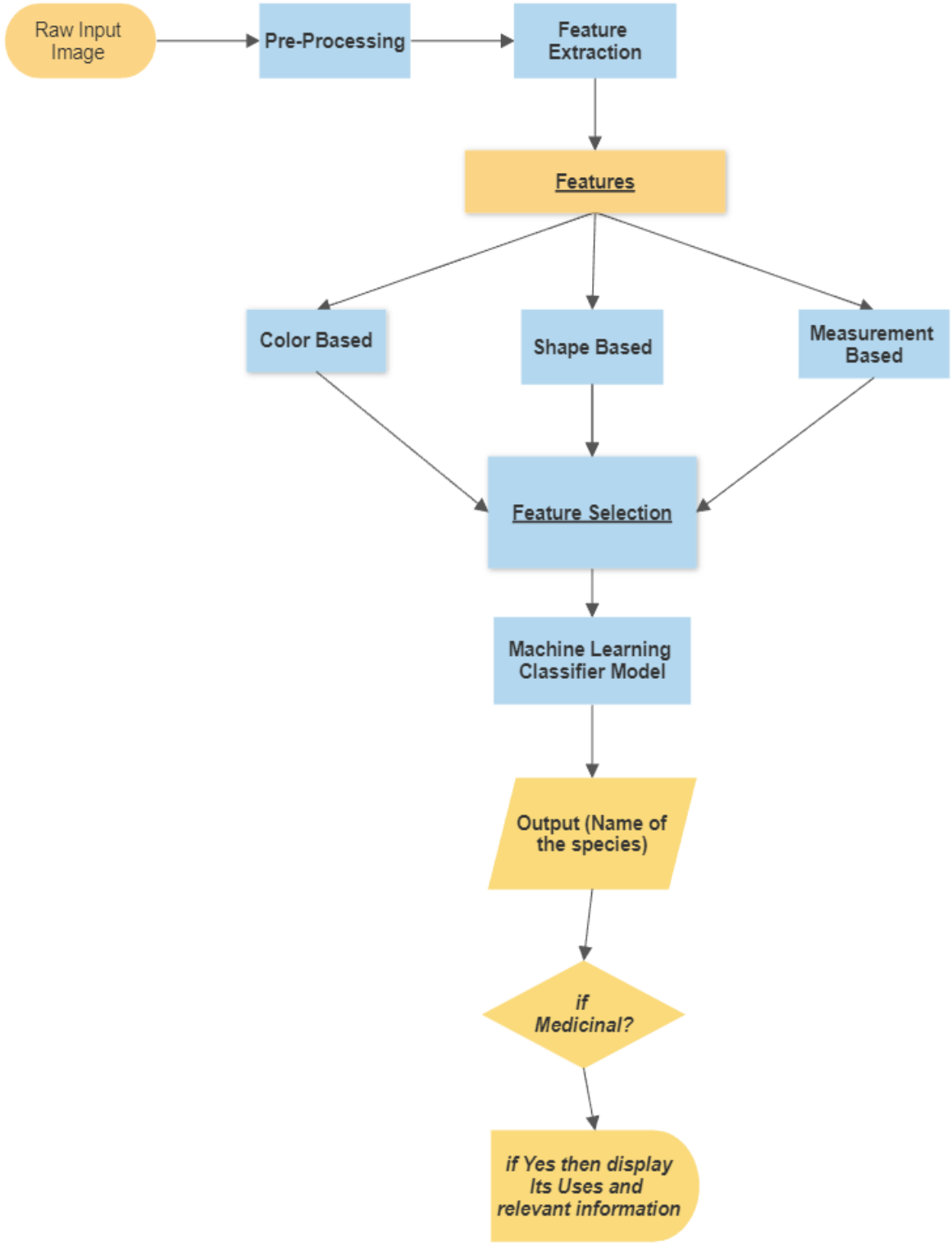
3) The selected feature is then extracted, based upon which we provide the feature as an input to a Machine Learning Classification Model.
4) The Machine Learning model classifies the image in one of the forty species (if applicable).
5) Based on the prediction, the relevant information about the use case and usage instruction is provided to the user.

FLOWCHART

STAGE 1:[1]



STAGE 2:


Raw Input Image
Pre-Processing
Feature Extraction
Features
Color Based
Shape Based
Measurement Based
Feature Selection
Machine Learning Classifier Model
Output (Name of the species)
if Medicinal?
if Yes then display Its Uses and relevant information

Technology Used

1) Jupyter Notebook
2) G1oogle Collab
3) Mendely Dataset
4) Machine Learning Algorithms
5) Libraries like OpenCV, Matplot Lib, SciKit Learn

DIAGRAMS

1)

Figure 1: Samples of the leaves, instance of "Tulsi"

2)

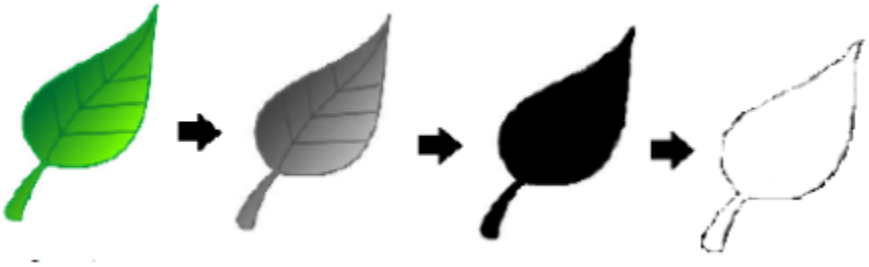


Figure 2: Herbal Plant Leaf Preprocessing Stage

3)

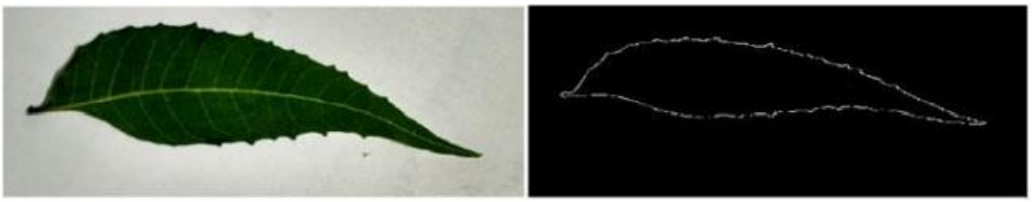


Figure 3: Length wise edge detection

CONCLUSION AND RESULT

1) The Project classifies the medicinal plant leaves, eliminates any human or manual need to identify and thus use Indigenous medicine.
2) The project also additionally, intends to educate and provide usage instructions for different medicinal herbs, thus, acting as a repository for herbal medicine.
3) This automatic classification system will help to identify the medicinal plants without getting any human support in various enterprise sectors such as botanists, taxonomists, Ayurveda manufacturing companies, and Ayurveda practitioners.
4) This will also help the general population in their day-to-day life to use herbal recipes not just limited to “medicine” rather as a supplement to maybe complement any (if any) disease.

REFERENCES

1) Putri, Yuanita A., Esmeralda C. Djamal, and Ridwan Ilyas. "Identification of medicinal plant leaves using convolutional neural network." *Journal of Physics: Conference Series*. Vol. 1845. No. 1. IOP Publishing, 2021.

2) Manoharan, J. Samuel. "Flawless Detection of Herbal Plant Leaf by Machine Learning Classifier Through Two Stage Authentication Procedure." *Journal of Artificial Intelligence and Capsule Networks* 3.2 (2021): 125-139.

3) Janani, R., and A. Gopal. "Identification of selected medicinal plant leaves using image features and ANN." *2013 international conference on advanced electronic systems (ICAES)*. IEEE, 2013.

4) Harakannanavar, Sunil S., Jayashri M. Rudagi, Veena I. Puranikmath, Ayesha Siddiqua, and R. Pramodhini. "Plant Leaf Disease Detection using Computer Vision and Machine Learning Algorithms." *Global Transitions Proceedings* (2022).