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
Ms. 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
Ms. 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)An Introduction to Machine Learning

By
Solveig Badillo, Balazs Bandai, Fabian Birzele, lakov I.Davydov, Lucy Hutchinson.
published in Clinical Pharmacology & Therapeutics Volume 107, Issue 4 2020
Via: Wiliey Online

- There have been huge advancements in the field of pharmacometrics and clinical pharmacology as having advancements in Artificial Intelligence and Machine Learning, this paper is supposed to give people in medical sciences an introduction to Machine Learning.
- This paper deals with the mechanisms and procedures by which Machine learning works, and its algorithms.
- In Data and Features section authors discuss what the input data is in the case of clinical settings for instance disease history, blood test results and gene makeup.
- They now move on to explain learning methods, that as Unsupervised Learning, in particular, they discuss the clustering of data, and how it is relevant in clinical practice.
- The authors discuss unsupervised learning methods, identifying what data can be "useful" or can act as training data to obtain output values.
- In the Discussion section, the authors put forth their final thoughts on how AI and ML in their entirety can be used in advancements in patient treatment and clinical practices as predictive methods.
- Conclusions derived: In the age of big data, there are many new opportunities for ML in clinical pharmacology. For example, data generated from wearable devices pose new challenges on how they can be linked to predictive modelling in the future. In addition, access to real-world data could provide strong evidence for covariates, supplement control datasets, and bolster models that have been trained on small datasets.
- The clinical pharmacology community will continue to base their analyses on pharmacological principles and will gradually build new ML elements to their workflow, strengthening their models further. In addition, the clinical pharmacology community will be able to enhance the range of questions they can address by using ML approaches.

2) Certification systems for machine learning: Lessons from sustainability

By:

Kira J.M. Matus, Michael Veale

Published in Regulation & Governance 2021

Via: Online Wiley Library.

- This paper discusses the need for standardized certification systems in the area of application as well as the development of Artificial Intelligence. Though some regulations exist in the forms of EU's AI acts and some similar but highly insufficient acts and legislations.
- The author discusses how these issues can be analyzed using information about software objects, and information about its deployment context and downstream effect.
- The author discusses how private governance can be the key to solving machine learning's adverse footprints in the environment.
- In the last section, the author discusses how we can learn from pre-existing Sustainability certification systems.
- In conclusion, the author reiterates the importance of measurement and regulations of data processing and the likes of processes in carbon footprints.
- The method or framework proposed by the author is divided into three key parameters, being, standard setting, behaviour modification and information collection, to be elaborate,
    - A set of *technical standards* (which can cover management systems, methods of production, outcomes, or some combination therein). These standards answer the question of *what* is being regulated (what behaviour needs to be modified).
    - A *certification process* (which includes monitoring and enforcement) to ensure that the standards are being met/followed. The certification process answers the question of *how* it will be regulated – specific mechanisms for monitoring and enforcement, including appropriate auditors, who are responsible for information collection.
    - A *labelling program* (to provide information on the credence qualities to consumers). This answers the question of *communication* of participation to the "market".

3) Abusive language detection from social media comments using conventional machine learning and deep learning approaches

By
Muhammad Pervez Akhter, Zheng Jiangbin, Irfan Raza Naqvi

published in Multimedia Systems 2021
Via: Wiliey Online

- This paper essentially deals with the rise of Hate speech and abusive language, with the rise of the Internet. Focusing on the Urdu language which is like Hindi.
- The model is trained and tested on two scripts, the first being Urdu and the second being Roman Urdu or rather Romanized Urdu.
- The author describes which are the people at receiving end of this cyberbullying and trolling, namely people from specific religions or nations.
- The author attempts to detect the abusive language in Urdu roman dialect with more than two thousand comments.
- The author gives a brief introduction to the Urdu alphabet and the basic method of writing.
- The authors discuss unsupervised learning methods, identifying what data can be "useful" or can act as training data to obtain output values.
- The author gives different approaches or Deep Learning model methods to detect comments by Machine Learning and Neural Networks like CNN (Convolutional Neural Network), LSTM (Long Short-Term Memory), BLST (Bidirectional long short-term memory), CLSTM (Convolutional long short-term memory).
- The authors also apply ML models to detect abusive language and compared the performance of these models with Deep Learning.
    - Naïve Bayes: Uses conditional probability and Bayes theorem.
    - K- Nearest Neighbours (K-NN)
    - Support Vector Machine (SVM)
    - Logistic Regression
- On performing comparisons on the above-mentioned models, the authors derive that CNN or Convolutional Neural Network outperforms the other algorithms on both datasets.
- The Deep Learning Model in general performed better than conventional Machine Learning Models.
- The author gives conclusive remarks about the future direction of this research, which is mainly the generation of more and more datasets,

the area of interest for these datasets are various social media websites, for instance, the likes of Facebook and Twitter to detect hate speech. Further, another research direction is to explore hybrid models of DL and ML to detect abusive language

4) Plant leaf disease detection using computer vision and machine learning algorithms

By
Sunil S. Harakannanavar , Jayashri M. Rudagi b , Veena I Puranikmathb , Ayesha Siddiquaa , R Pramodhini a
published in Global Transitions Proceedings
Via: Science Direct

- This paper essentially deals with the detection of plant leaf diseases using computer vision and machine learning algorithm.
- The authors propose a model based on IP (Image Processing) and ML (Machine Learning) approaches for detection of leaf disease.
- The authors use image processing tools like DWT, PCA and GLCM to extract the informative regions and feature of the samples.
- The authors, in the next stage, as a part of machine learning approaches, use the SVM, KNN and CNN to classify the features and the performance of the model is recorded.
- In the preprocessing stage, K-means clustering is applied on leaf images to find out the infected region. The K-mean clustering is used to get the data center of the image and make the clusters of that image and calculates the center distance from the other cluster.
- Color tracing is performed on digital leaf samples to extract their general shape information. After extracting the contour, its characteristics is analyzed and used for pattern classification.
- In the feature extraction stage, Discrete Wavelet Transform or DWT is applied on enhanced tomato leaf samples to extract useful features.
- Then, GLCM derives several properties for the extraction of leaf features. The most used feature-based features are as listed.
- The features obtained using DWT, GLCM, and PCA are combined to form feature vector which are provided as an input sample to the classifiers to recognize classify the images.
- In the classification section, the techniques such as SVM (support vector machines), KNN (K-nearest neighbor) and CNN (Convolutional Neural Network) are used for classifying the samples.

- In the conclusion section of the paper, the author describes the advantages of the proposed method in comparison to other publications and methods. The author argue that the analysis of the proposed model is well suited for CNN are used to distinguish diseased or non-diseased leaf.

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

By Chanchal Kumar, Taran Singh Bharati, Shiv Prakash

Published in Neural Processing Letters 2021

Via: Springer.

- 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].

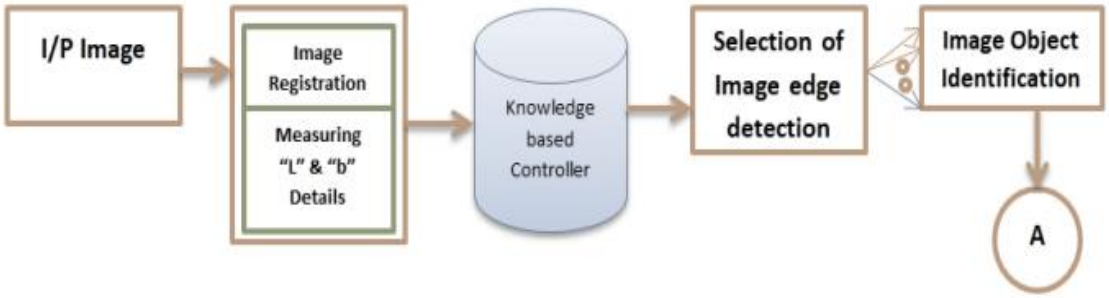
1) The raw input image is taken.
2) The input image is measured for dimensions, that is, for its length and breadth.
3) The image is processed through edge-based detection, giving an edge image outline as an output.
4) The edge outline then undergoes through a Image identification process.

Stage 2

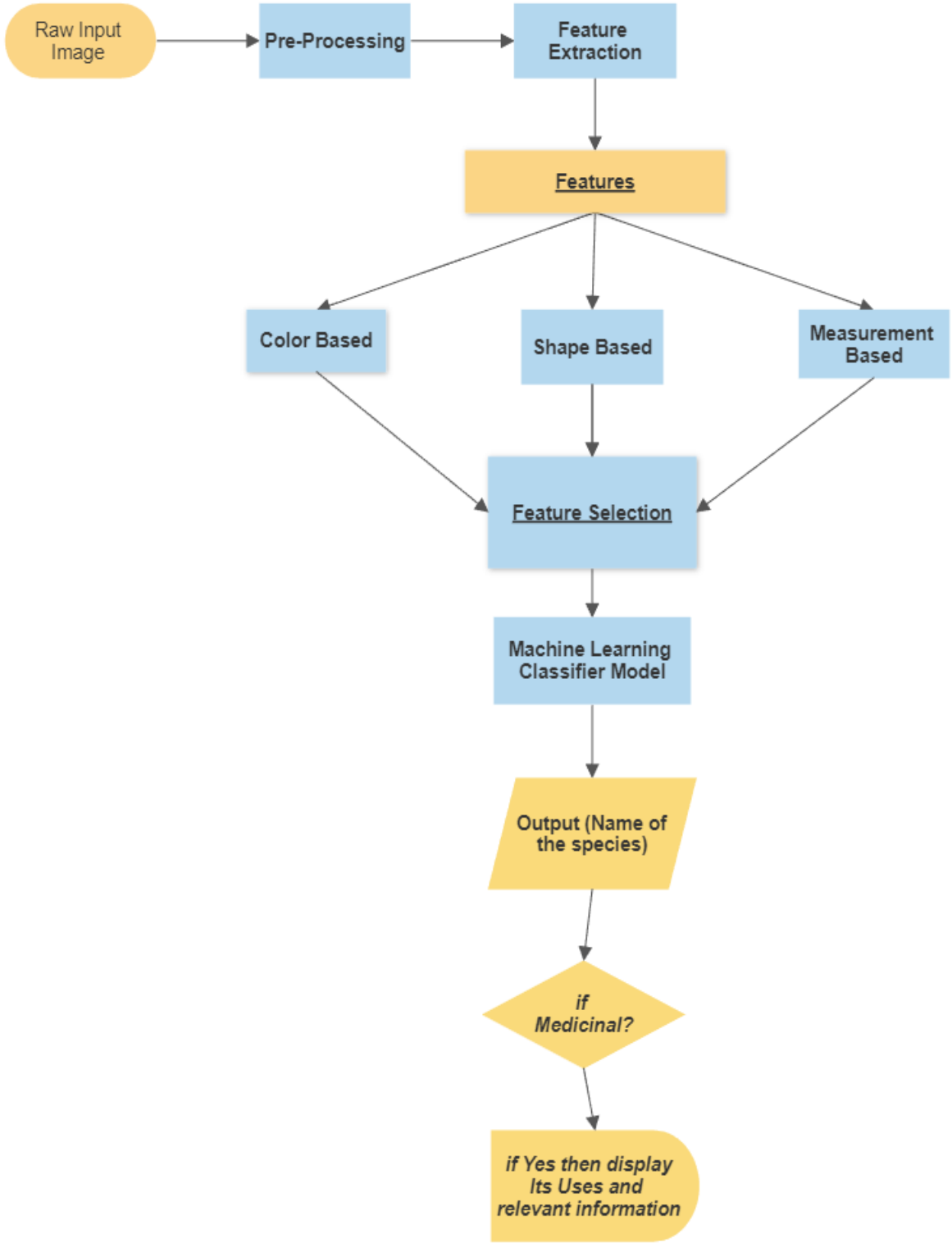
1) The input raw image will be preprocessed to improve the image clarity.
2)Then, the preprocessed image is processed for feature extraction, there are mainly 3 features, namely,
   a) Color
   b) Shape
   c) Measurement
3) The selected feature is then extracted, based upon which we provide the feature 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)

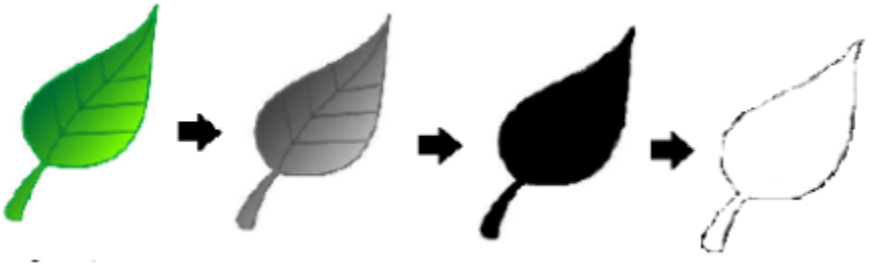


Figure 2: Herbal Plant Leaf Preprocessing Stage

3)

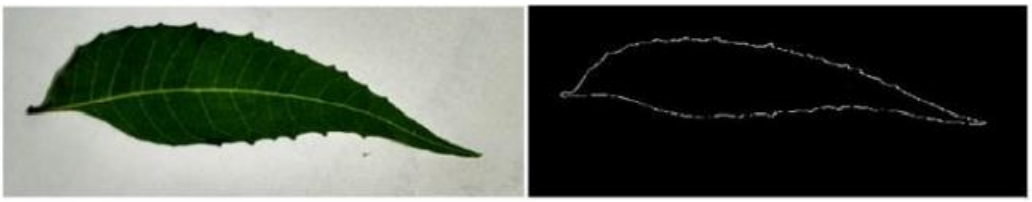


Figure 3: Length wise edge detection

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).