AROGYA - AN INTELLIGENT AYURVEDIC HERB MANAGEMENT PLATFORM

Project ID: 2020-112

Project Proposal Report

N.J. Pathiranage – IT17129404

Nilfa M.S.F - IT17145930

R.A. Mithula Nithmali - IT17089500

K.A.G.Y. Nadee Kumari - IT17014250

BSc (Hons) in Information Technology

Specializing in Software Engineering

Department of Software Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

February 2020

AROGYA – AN INTELLIGENT AYURVEDIC HERB MANAGEMENT PLATFORM

Project ID: 2020-112

Project Proposal Report

(Proposal documentation in partial fulfilment of the requirement for the Degree of Bachelor of Science Special (honors) In Information Technology Specializing in Software Engineering)

BSc (Hons) in Information Technology

Specializing in Software Engineering

Department of Software Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

February 2020

DECLARATION OF THE CANDIDATES AND SUPERVISOR

We declare that this is our own work and this proposal does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Name	Student ID	Signature
N.J. Pathiranage (Leader)	IT17129404	
Nilfa M.S. F	IT17145930	
R.A. Mithula Nithmali	IT17089500	
K.A.G.Y. Nadee Kumari	IT17014250	

The supervisor/s should certify the proposal report with the following declaration.

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

…….. ……

Signature of the supervisor:

(Mrs. Lokesha Weerasinghe) Date

…….. ……

Signature of the co-supervisor:

(Ms. Ishara Weerathunga) Date

ABSTRACT

Ayurvedic means a science of Life well- being with its unique approaches of social and spiritual life is in practice science centuries in the Indian sub-content. There is a considerable species of plants in our environment which will serve us as home remedies for almost all diseases and keeps us healthy when taken for daily consumption. Especially in Sri Lanka we have our own set of rare Ayurvedic herbs which have been utilized by generations as medicinal treatments for a variety of diseases. Detection and classification of medicinal plants, knowing the accurate medical recipes, as well as identifying the category of disease are essential for better treatment. But most of us are unable to identify these plants, and also are with less awareness about valuable Ayurveda recipes due to lack of knowledge. Lack of experts in this field makes proper identification and classification of medicinal plants a tedious task. Hence, a fully automated system for medicinal plant detection and classification, information visualization regarding medicine, track geographical locations of newly identified plants, as well as display the most used herbal sample of each category of disease and visualizing some of herbal remedies related to them are highly desirable. There are existing applications which can identify plants with low prediction accuracies, as well as to give information regarding them. However, these applications are based on foreign plant data sets that do not include valuable herbs and shrubs with medicinal qualities. Hence this research proposes a mobile application which can identify a group of Ayurvedic herbs in Sri Lanka, give information of them through information extraction from sources all over the world with the facility of summary generation, track the locations of newly identified medical plants, display the most used herbal sample of many categories of diseases and visualize some of herbal remedies that are included. Conclusively, the outcome of this study will be used by locals to identify, track and use herbal plants, recipes and remedies precisely.

Keywords: Ayurvedic Leaf Detection and Classification, Computer vision, Geographical Information System, Information extraction

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ANN-Artificial Neural Network

CNN-Convolutional Neural Network

MSF-CNN-Multi-Scale fusion Convolutional Neural Network

DNN-Deep Neural Network

WHO-World Health Organization

HOG- Histograms of Oriented Gradients

LBP- Local Binary Patterns

SVM- Support Vector Machines

ELM- Extreme Learning Machines

PCA- Extreme Learning Machines

KNN- K-Nearest Neighbor

PNN- Probabilistic Neural Network

CBIR- Content-Based Image Retrieval

LSTM – Long Short-Term Memory

NLP – Natural Language Processing

AR – Augmented Reality

GIS – Geographical Information System

NLP: Natural Language Processing

NLTK: Natural Language Toolkit

Mask RCNN – MASK Convolutional Neural Network

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Ayurveda is an ancient medicinal system evolved in India around thousands of years ago, still followed by many people as it is purely natural and has no side effects [2]. It is not only a form of medication – it is a way of life known to generations of Sri Lankans for over 3000 years [2]. It is very relevant from ancient to this most modern time because of its power to cure chronic diseases [2]. According to the World Health Organization (WHO) in 2009, 80% of the people around the world still rely on ayurvedic botanical drugs or medicine. The health-conscious today is searching for effective alternatives to the spiraling costs and side effects that result from the use of modern medicine [2]. Sri Lankans, in the last couple of millennia has made use of the "user-friendly and traditional medicine – Ayurveda" which over 75% of the island's population depend on because of its reliance on natural plants, herbs and oils [2]. They have significant contributions towards human lives and play a predominant role in the well-being of the global population.

According to Ayurveda every plant on earth has some medicinal value, so it is important to protect the plant and identify its medicinal values [30]. Studies have proved that consuming so much of allopathic medicines may lead to side effects as it carries out many chemical reactions within the body [30]. A general fact about Allopathy is that once it is taken, it requires taking another medicine to cure the side effects which has happened due to the previous medicine [30]. In general, process of consuming medicines will not end. Allopathic treatments are meant to treat the Symptoms of a disease whereas Ayurveda treats the root of the disease [30].

One of the major advantages of Ayurveda is that it does not have any side effects as it is purely natural, that is relevant in this most modern time as new diseases evolve due to changed lifestyle and changed diet [4]. Also, majority prefer Ayurveda medicine over Western medicine due to some versatile factors such as Completely natural, Massage treatment, Positive influence on mental health, Fights inflammation, adds to weight loss, Improves heart health, Healthy detox, Individually prescribed methods, etc. [4]. So, it is important for every human being to return to Ayurveda. Almost all general diseases can be cured through Ayurveda using the shrubs and herbs that are

around us. Ayurveda also brings lots of foreign money to the country since many foreign countries are inclining towards it. Nowadays there is a potential class of customers who prefer ayurvedic medicines than other medicines. So, the knowledge about medicinal plants carried by generations must be preserved and protected.

Computer vision, pattern recognition, and image processing technologies provide promising results for identification and classification of medicinal plants. Identifying a medicinal plant with required medicinal values is one of the major challenging tasks [20]. It plays a crucial role in the preparation of ayurvedic medicines. Additionally, proper classification of medicinal plants, important information sources regarding them, knowing about the locations of their growth in Sri Lanka as well as providing medical recipes for diseases of specialized categories are important for agronomist, botanist, ayurvedic medicinal practitioners, forest department officials and those who are involved in the preparation of ayurvedic medicines [20]. But lack of expert taxonomists is a major issue in this area. Even though herbal medicine has no side effects, treatment using a wrongly identified medicinal plant may claim the life of a patient. Hence, a fully automated system to satisfy all the above-mentioned requirements regarding the local medicinal plants is inevitable at this point of time.

1.1 Background & Literature Survey

An overall idea about the importance of Ayurveda in Sri Lanka, current problems faced by people in this domain, apparent tasks that should be there in any automated identification and classification, information resources, tracking locations of existing plants, herbal remedies for categories of diseases with specific characteristics as well as the limitations of previous researches carried out regarding herbal plant management and preservation worldwide have been discussed in the introduction section. This section brings several works and important attention in the focus of this research.

Following are some literature reviews done with some existing mobile applications regarding medicinal herb management, which are somewhat similar to our platform to be implemented.

In 2013, the study of "Pl@ntNet Mobile App" [40] has been proposed as an image sharing and retrieval application for the identification of plants, available on iPhone and iPad devices. Contrary to previous content-based identification applications it can work with several parts of the plant including flowers, leaves, fruits and bark. It also allows integrating user's observations in the database thanks to a collaborative workflow involving the members of a social network specialized on plants. Data collected so far makes it one of the largest mobile plant identification tools.

The iPhone app itself basically contains four functionalities: an image feeds reader to explore the last contributions of the community, a taxonomic browser with full text search options, a user profile and personal contents management screen and the image-based identification tool itself. This one first asks the user to take a picture and then let him choose among 4 icons representing a flower, a leaf, a fruit and a bark. Up to five pictures of the same plant can be acquired in this way and the complete set of pictures can finally be submitted as a query plant to the remote visual search engine. Retrieved species with confidence scores and matched images are finally returned to the device and displayed on the result screen by decreasing confidence. Selecting one of the

retrieved species opens a detailed view screen with all matched pictures (classified by organ galleries) allowing a first stage of refinement in the determination process. A second stage of refinement can be achieved by accessing either e Flore fact sheets (the most complete db on France flora) or Wikipedia mobile pages. If the user believes having found the right species, he can finally contribute by sending his observation with pictures, date, gps and author's name (in Creative Commons). The observation will then be integrated in the collaborative web tool.

In 2013, the research work named "MedLeaf: Mobile Application for Medicinal Plant Identification Based on Leaf Image" [42] has proposed MedLeaf as a new mobile application for medicinal plants identification based on leaf image. The application runs on the Android operating system. MedLeaf has two main functionalities, i.e. medicinal plants identification and document searching of medicinal plant. They have used Local Binary Pattern to extract leaf texture and Probabilistic Neural Network to classify the image. In this research, it has used 30 species of Indonesian medicinal plants and each species consists of 48 digital leaf images. To evaluate user satisfaction of the application they have used questionnaire based on heuristic evaluation. The evaluation result shows that MedLeaf is promising for medicinal plants identification. MedLeaf will help botanical garden or natural reserve park management to identify medicinal plant, discover new plant species, plant taxonomy and so on. Also, it will help individual, groups and communities to find unused and undeveloped their skill to optimize the potential of medicinal plants. As the results, MedLeaf will increase of their resources, capitals, and economic wealth. Keywords— Heuristic evaluation, medicinal plant, identification, Local Binary Patterns, Probabilistic Neural Network.

In 2013, in the research study of "ApLeafis: An Android-Based Plant Leaf Identification System" [43], an Android-based mobile application has been designed to automatically identify plant species by the photographs of tree leaves is described. In this application, one leaf image can be either a digital image from the existing leaf image database or a picture collected by a camera. The picture should be a single leaf placed on a light and un-textured background without other clutter. The identification process consists of totally three steps: leaf image segmentation, feature extraction, and species identification. The demo system is evaluated on the ImageCLEF2012 Plant

Identification database which contains 126 tree species from the French Mediterranean area. The output of the system to users is the top several species which match the query leaf image the best, as well as the textual descriptions and additional images about leaves, flowers, etc., of theirs. The system works well with state-of-the-art identification performance.

In 2012, the study of “Leafsnap: A Computer Vision System for Automatic Plant Species Identification” [41] describes the first mobile app for identifying plant species using automatic visual recognition. The system called Leafsnap identifies tree species from photographs of their leaves. Key to this system are computer vision components for discarding non-leaf images, segmenting the leaf from an untextured background, extracting features representing the curvature of the leaf's contour over multiple scales, and identifying the species from a dataset of the 184 trees in the Northeastern United States. Their system obtains state-of-the-art performance on the real-world images from the new Leafsnap Dataset, which is the largest of its kind. Throughout the paper, it has been documented about many of the practical steps needed to produce a computer vision system such as this, which currently has nearly a million users.

1.1.1 Selecting the Most Accurate and the Highest Performance Segmentation Technique using Experimental Outcomes of Plants in Sri Lanka

Biodiversity is declining steadily throughout the world [52]. The current rate of extinction is largely the result of direct and indirect human activities [53]. Building accurate knowledge of the identity and the geographic distribution of plants is essential for future biodiversity conservation [54]. Therefore, rapid and accurate plant identification is essential for effective study and management of biodiversity.

In a manually identification process, botanist use different plant characteristics as identification keys, that are used sequentially and adaptively to identify plant species. In essence, a user of an identity key's answering a sequence of questions about one or more attributes of an unknown plant constantly focusing at the maximum discriminating characteristics and narrowing down the set of candidate species. This series of answered questions leads ultimately to the desired species. However, the determination of plant species requires a substantial botanical expertise. Sometimes botanists themselves species identification is often a difficult task. The situation is further exacerbated by the increasing shortage of skilled taxonomists [55]. The declining and partly nonexistent taxonomic knowledge within the general public has been termed “taxonomic crisis” [56]. And other hand, it decreases user satisfaction about the application. Because these applications allow users to identify the plant using traits such as flower structure, leaf shape, flower color or fruit color. Sometimes users would not prefer to go several steps to identify the plant.

To overcome these issues, we are going to use object identifications methods. So that user would not be unsatisfied about the product because user just has to take an image of a picture. After capturing the image, using image segmentation we can detect the object with its shape, color and with other features. Because our application gives deepest attention to identify the objects (leaf, fruit or flower) accurately reason is why identifying objects correctly increases the application accuracy, performance and user satisfaction. Otherwise, it will give incorrect results at the end.

Shape is the most popular feature used in plant identification, be it manual or automatic plant identification [57]. So that we are going to use image segmentation techniques rather than using object detection technique. The reason is that Object detection [58] builds a bounding box corresponding to each class in the image. But it tells us nothing about the shape of the object. But image segmentation creates a pixel-wise mask for each object in the image. This technique gives us a far more granular understanding of the object(s) in the image. So, we can get clear idea of the shape of leaf, flower and etc.

Researchers have devised and proposed many techniques for segmentation, but no general technique exists, which may be used for all images.

However, Keri Wood [63] suggested that good image segmentation should meet the following requirements:

1. Every pixel in the image must belong to a region and each region should be homogeneous with respect to a chosen characteristic, which could be syntactic e.g. color, intensity or texture or the characteristic based on semantic interpretation.

2. Every region should be connected and non- overlapping i.e. any two pixels in a particular region should be connected by a line that does not leave the region.

3. It should not be possible to merge two adjacent regions to form a single homogeneous region.

These characteristics also include in our segmentation process in order to get successful results.

“Color images can increase the quality of segmentation, but complexity of the problem is also increased” [63]. Segmentation of a colored image having different kinds of texture regions is a hard problem, particularly if an exact texture field is to be computed and a decision is to be made regarding optimum number of segments in the image. The problem becomes further complicated if the image contains similar and/or “non-stationary texture fields”. Each pixel of colored images is denoted by three component values i.e. Red, Green and Blue and as such these are more complex as far as

segmentation is concerned, than gray scale images which have a single intensity value for a pixel. Colored image segmentation can solve many contemporary problems in medical imaging, mining and mineral imaging, bioinformatics, and material sciences [57]. It leads to lost lots of information of the leaf, flower or fruits. The lots of existing plant identifications applications regenerate the user entered pictures into grey scale images due to complexness of the color images. It results in lost voluminous info of the leaf, flower or fruits.

K. Deepak and A.N. Vinoth proposed to develop an application to identify plant species on android platform [59]. They used edge detection as there segmentation technique. This proposed method contains several major drawbacks. It converts the image into gray scale image. It ends with lost plenty of information approximately the detected object. It will be a major disadvantage when the detected object goes to classification model. And also, in order to use this technique, there should be better contrast between objects. Otherwise, this technique isn't going to work, and it doesn't detect the object accurately. It leads the application failed among the users. This indicates how the importance of detecting object accurately.

Lin-Hai Ma, Zhong-Qiu Zhao and Jing Wang designed plant identification system called APLeafis [60]. They used threshold segmentation for the segmentation process. It is the simplest segmentation method. This application contains several drawbacks. In this application, one leaf image can be either a digital image from the existing leaf image database or a picture collected by a camera. The picture should be a single leaf placed on a light and untextured background without other clutter. Those facts reduce the user satisfaction because before taking a picture, user has to make the background that they mentioned about otherwise this application not going to detect the particular leaf from the image. If the application doesn't select it accurately other all processing things are not going to work. The application is failed. Because now a days, modern people always do and buy things which their lives make easier.

Desta Sandya Prasvita and Yeni Herdiyeni proposed MedLeaf as a new mobile application for medicinal plants identification based on leaf image [61]. Neeraj Kumar, Peter N. Belhumeur, Arijit Biswas and David W. Jacobs proposed a system called

Leafsnap which is a mobile app that helps users identify trees from photographs of their leaves [62]. Above two authors proposed systems they identify the plant only using the leaf. The major drawback of that kind of system is that sometimes some leaves have similar features, but they are different plants. Because Ayurvedic plants are used by people as their medicine. Thus, it affects to human health also. We have to identify those plants separately. So that our proposed system detects not only leaf but also flower and fruit also. It leads to get best accurate result.

As mentioned above several problems occur when detecting the objects (leaf, flower or fruit). Therefore, our proposed method is going to use the most suitable image segmentation techniques to detect the objects even if the image background is complex. In computer vision, we cannot directly apply an algorithm by believing that matches well with our application. We have to do several experiments and get outputs of those and then we should perform some evaluations on final outcomes to select best approach.

1.1.2 Identification and Classification of Ayurvedic Plants in Sri Lanka Using A Mobile Application in An Offline Environment Approach

According to many known sources, researchers have tried many methodologies to extract the features and identify the plant species automatically. According to previous work done, there are existing applications which can identify plants with low prediction accuracies. However, these applications are based on foreign plant data sets that do not include valuable herbs and shrubs with medicinal qualities. Additionally, most of these methods make use of the combination of many parameters like color, shape, texture features etc.

In 2019, the research paper "Oak Leaf Classification: An Analysis of Features and Classifiers" [14] has presented a new oak leaf dataset and preliminary results for classification of 8 types of oak trees. The novelties include comparative analysis of a small set of hand-crafted geometric features and popularly used high-dimensional appearance features, such as Local Binary Patterns (LBP) and Histograms of Oriented Gradients (HOG). They have further compared commonly used Support Vector Machines (SVM) classifier with a recently popular, fast and robust learner called Extreme Learning Machines (ELM). Results indicate that a small set of geometric features reach an accuracy of 75%, while high dimensional appearance features can boost the performance up to92%.

In 2019, in the study of "Combination of Deep Features and KNN Algorithm for Classification of Leaf-Based Plant Species" [15], they have proposed an approach based on the combination of deep architectures. Deep features were extracted from the plant leaves using the fc6 layer of the previously trained AlexNet and VGG16 models. Then, the reduction of the number of deep features by using the Principal Component Analysis (PCA) method was done quickly and the best distinguishing features were obtained. Finally, the classification performances were calculated using the K-Nearest Neighbor (KNN) method. Flavia and Swedish plant leaf data sets were used to test the proposed system. According to the experimental results, the accuracy scores for Flavia and Swedish data sets was obtained as 99.42% and 99.64%, respectively.

In 2019, in the study of "AyurLeaf: A Deep Learning Approach for Classification of Medicinal Plants"[20], they have proposed AyurLeaf, a Deep Learning based Convolutional Neural Network (CNN) model, to classify medicinal plants using leaf features such as shape, size, color, texture etc. This research work also has proposed a standard dataset for medicinal plants, commonly seen in various regions of Kerala, the state on southwestern coast of India. The proposed dataset contains leaf samples from 40 medicinal plants. A deep neural network inspired from Alexnet is utilized for the efficient feature extraction from the dataset. Finally, the classification is performed using Softmax and SVM classifiers. Their model, upon 5-cross validation, achieved a classification accuracy of 96.76% on AyurLeaf dataset.

In 2019, in the study of "Fine-Grained Plant Identification using wide and deep learning model" [21], they have proposed a model to address the fine-grained plant image classification task by using the wide and deep learning framework which combines a linear model and a deep learning model. Proposed method sums the result of the wide and deep learning model using a logistic function so that discrete features can be considered simultaneously with continuous image content. Their works have used metadata such as the date of flowering and locational information for the wide model. Their experiment shows that the proposed method gives better performance than a baseline method.

In 2018, the study of "A Multiscale Fusion Convolutional Neural Network for Plant Leaf Recognition" [16], a multiscale fusion convolutional neural network (MSF-CNN) is proposed for plant leaf recognition at multiple scales. First, an input image is down-sampled into multiples low resolution im- ages with a list of bilinear interpolation operations. Then, these input images with different scales are step-by-step fed into the MSF-CNN architecture to learn discriminative features at different depths. At this stage, the feature fusion between two different scales is realized by a concatenation operation, which concatenates feature maps learned on different scale images from a channel view. Along with the depth of the MSF-CNN, multiscale images are progressively handled, and the corresponding features are fused. Third, the last layer of the MSF-CNN aggregates all discriminative information to obtain the final feature for predicting the plant species of the input image. Experiments show the proposed

MSF-CNN method is superior to multiple state-of-the art plant leaf recognition methods on the MalayaKew Leaf dataset and the Leaf Snap Plant Leaf dataset.

In 2018, the study of "Improving Plant Recognition using Hybrid features from Connectionist and Knowledge-Based Approaches" [22] has proposed architecture that combined knowledge-based approach to improve the accuracy of plant recognition. Towards this, hybrid features are constructed by merging three types of knowledge-based features; morphological feature, texture feature and color feature with convolutional neural network extracted features. Their architecture consists of three main stages which are data pre-processing, feature extraction and classification. Before features are extracted, images will be resized and augmented in the pre-processing stage. To classify the species of leaf, they consider decision tree and artificial neural network as a classifier. They have experimented on two datasets: Flavia and Swedish dataset. The experimental result shows that the proposed architecture can predict unseen images correctly more than existing models.

In 2017, the study of "Plant Classification using Convolutional Neural Networks" [23] has proposed a Convolutional Neural Network (CNN) architecture to classify the type of plants from the image sequences collected from smart agro-stations. First challenges introduced by illumination changes and deblurring are eliminated with some preprocessing steps. Following the preprocessing step, Convolutional Neural Network architecture is employed to extract the features of images. The construction of the CNN architecture and the depth of CNN are crucial points that should be emphasized since they affect the recognition capability of the architecture of neural networks. In order to evaluate the performance of the approach proposed in this paper, the results obtained through CNN model are compared with those obtained by employing SVM classifier with different kernels, as well as feature descriptors such as LBP and GIST. The performance of the approach is tested on dataset collected through a government supported project, TARBIL, for which over 1200 agro-stations are placed throughout Turkey. The experimental results on TARBIL dataset confirm that the proposed method is quite effective.

In 2016, the study of "Ayurvedic Plant Species Recognition using Statistical Parameters on Leaf Images" [30], has proposed a simple and efficient methodology for Ayurvedic plant classification using digital image processing and machine vision technology. The three major phases in proposed methodology are pre-processing, feature extraction and classification. Pre-processing is done in order to highlight the relevant features to be used in the proposed methodology as well as to reduce unwanted noise from the input image, which reduces the chance of getting optimal feature values. In feature extraction phase, different morphologic features such as mean, standard deviation, convex hull ratio, isoperimetric quotient, eccentricity and entropy are extracted from the pre-processed leaf image. In the third phase, a new approach to classify ayurvedic plant species is adopted to recognize plant species by calculating the leaf factor of the input leaf using the extracted feature values and it is compared with the trained values that are stored in the database. An accuracy of 93.75% is obtained for the proposed methodology.

In 2015, the study of "Recognition of Whole and Deformed Plant Leaves using Statistical Shape Features and Neuro-Fuzzy Classifier" [24] has proposed a methodology for recognition of plant species by using a set of statistical features obtained from digital leaf images. As the features are sensitive to geometric transformations of the leaf image, a preprocessing step is initially performed to make the features invariant to transformations like translation, rotation and scaling. Images are classified to 32 pre-defined classes using a Neuro fuzzy classifier. Comparisons are also done with Neural Network and k-Nearest Neighbor classifiers. Recognizing the fact that leaves are fragile and prone to deformations due to various environmental and biological factors, the basic technique is subsequently extended to address recognition of leaves with small deformations. Experimentations using 640 leaf images varying in shape, size, orientations and deformations demonstrate that the technique produces acceptable recognition rates.

In 2015, the study of "A Convolutional Neural Network for Leaves Recognition Using Data Augmentation" [25], a seven-layer ConvNet using data augmentation is proposed for leaves recognition. First, they implement multiform transformations (e.g., rotation and translation etc.) to enlarge the dataset without changing their labels. This novel

technique recently makes tremendous contribution to the performance of ConvNets as it is able to reduce the over-fitting degree and enhance the generalization ability of the ConvNet. Moreover, in order to get the shapes of leaves, they sharpen all the images with a random parameter. This method is similar to the edge detection, which has been proved useful in the image classification. Then have trained a deep convolutional neural network to classify the augmented leaves data with three groups of test set and finally find that the method is quite feasible and effective. The accuracy achieved by their algorithm outperforms other methods for supervised learning on the popular leaf dataset Flavia.

In 2015, the study of "DEEP-PLANT: PLANT IDENTIFICATION WITH CONVOLUTIONAL NEURAL NETWORKS" [26] studies convolutional neural networks (CNN) to learn unsupervised feature representations for 44 different plant species, collected at the Royal Botanic Gardens, Kew, England. To gain intuition on the chosen features from the CNN model (opposed to a 'black box' solution), a visualization technique based on the deconvolutional networks (DN) is utilized. It is found that venations of different order have been chosen to uniquely represent each of the plant species. Experimental results using these CNN features with different classifiers show consistency and superiority compared to the state-of-the art solutions which rely on hand-crafted features.

In 2014, the study of "Ayurvedic leaf recognition for Plant Classification" [31], the performance of different features extraction methods are compared, different combinations of features and a number of classifiers applied for leaf identification process are also discussed.

In 2012, the study of "An Efficient Leaf Recognition Algorithm for Plant Classification Using Support Vector Machine" [27] uses an efficient machine learning approach for the classification purpose. This proposed approach consists of three phases such as preprocessing, feature extraction and classification. The preprocessing phase involves a typical image processing steps such as transforming to gray scale and boundary enhancement. The feature extraction phase derives the common DMF from five fundamental features. The main contribution of this approach is the Support

Vector Machine (SVM) classification for efficient leaf recognition. 12 leaf features which are extracted and orthogonalized into 5 principal variables are given as input vector to the SVM. Classifier tested with flavia dataset and a real dataset and compared with k-NN approach, the proposed approach produces very high accuracy and takes very less execution time.

In 2012, the study of "Classification of Selected Medicinal Plants Leaf Using Image Processing" [32] has aimed at implementing a system using image processing with images of the plant leaves as a basis of classification. The software returns the closest match to the query. The proposed algorithm is implemented, and the efficiency of the system is found by testing it on 10 different plant species. The software is trained with 100 (10 number of each plant species) leaves and tested with different plant species) leaves. The efficiency of the implementation of the proposed algorithms is found to be 92%.

1.1.3 Extractive Information Summarization and Location Mapping with Smart Visualization on Ayurvedic Plants in Sri Lanka Using a Mobile Application in an Offline and Online Environment Approach

Most of the present Ayurvedic based applications [61] follows only the medicinal plants identification process. There are no smart visualization techniques used like augmented reality to display information to the users. And also, the proposed system Arogya is capable of generating dynamic summaries regarding the identified medicinal plants. So, the system has to work with the online database, whenever the web pages get updated, the generated summary also should be adjusted in order to the current version of the changed website page. Hence it is kind of a real time information extraction, users also will get updated automatically. Currently the existing systems are [62] just managing with the offline database of their own system, so the information will not be changed according to the timeline. So, the users will not get updated with the real time information of that particular medicinal plants.

In consequence, Arogya acts as a single point of resource access for multiple features to overcome all these difficulties in the current products in the market [63]. Arogya's main goal is to be the feasible single unit of medicinal plants information hub which will satisfy all the user requirements regarding ayurvedic plants.

Some of the existing system's functionalities in comparison to Arogya can be shown in the below table.

References	Smart Visualization	Real time Summary generation on information	Using GIS for Location tracking
H. P. Edmundson, "New methods in automatic extracting"[64]	✗	✗	✗
J. Kupiec, J. Pedersen, F. Chen, "A trainable document summarizer" [65]	✗	✗	✗
Neeraj Kumar, Peter N Belhumeur, Arijit Biswas, David W Jacobs, W John Kress, Ida C Lopez, João VB Soares, "Leafsnap: A computer vision system for automatic plant species identification" [50] research was done on Identification of plant leaf using computer vision techniques	✗	✗	✗
D. R. Radev, H. Jing, M. Stys, D. Tam, "Centroid-based summarization of multiple documents" [66], Information Processing and Management	✗	✗	✗
Arogya our proposed System	✓	✓	✓

Table 1.1.3:1 Comparing with existing system

1.1.4 The Text classification of identifying a category type of diseases and visualizing ingredients in most probably used Ayurvedic prescription/recipes

Arogya Mobile app is going to overcome some of the problems mentioned in above. apart from there is a special feature of texting a description including characteristics of herbal remedies/ plants. Mainly, that option will be provided for identifying each type of herbal remedies / plants which represent in what category of diseases who may be seen or ever seen either know the characteristics well.

In this App going to implement an automation text classification on Machine Learning algorithms related applications.

So that, this proposed system is more important to identify the herbal remedies when people failed in its identification and not understanding in better ayurvedic samples/prescriptions medicine even though in Offline status.

In ancient Indian medical system, also known as Ayurveda, is based on ancient writings that rely on a natural and holistic approach to physical and mental health. This is also world's oldest medical systems and remains one of India's traditional health care system. Every person engaged in various type of activities/ treatments internet either by using social media or online identification activities. Many numbers of activities go through on internet as well as treatments in Ayurvedic and western approach medicine for health-related issues [44].

Nowadays, many online communities are focused on health-related issues in western medicine approach and Ayurveda treatments. In order to that it clarify different types of diseases such as diabetes, Arthritis, Cancers etc [47]. However, they introduced dynamic exercise plans, nutritious to treatment and diagnosis for curving everything else related diseases [45]. Therefore, today Created many no of application for each of different tasks but our proposed system will give special options as clarifying a texts identifying its category type. According to this Text classification and picture extraction function is somewhat different from the extract process of text classification apply in this application. There interface displays with dropdown icons for selecting needed plants. Hence, there Analyzing a categorized process is on feature extraction stage using Naive Bayes classifiers [46].

1.2 Research gap

In computer vision, despite many efforts [8–13], plant identification, classification, tracking locations of valuable plants, defining information sources with summary reports, as well as details of medical recipes for specific categories of diseases are still considered as challenging and unsolved problems. It is very challenging since the unavailability of a single platform which can manage all of the above requirements regarding Ayurveda in Sri Lanka.

In Sri Lanka we have our own set of rare Ayurvedic herbs. But most of us are unable to identify these plants due to lack of knowledge. Dissemination of knowledge regarding herbal plants is restricted mainly to very limited group of people and is passed down from generation to generation who practice traditional medicine.

So, recognizing these endemic herbal plants is a challenging problem in the fields of Ayurvedic medicine, computer vision, and machine learning, because although leaf classification apps have been implemented for leaves in other countries, such an automated app with a reliable mechanism has not still been implemented for the purpose of locals. There are existing applications which can identify plants with low prediction accuracies. However, these applications are based on foreign plant data sets that do not include valuable herbs and shrubs with medicinal qualities. Therefore, it is the main research gap which can be identified through the existing researches and implemented automated.

If the main research gap is summarized, it is the “Lack of a single platform for the management and preservation of Ayurvedic plants in Sri Lanka using computer vision, and unavailability of a full-automated mobile application for the particular purpose.” This is mostly important in the system of medicine called Ayurveda in Sri Lanka, because identification and classification of medicinal plants is considered an important activity in the preparation of herbal medicines. Because treatment using a wrongly identified medicinal plant may claim the life of a patient. Additionally, it is important for Ayurveda practitioners and also traditional botanists to know how to identify and

classify the medicinal plants through computers [5]. They should apply their knowledge of Ayurveda with technical approach too.

Following table shows the comparison of several existing mobile applications with plant classification functionality with our system, and the gaps are well defined through this structure.

	Agrobase	**Plantex**	**PlantSnap**	**Pl@ntNet**	**MedLeaf**	**Arogya**
Search Plants	✓	✓	✓	✓	✓	✓
Identify and classify plants	✓	✓	✓	✓	✓	✓
Filters whether the searched plant is a medical plant or not	✓	✓	×	×	✓	✓
Identify and classify plants offline	×	×	×	×	×	✓
Mapping the locations of newly identified plants	×	×	×	×	×	✓
Show directions from users' current location	×	×	×	×	×	✓
Ability to share plant location	×	×	×	×	×	✓
Gives a Summarized information about the classified plant	✓	✓	✓	×	✓	✓

Gives the user an AR experience	×	×	×	×	×	✓
Gives remedies for specific categories of diseases and propose medical recipes for them	×	×	×	×	×	✓
Gives user a novel social media experience limited to trees (upload/view/share posts only related to plants) and makes users aware of the importance of plants to our world	×	×	×	×	×	✓

Table 1.2.1: Comparison with existing systems and identifying gaps

1.3. Research problem

The problem regarding the identification and familiarization of the plants, getting information about the value of them precisely, tracking locations of their growth and availability, providing remedies and recipes for specific categories of diseases limited to the specified herbal medicine plant leaves, along with the seen necessity for the development of herbal medicine plant identification should be addressed by a study. Especially in Sri Lanka we have our own set of rare Ayurvedic herbs. But most of us are unable to identify these plants due to lack of knowledge. Though many are aware of the existence of different herbal medicine plants and their familiar applications, many are still unable to identify which are these herbal medicine plants from the vast diversity of plants in the environment, so as their noted and approved applications by health professionals.

There are several methods to recognize a plant and use medical recipes assigned with. At present, plants are identified, as well as recipes are predicted manually by taxonomist, which are prone to human errors. But manual process requires prior knowledge and also it is a lengthy process. Leaf Identification by mechanical means often leads to wrong identification. Identification and classification of medicinal plants are essential for better treatment. Lack of experts in this field makes proper identification and classification of medicinal plants a tedious task. This is one of the major problems in this domain.

Additionally, researchers in the field of botany, medicine, chemical structure analysis, agriculture, and ayurvedic medicinal practitioners, forest department officials, those who are involved in the preparation of ayurvedic medicines and others who are concerned with plant studies are faced with the application of considerable effort on identifying plants. It is stated that plant identification demands extensive knowledge and uses complex terminology, even professional botanists need to take much time in the field to master plant identification [7]. Identifying a medicinal plant with required medicinal values is one of the major challenging tasks faced by them. It plays a crucial role in the preparation of ayurvedic medicines. But lack of expert taxonomists is a

major issue in this area. Even though herbal medicine has no side effects, treatment using a wrongly identified medicinal plant may claim the life of a patient.

Below are the summarized research problems those are going to be addressed through our research study.

1.Lack of technological research carried out on the Recognition and Classification of Ayurvedic plants in Sri Lanka using machine learning and computer vision.

2.Absence of a single platform to detect and classify Sri Lankan ayurvedic leaves, give important information about them and to show the locations from where those plants can be found in Sri Lanka.

3.Lack of a system which provides many features as services of giving Traditional treatments.

4.Lack of a system which identifies the category type of disease in an unknown herbal remedy by giving its special characteristics/properties and which displays the most used herbal sample of each category and visualizing some of herbal remedies that are included.

5.Current mapping systems used are inaccurate and lacks use of a common GIS platform.

6.Only manual location records are available.

7.Dissemination (Circulation) of knowledge regarding herbal plants is restricted mainly to very limited group of people and is passed down from generation to generation who only practice traditional medicine.

Plants are an indispensable part of our ecosystem and the dwindling number of plant varieties is a serious concern. Another major problem is that, there is a growing scientific consensus that plant habitats have been altered and species are disappearing at rates never witnessed before. The biodiversity crisis is not just about the perilous state of plant species but also of the specialists who know them This initially requires

data about various plant varieties, so that they could be monitored, protected and can be used for future. Plants form the backbone of Ayurveda and today's modern-day medicine and are a great source of revenue. Due to Deforestation and Pollution, lot of medicinal plant leaves have almost become extinct. So, there is an urgent need for us to identify them and regrow them for the use of future generations. Due to growing illegal trade and malpractices in the crude drug industry on one hand and lack of sufficient experts on the other hand, an automated and reliable identification and classification mechanism in order to handle the bulk of data and to curb the malpractices is needed.

2. OBJECTIVES

2.1 Main Objective

The primary objective of the proposed solution is to develop a Centralized Social Media platform (android mobile application) that gives user the ability to identify and classify a group of selected important ayurvedic plant species in Sri Lanka, based on photographs of the plant's leaf as well as other parts such as digested root and fruit, taken with mobile phone, to share information sources of them through information extraction, get information about remedies and recipes for specific disease categories as well as the tracked locations with the aim of preservation of valuable Ayurvedic plants in Sri Lanka.

Therefore, anyone without prior knowledge will be able to identify and classify the particular ayurvedic plant hopefully. The development strategy and methodology used in this approach will be able to be used and extended to identify any ayurvedic herb in Sri Lanka furthermore.

2.2 Specific Objectives

Following are some specific objectives those are going to be achieved through our new product

1.Recognizing and classifying the leaves of the herbs/plants used for Ayurvedic treatment once the user inputs an image of a leaf captured by mobile camera.

2.Providing a description of the identified herb's medicinal usage in reference to illnesses by going through feed data.

3.Dynamic search for the names of the plant leaves through a word search, or to do either a web search or generate a summary of the plant.

4.Use augmented reality to display onsite information on medicinal plants and provide a summarized detail regarding the values and uses of them through abstractive information extraction and summarization.

5.Identify the category type of disease in an unknown herbal remedy by giving its special Characteristics/properties. After displaying the category, The System will display the most used herbal sample of each category of disease.

6.User is cable of adding the location of plants into the Map, as well as to get use of the locations of the previously tracked plants (availability and growth)

7.After identification process user can choose to publish the post and other users can comment on them.

8.Can be seen rare or unknown medicine plants which are included in as ingredients of medicine recipes.

On a mobile device, the Ayurveda plant detection has to be done with time, battery life critical manner, especially when it has to be done in a forest area. In the proposed system the whole identification process will take place on mobile devices and it doesn't require internet. Therefore, this will be a great solution to identify Ayurveda plants in deep forest areas, where mobile network coverage is not available.

3. METHODOLOGY

3.1 Understanding the key pillars of the research domain

The system mainly relies on the key pillars of Deep Learning, Transfer Learning based on deep CNN and Data Augmentation.

3.1.1 Deep Learning

According to many previous research works done [10-30], identification and classification of plants have been carried out for the last few decades using classical image processing and classification methods. These methods have used shape, texture and color-based features to perform classification. Aspect ratio, eccentricity, kurtosis, skewness, energy, correlation, sum variance, entropy and compactness are some of these features [20]. Excessive computation time required for handcrafted feature extraction is the major problem associated with these classical methods. Nowadays all the classical methods are replaced by machine learning techniques.

Deep learning is a class of machine learning in which a computational model learns to perform classification tasks directly from images. It is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example [37]. Higher accuracy, ability to handle huge volume of image data, inbuilt ability to use GPUs for parallel computation and availability of inbuilt pre-trained Convolutional Neural Networks contribute towards the popularity of deep learning. Since Arogya performs classification on huge volume of image data, deep learning approach will be the best choice.

3.1.2 Convolutional Neural Network (CNN)

A convolutional neural network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data [39]. A CNN uses a system much like a multilayer perceptron that has been designed for reduced processing requirements. The layers of a CNN consist of an input layer, an output layer and a hidden layer that includes multiple convolutional layers, pooling

layers, fully connected layers and normalization layers. The removal of limitations and increase in efficiency for image processing results in a system that is far more effective, simpler to trains limited for image processing and natural language processing [39].

CNN has been recognized as a competent method for image recognition in past few years, on mobile devices in an offline environment. CNNs work somewhat similar to neurons in brain. It also performs image transformation with a certain degree of rotation and distortion [35]. Since this network is avoiding the complex preprocessing of the image, we will able to input the original image of the suspected plant directly.

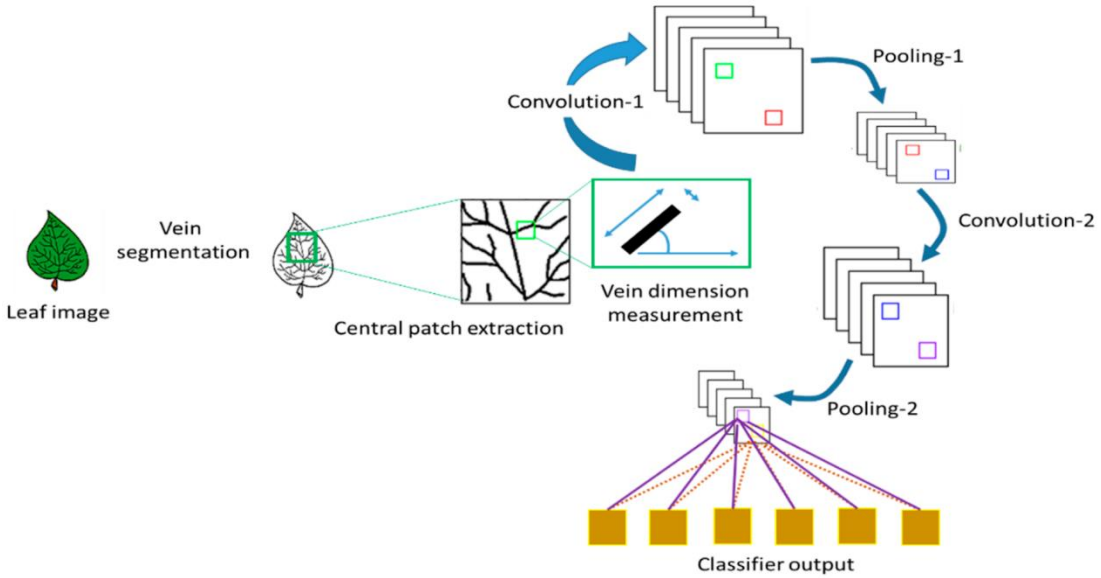


Figure 3.1.2.1: Review on techniques for plant leaves CNN:

Source: https://www.google.com/https://www.mdpi.com

3.1.3 Transfer Learning based on deep CNN

An interesting benefit of deep learning neural networks is that they can be reused on related problems. Transfer learning refers to a technique for predictive modeling on a different but somehow similar problem that can then be reused partly or wholly to accelerate the training and improve the performance of a model on the problem of interest [38]. In deep learning, this means reusing the weights in one or more layers

from a pre-trained network model in a new model and either keeping the weights fixed, fine tuning them, or adapting the weights entirely when training the model [38].

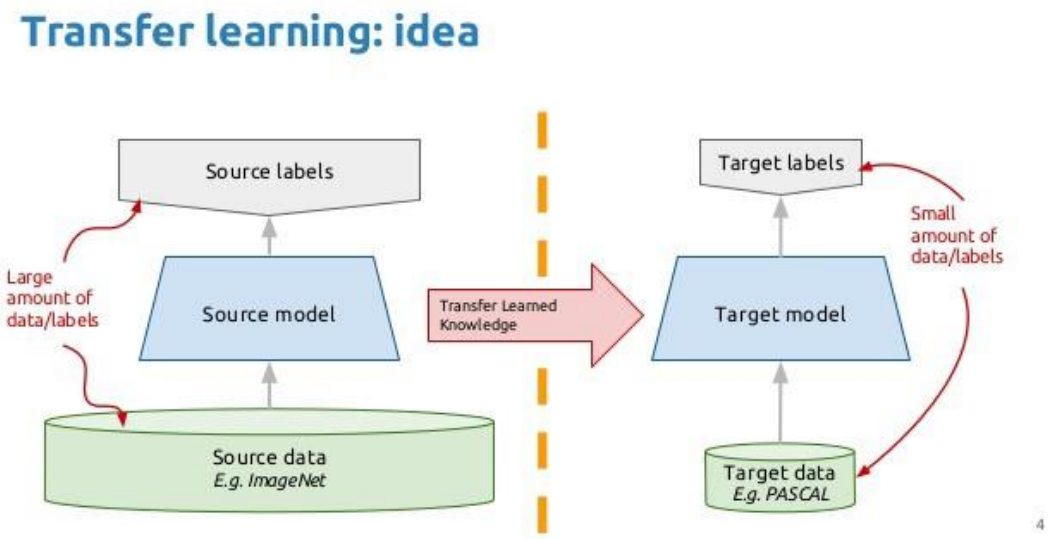


Figure 3.1.3.1: Transfer Learning with Convolutional Neural Networks in PyTorch:

Source: https://www.google.com/towardsdatascience.com/transfer-learning-with-convolutional-neural-networks-in-pytorch

Following is the general outline for transfer learning for object recognition [18]:

1. Load in a pre-trained CNN model trained on a large dataset

2. Freeze parameters (weights) in model's lower convolutional layers

3. Add custom classifier with several layers of trainable parameters to model

4. Train classifier layers on training data available for task

5. Fine-tune hyperparameters and unfreeze more layers as needed

3.1.4 Data Augmentation

As mentioned in many researches in this domain, among many of DNN structures, the Convolutional Neural Networks (CNN) are currently the main tool used for the image analysis and classification purposes [17]. Although great achievements and perspectives, deep neural networks and accompanying learning algorithms have some relevant challenges to tackle [17]. The most frequently mentioned problem in the field of machine learning, is the lack of sufficient amount of the training data or uneven class balance within the datasets [17]. One of the ways of dealing with this problem is so called data augmentation [17]. For training the convolutional neural network that we use for extracting features, there is an idea for training images to be augmented. As a CNN requires a large number of images to train it, various data augmenting methods is planned to be applied to the collected dataset to increase the size of the dataset.

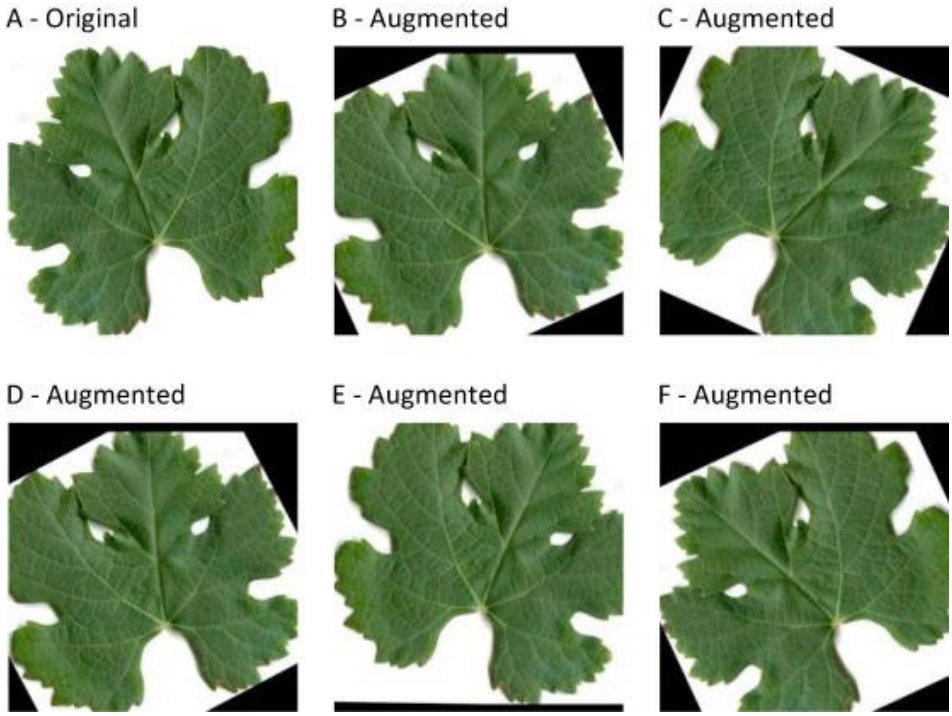


Figure 3.1.4.1: Data Augmentation with leaves:

Source: https://www.google.com/https/www.sciencedirect.com/science/article/pii

3.2 Methodologies of each Approach

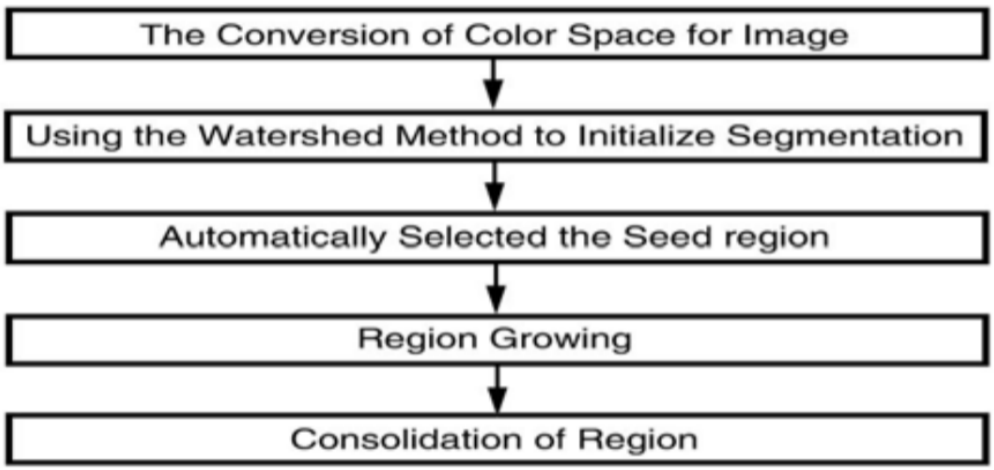
3.2.1 Selecting the Most Accurate and the Highest Performance Segmentation Technique using Experimental Outcomes of Plants in Sri Lanka

3.2.1.1: A color image segmentation method of automatic seed region growing on basis of the region with the combination of the watershed algorithm.

Our expectation is to build a segmentation process that can detect the leaf, fruit or flower even if it is in the complex background. And also, we are not going convert image into a grayscale image reason is that it leads to lost the most important information about the detected objects. This provides a complementary approach to the segmentation of objects. It is especially useful for segmenting objects that are touching one another. But if we use directly watershed algorithm, it views only a grayscale image as a topological surface. Thus, we use our proposed approach 1 mentioned above as our first segmentation process.

Existing applications have a major drawback of segmenting that kind of user captured images. When image with overlapping leaves, there segmentation process didn't work. At the end, it produced the wrong result because at the begging they didn't correctly identify the objects. So, our proposed method overcome this issue as well.

The flow chart for the algorithm as shown



3.2.1.1.1 - An overview of watershed segmentation

Approach 2: Mask RCNN (Mask Convolutional Neural Network)

Mask RCNN presents a conceptually simple, flexible, and general framework for object instance segmentation. This approach 2 efficiently detects objects in an image while simultaneously generating a high-quality segmentation mask for each instance. The method, called Mask R-CNN, extends Faster R-CNN by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition. Mask R-CNN is simple to train and adds only a small overhead to Faster R-CNN, running at 5 fps.

The fact which we proposed this technique is that the network with the Mask R-CNN architecture makes possible to select the outlines (“masks”) of different objects in photographs, even if there are several such instances, they have different sizes and partially overlap. The network is also capable of recognizing the poses of objects in an image.

Main steps of approach 2:

1) Input an image as input and pass it to the ConvNet, which returns the feature map for that image.

2) Region proposal network (RPN) is applied on these feature maps. This returns the object proposals along with their objectness score.

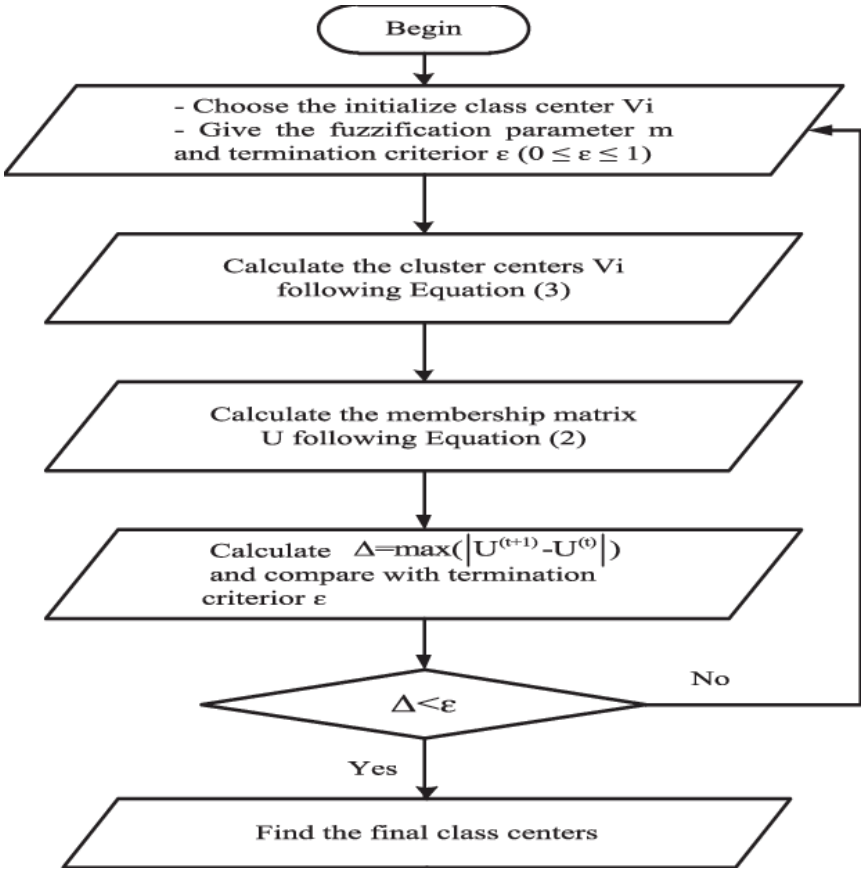
3) A RoI pooling layer is applied on these proposals to bring down all the proposals to the same size.

4) Finally, the proposals are passed to a fully connected layer to classify and output the bounding boxes for objects. It also returns the mask for each proposal.

Approach 3: Fuzzy c means clustering algorithm

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. FCM is the advanced version of K-means clustering algorithm and doing more work than K-means. This algorithm best result for overlapped data set and comparatively better than k-means algorithm. Unlike k-means where data point must exclusively belong to one cluster center here data point is assigned membership to each cluster center as a result of which data point may belong to more than one cluster center.

Main steps of approach 3:



3.2.1.1.2 Fuzzy c-means clustering algorithm steps.

After applying these segmentation techniques, background will be removed. Segmented result will be sent to the classification model. Our proposed method is not just finding objects in light and untextured background but also it finds the objects even if they are in a complex background. And it would be done with high accuracy and best performance.

3.2.2 Identification and Classification of Ayurvedic Plants in Sri Lanka Using A Mobile Application in An Offline Environment Approach

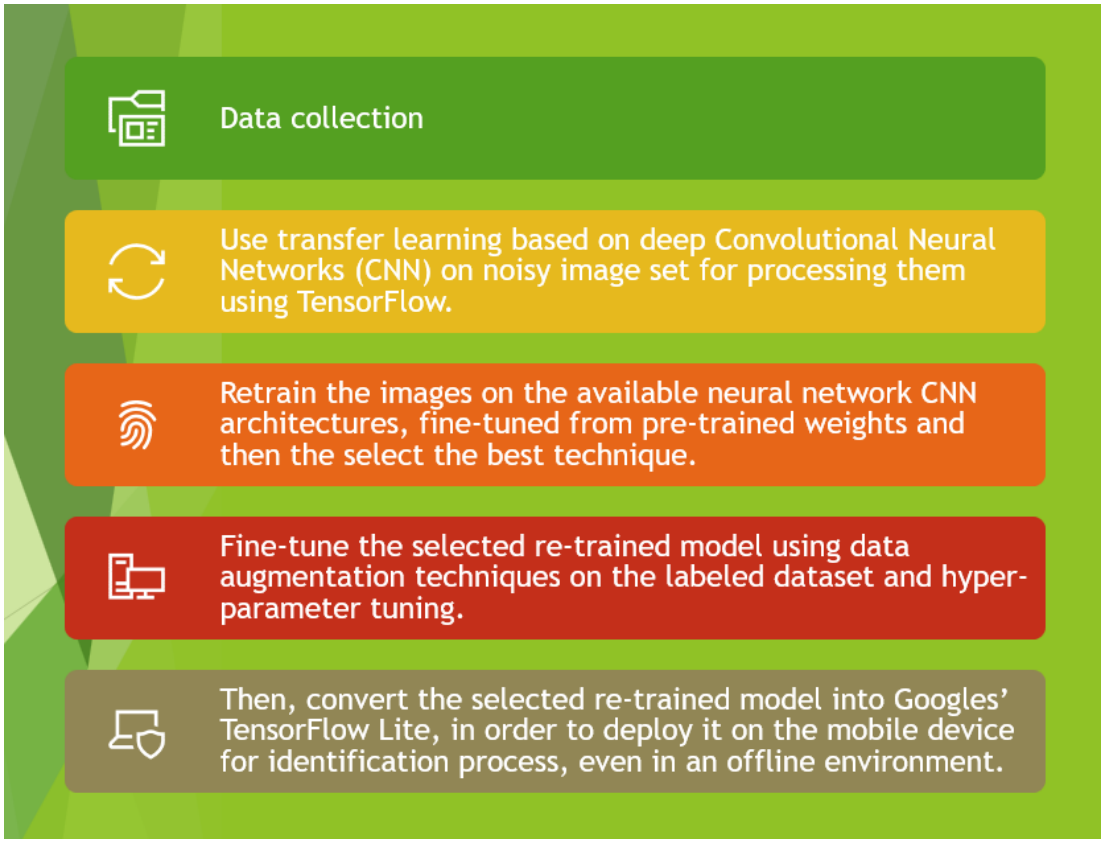


Figure 3.2.2.1 Methodology of the approach

In this research, the images of the leaf/digested root/fruit that will be scanned using the application will act as the input for this phase. Among many herbal plants, 5 most important ayurvedic plants in Sri Lanka will be chosen to analyze further in detail and the images of the plants will be acquired from plant nursery of Navinna Ayurveda Medical Hospital, social media, Institute of Ayurveda, Alternative medicine website and blogs related to Sri Lankan herbal plants creating a noisy web data set.

After acquiring the images of leaves/fruit/digested root of the selected Ayurveda plants, they will be labelled and annotated using multi-class classification, by forming 5 classes of the particular plants. After the dataset is prepared, transfer learning based on deep Convolutional Neural Networks (CNN) will be used on noisy image set for processing them using TensorFlow, in a local computer. However, training a model

from scratch is too costly. To overcome this challenge transfer learning will be applied in the Ayurvedic dataset. The proposed identification method will be based on running a CNN.

After that, the dataset will be retrained on the available neural network CNN architectures (such as Google Net, Inception v2, Inception v4, VGG-16, VGG-19 architectures), then will be fine-tuned from pre-trained weights and then the best technique with the highest accuracy will be selected.

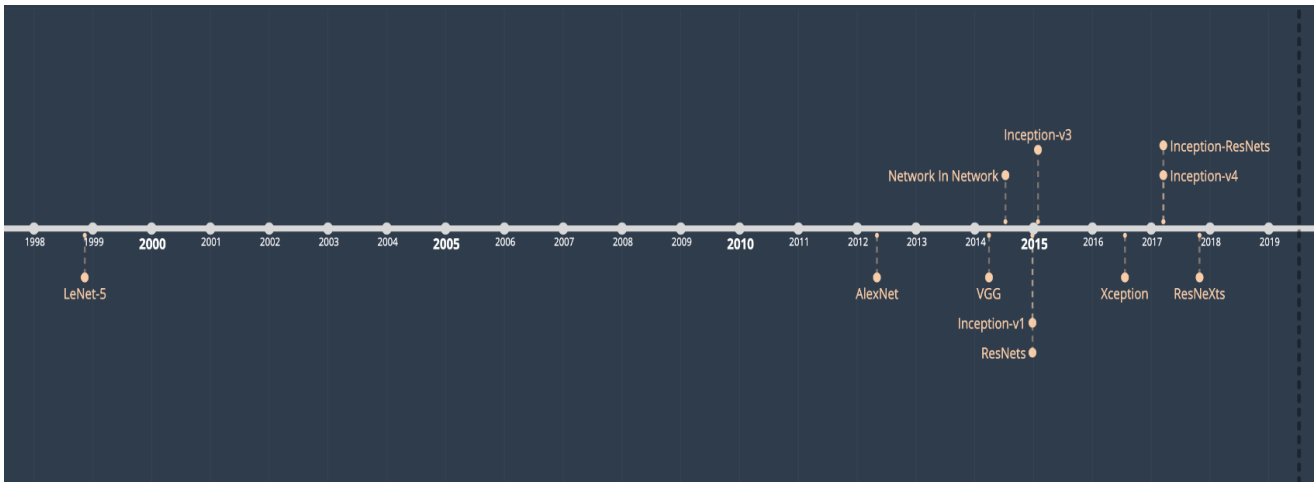


Figure 3.2.2.2 The 10 architectures and the year their papers were published.

Source: https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614

Then, the selected re-trained model will be fine-tuned using data augmentation techniques on the labeled dataset and hyper-parameter tuning. It will be re-trained with the dataset using Keras which is an open source neural network library, written in python. It is user friendly, modular and extensible since it has been designed for fast experimentations with deep neural networks.

Then the re-trained model will be converted for the deployment on mobile devices with Googles' TensorFlow Lite. TensorFlow Lite is a TensorFlow's lightweight solution for mobile devices which provides on-device machine learning inference with a small binary size and low latency. TensorFlow provides an interface for expressing machine learning algorithms and an application for executing these algorithms [36].

One of the main challenges in detection of Ayurveda plants are similarity between different plant species in different phases of their life cycles and complexity of the background since the end user will be using this application in deep forest environment. Therefore, we cannot just rely on a single feature, such as color, texture or shape to distinguish them. The same species of leaves will be different because of the shades of colors, shape, scale, viewpoint etc. [35].

On a mobile device the Ayurveda plant detection has to be done with time, battery life critical manner, especially when it has to be done in a forest area .In the proposed system the whole identification process will take place on mobile devices and it doesn't require internet .Therefore this will be a great solution to identify invasive plants in deep forest areas, where mobile network coverage is not available .The mobile application will be built using android.

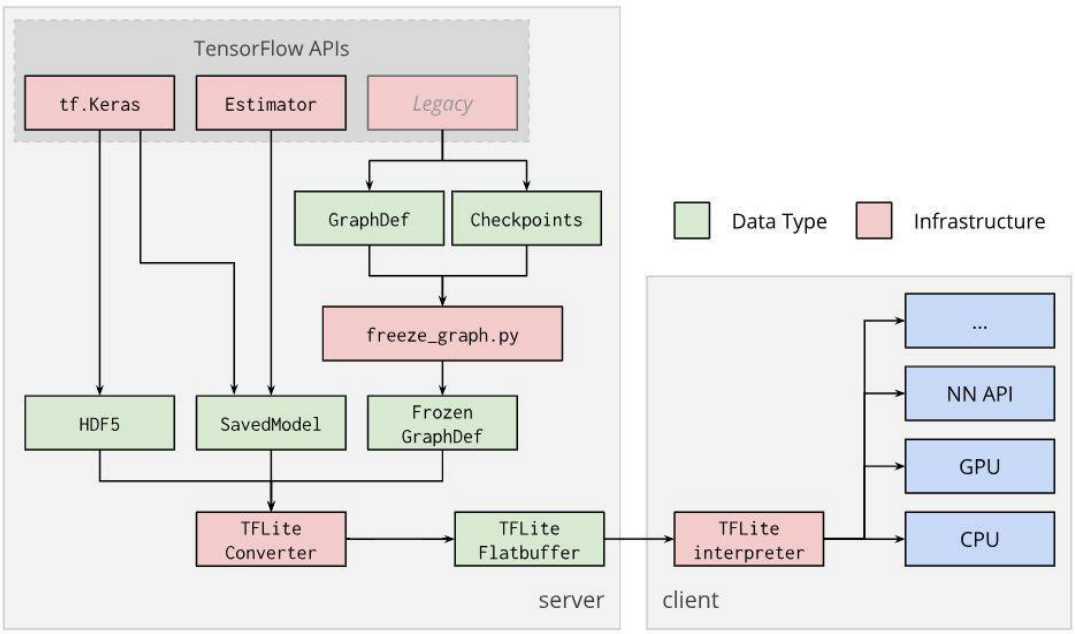


Figure 3.2.2.3: TensorFlow Lite Converter

Source: https://www.tensorflow.org/lite/convert

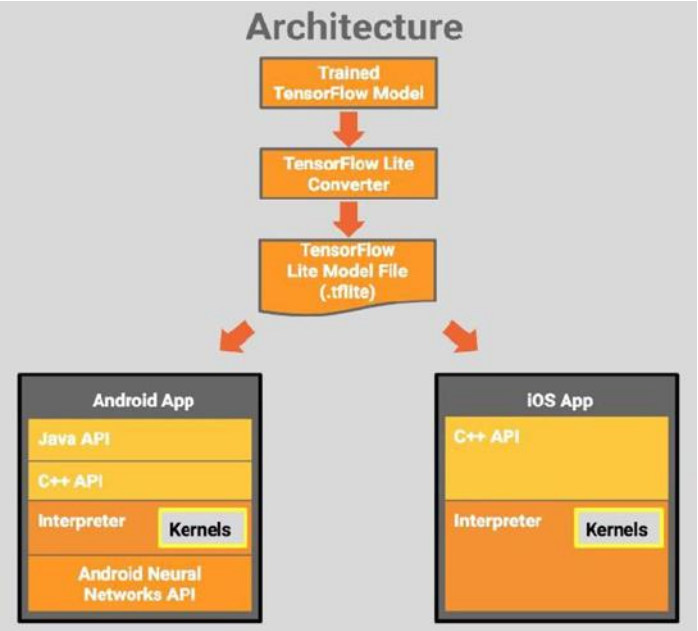


Figure 3.2.2.4: TensorFlow Lite Architecture

Source: https://www.tensorflow.org/lite/guide

3.2.2.1 High level system architecture diagram

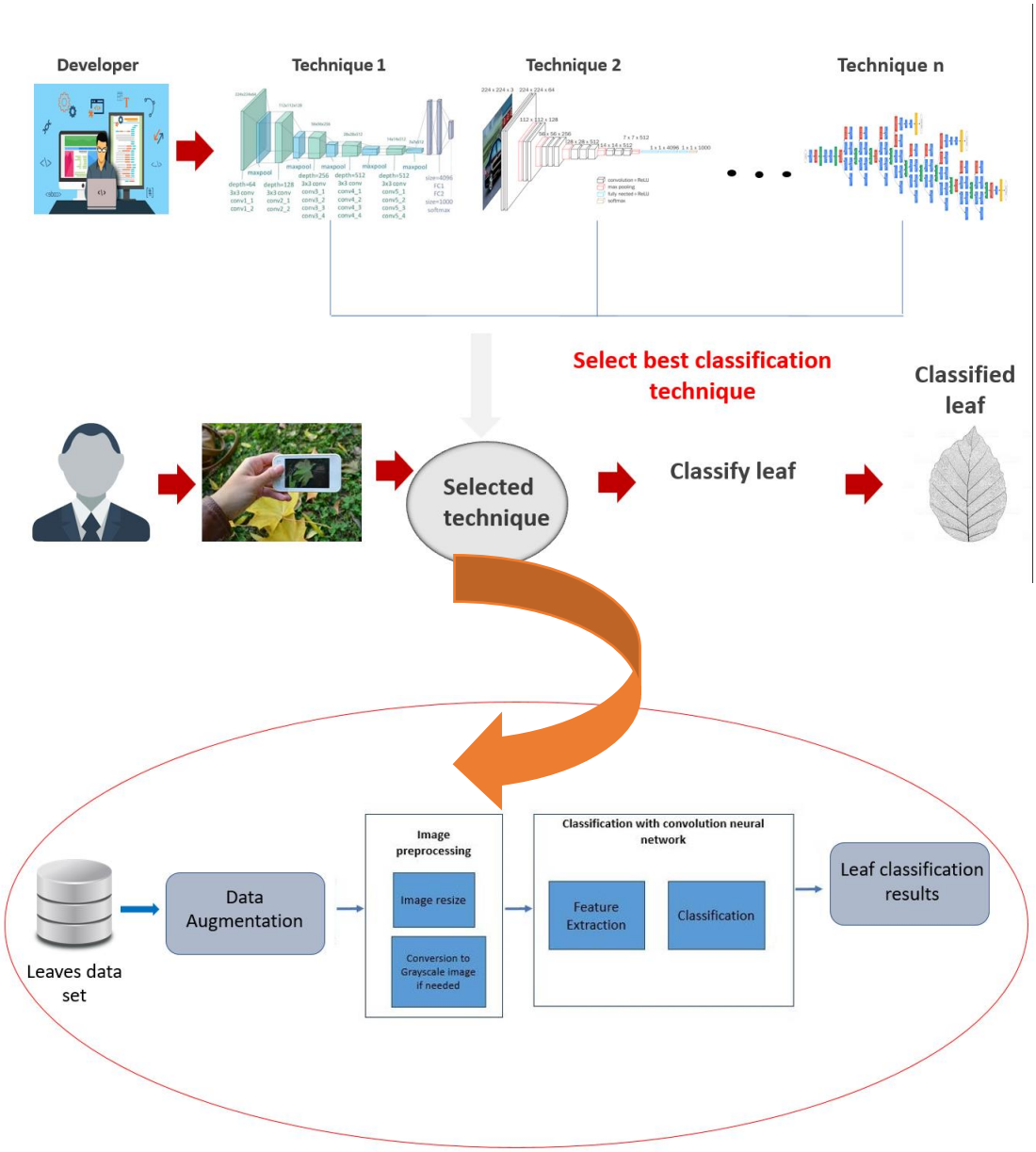


Figure 3.2.2.1.1 High level system architecture diagram

3.2.2.2 Expected Ways of Testing the Accuracy of the selected predictive models

From all the data collected (captured photos of the tree leaves/digested roots/fruits, digital images, web images, etc.), after the process of data annotation, labelling into 5 classes for the purpose of annotation and totally pre-processing; the prepared dataset will be divided into 3 parts as:

1. 70% of data for training the model – **Training dataset**
2. 15% of data for validating the model – **Validation dataset**
3. **15% of data for testing the model – Test dataset**

3.2.3 Extractive Information Summarization and Location Mapping with Smart Visualization on Ayurvedic Plants in Sri Lanka Using a Mobile Application in an Offline and Online Environment Approach

Main research area of this component is regarding dynamic information extraction and information summarization. So, in order to achieve this goal, many techniques should be followed to select the best model out of them. Thus, several existing algorithms will be analyzed, and accuracy will be tested in order to select the suitable algorithms to extract the required Information from dynamic web pages to generate a summary on medicinal plants. Here is a succinct definition to summarization according to Text Summarization Techniques, A Brief Survey, 2017, "Automatic text summarization is the task of producing a concise and fluent summary while preserving key information content and overall meaning". In general, there are two different approaches that are used for text summarization called Extractive Summarization and Abstractive Summarization. In this research Abstractive Summarization will be used as it is a very interesting approach. Here, we generate new sentences from the original existing text. In contrast, extractive approach uses only the sentences that were present in the text. The sentences generated through abstractive summarization might not be present in the original text.

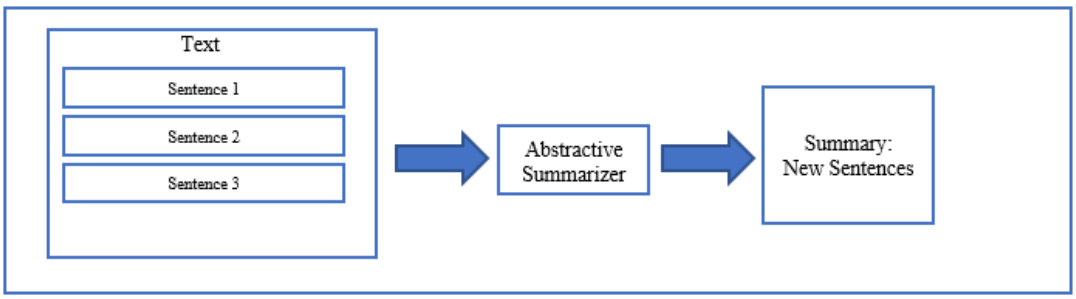


Figure 3.2.3.1: Abstraction summarization process

Our expectation is to build a text summarizer where the input is a long sequence of words extracted from web pages and the output is a short summary of sequence of words. Thus, we can model this as a Many-to-Many Seq2Seq problem. Below is a typical Seq2Seq model architecture:

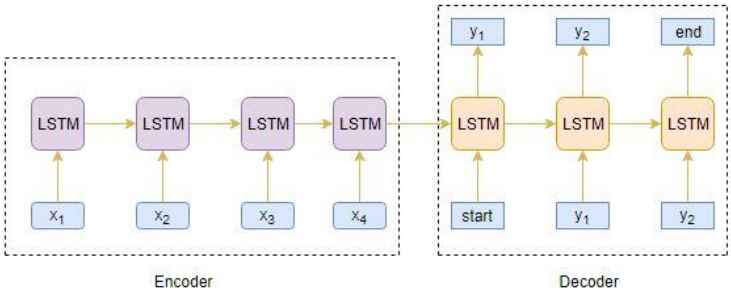


Figure 3.2.3.2: Seq2Seq model

There are two major components of a Seq2Seq model as, Encoder and the Decoder [51].

Generally, variants of Recurrent Neural Networks (RNNs), i.e. Gated Recurrent Neural Network (GRU) or Long Short-Term Memory (LSTM), are preferred as the encoder and decoder components. This is because they are capable of capturing long term dependencies by overcoming the problem of vanishing gradient.

We can set up the Encoder-Decoder in 2 phases:

- Training phase
- Inference phase

3.2.3.1 Training phase

In the training phase, we will first set up the encoder and decoder. We will then train the model to predict the target sequence offset by one timestep. Let us see in detail on how to set up the encoder and decoder.

3.2.3.1.1 Encoder

An Encoder Long Short-Term Memory model (LSTM) reads the entire input sequence wherein, at each timestep, one word is fed into the encoder. It then processes the information at every timestep and captures the contextual information present in the input sequence. Below diagram which illustrates this process:

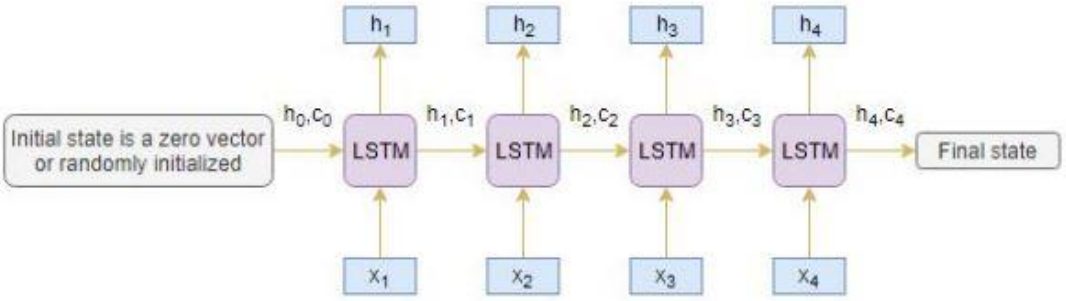


Figure 3.2.3.1.1.1: Encoder Process

The hidden state (hi) and cell state (ci) of the last time step are used to initialize the decoder. Remember, this is because the encoder and decoder are two different sets of the LSTM architecture.

3.2.3.1.2 Decoder

The decoder is also an LSTM network which reads the entire target sequence word-by-word and predicts the same sequence offset by one timestep. The decoder is trained to predict the next word in the sequence given the previous word.

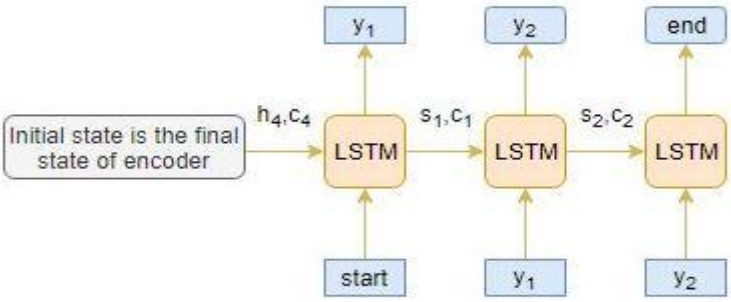


Figure 3.2.3.1.1.2: Decoder Process

<start> and <end> are the special tokens which are added to the target sequence before feeding it into the decoder. The target sequence is unknown while decoding the test sequence. So, we start predicting the target sequence by passing the first word into the decoder which would be always the <start> token. And the <end> token signals the end of the sentence

3.2.3.2 Inference Phase

After training, the model is tested on new source sequences for which the target sequence is unknown. So, we need to set up the inference architecture to decode a test sequence:

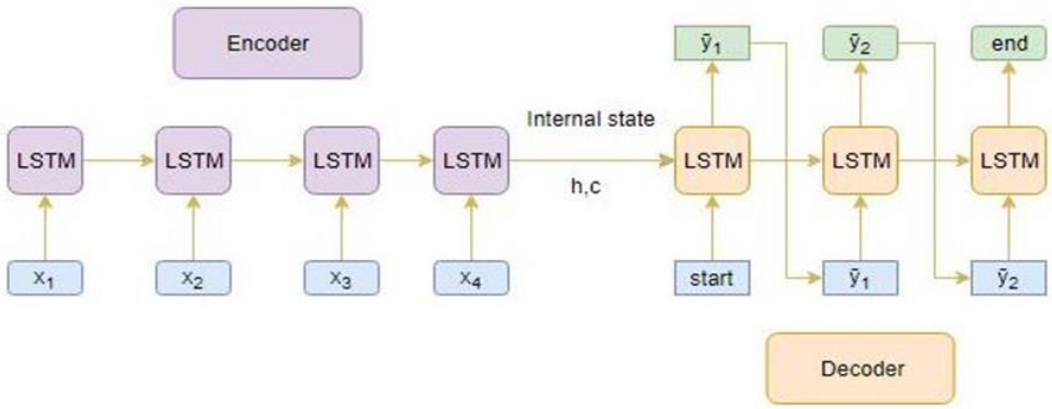


Figure 3.2.3.2.1: Encoder Process

Here are the steps to decode the test sequence:

1.Encode the entire input sequence and initialize the decoder with internal states of the encoder

2.Pass <start> token as an input to the decoder

3.Run the decoder for one timestep with the internal states

4.The output will be the probability for the next word. The word with the maximum probability will be selected

5.Pass the sampled word as an input to the decoder in the next timestep and update the internal states with the current time step

6.Repeat steps 3 – 5 until we generate <end> token or hit the maximum length of the target sequence

So that, with the support of encoder-decoder architecture Arogya will be able to manage the summary generation functionality.

3.2.4 Text classification Herbal diseases and Feature Extraction of Ayurveda medicine

3.2.4.1 Process Model for supervised Text classification (High-Level Diagram)

Text Classification is an example of supervised machine learning task since a labelled dataset containing text documents and their labels is used for train a classifier.

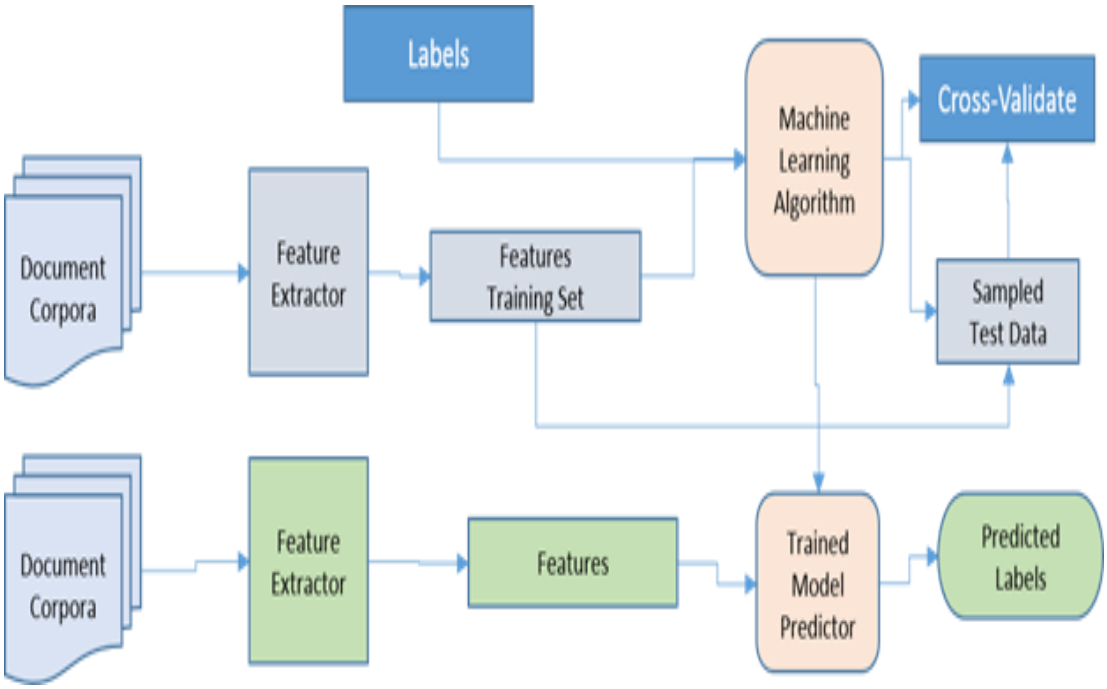


Figure 3.2.4.1.1:Supervised Text Classification

This is using for a Text classification process while typing its sentiments of needed herbs. Generated data has a variety of tabular data columns have either numerical or categorial data. NLP (Natural Language Processing) is applicable in several problematic from speech recognition, language translation. Classifying documents to information extraction. This is helps identified sentiment of herbs, finding entities in the sentence, category of texts. NLP enables the computer to interact with human in a natural manner. It helps the computer to understand the human language and derive meaning from it.

3.2.4.2 System Architecture

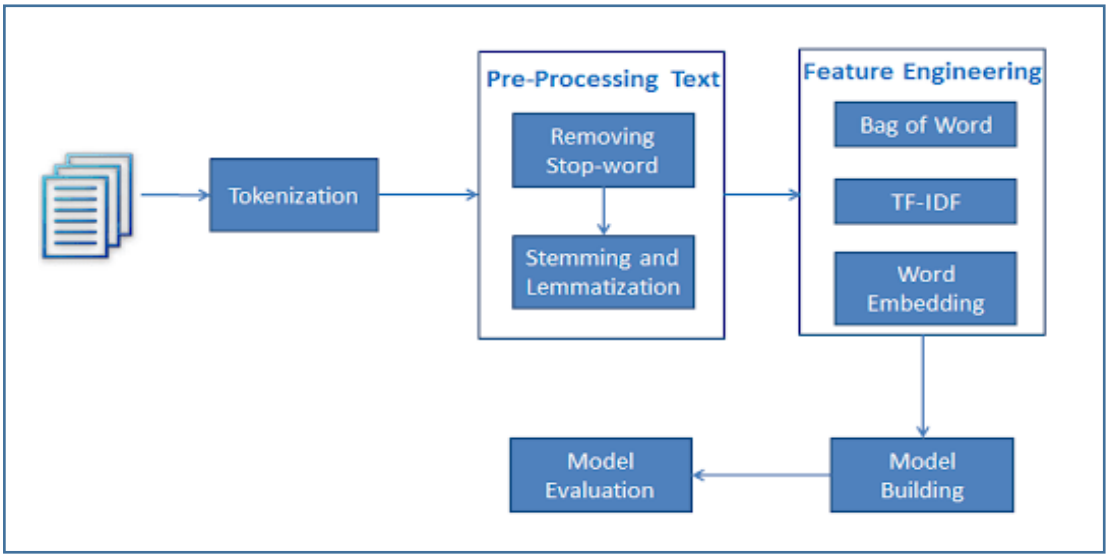


Figure 3.2.4.2.1: System Architecture

Text classification process regarding the flow is implemented under the NLP. In order to that, NLP supports here to identify the category of texts and Information extraction. There end to end text classification which is a pipeline consists three main components. According to the Figure 3.2. It represents Dataset Preparation (Pre-processing Text), Feature Engineering, and Model Training are the major components of this classification process. Therefore, consisting some points regarding the each of levels describing briefly as per in below.

1. **Dataset Preparation**:

Loading a relevant dataset and performing basic pre-processing. If there including unwanted things which should be ejected on this stage. Hence, the dataset split into train and validation sets. The Traditional text classifiers usually break the documents into small word fragments(n-grams) and locate them as separate dimensions in the fragment hyperspace. These steps will be used for a typing field provide for special properties/features

According to the application, Giving an Interface including a separated dropdown buttons for selecting most appropriate words related to the medicine plants. After selecting all the dropdowns, there will be consisted a space for typing a special property if there have.

2. Feature Engineering:

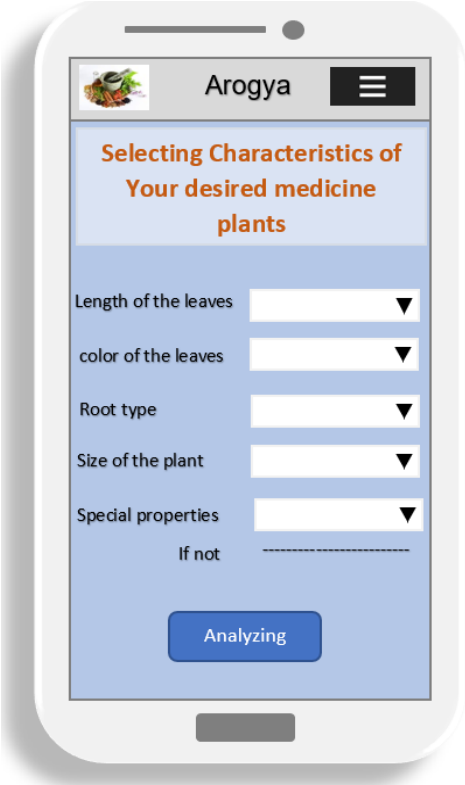
Raw dataset transforms into flat features using Machine Learning model and This process is going on creating new features from the existing data. Raw text data will be transformed into feature vectors and new features will be created using an existing text data. According to the new features from dataset can be identified different ideas. Such as, Count vectors as features, TF-IDF Vectors as features, Word as features, Text NLP based features, Topic model as features. But this function not able to do such type of bored things to classify. Hence, Naïve bayes is used for this all the NLP based TF-IDF.

According to the application, providing a same dropdown option to selecting the specific characteristics if the plants be consisting on. If this dropdown will not be included in needed characteristics, then giving a limited space for typing it with main points. So that, there only typing field must consider for the feature extraction part. There is a space for typing an any special feature of the medicine plant in provided area. In there should be considered about filtering options such as removing emoji, punctuation marks, spaces, special characteristics etc. Then it will be transformed into feature vector. This field is not a required filed and it's not use in always because there are most of the special characteristics are included in dropdown field.

3. Model Training:

Machine learning model is trained on a labeled dataset. In there, data which are taken from dropdown fields should be moved to analyze options using Algorithms.

Other data which are from typing area should have to label and categorize according to the relevant types of category diseases. Before that these types of data coming after doing previous steps. (Pre-processing and Feature Extraction). After going on both components, should have to Improve the performances of text classifiers.



According to the application,

This is the Planned Mobile UI of this Analyzing the category of diseases part. All the Fields should be required excepting a provided typing area. Finally, all the filed be filed then pressing a button "Analyzing" move to the next interface. Then displaying a Category after taking some time period of loading. Apart from that this interface will be displayed some recipes related to the displayed category. Then user can select most usable recipe and giving a chance of visualizing there some of rare ingredients which are included.

3.2.4.3 Text Classification Using Tools

Both NLTK and TextBlob performs well in Text Classification processing.

Text Classification - NLTK

NLTK is a very big library holding 1.5GB and has been trained on a huge data with proving different dataset in multiple languages which can deploy according to the functionality its be required. NLTK is a powerful Python package that provides a set of diverse natural language algorithms.

Text Classification - TextBlob

TextBlob library which is a python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as noun phrase extraction, sentiments analysis, classification, translation etc. this function will be applied for classification task. Under the Features there are some points to follow on.

1. Classification – Using Naïve Bayes and Decision Tree
2. Tokenization – Splitting text into words and sentences
3. Spelling correction
4. Emojis

3.2.4.3.1 Comparison of Text classification tools

There are lots of tools to work with NLP. NLTK, Spacy, Stanford Core NLP. These are the comparison of properties of tools

	NLTK	Spacy	Stanford Core NLP	TensorFlow	Allen NLP
Build an end-to-end production application	✓	✓	✓	✓	✗
Efficiency on CPU	✓	✓	✗	✗	✗
Train models from own data	✓	✓	✓	✓	✓
Different neural network architectures for NLP	✗	✗	✗	✓	✓

Table 3.2.4.3.1.1 : Comparison of Tools

3.2.4.4 Text Classification Using Algorithms

Consisting of the most common algorithms such as *tokenizing*, part-of-speech tagging, topic segmentation, named entity recognition, etc. when considering this function, it helps the computer analysis, pre-process, and understand the written text. In This part, only use for the typing field and this field is consist with limited words as key words of herbal characteristics. Because if we give a large description there should have more stuffs further.

Text Classification Steps for typing field:

1. Loading data and Creating classifiers:

First create custom classifiers using TextBlob module. Before that creating some training and test data. Then creating a Naïve Bayes Classifier for passing the training data into Constructor.

2. Loading data from Files:

Loading data from common file formats including CSV, JSON, and TSV

3. Classifying Text or Classifying TextBlobs
4. Evaluating Classifiers

   Compute the accuracy of the test data using a relevant method

5. Updating Classifiers with New Data
6. Feature Extractions.

We can use our own extractors to identify each specialty.

Text Classification – Flow Chart

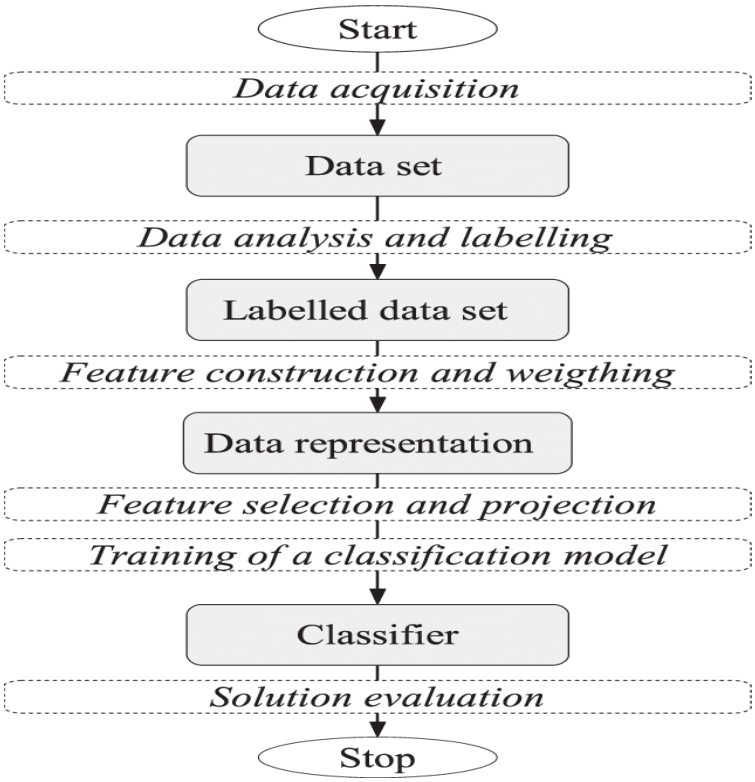


Figure 3.2.4.4.1:Text Classification flowchart

After that selecting a Most usable medicine recipe and visualize ingredients as Feature extraction. Here take those images from the database which is used to store all the images in initial stage.

3.4 Architecture Diagram

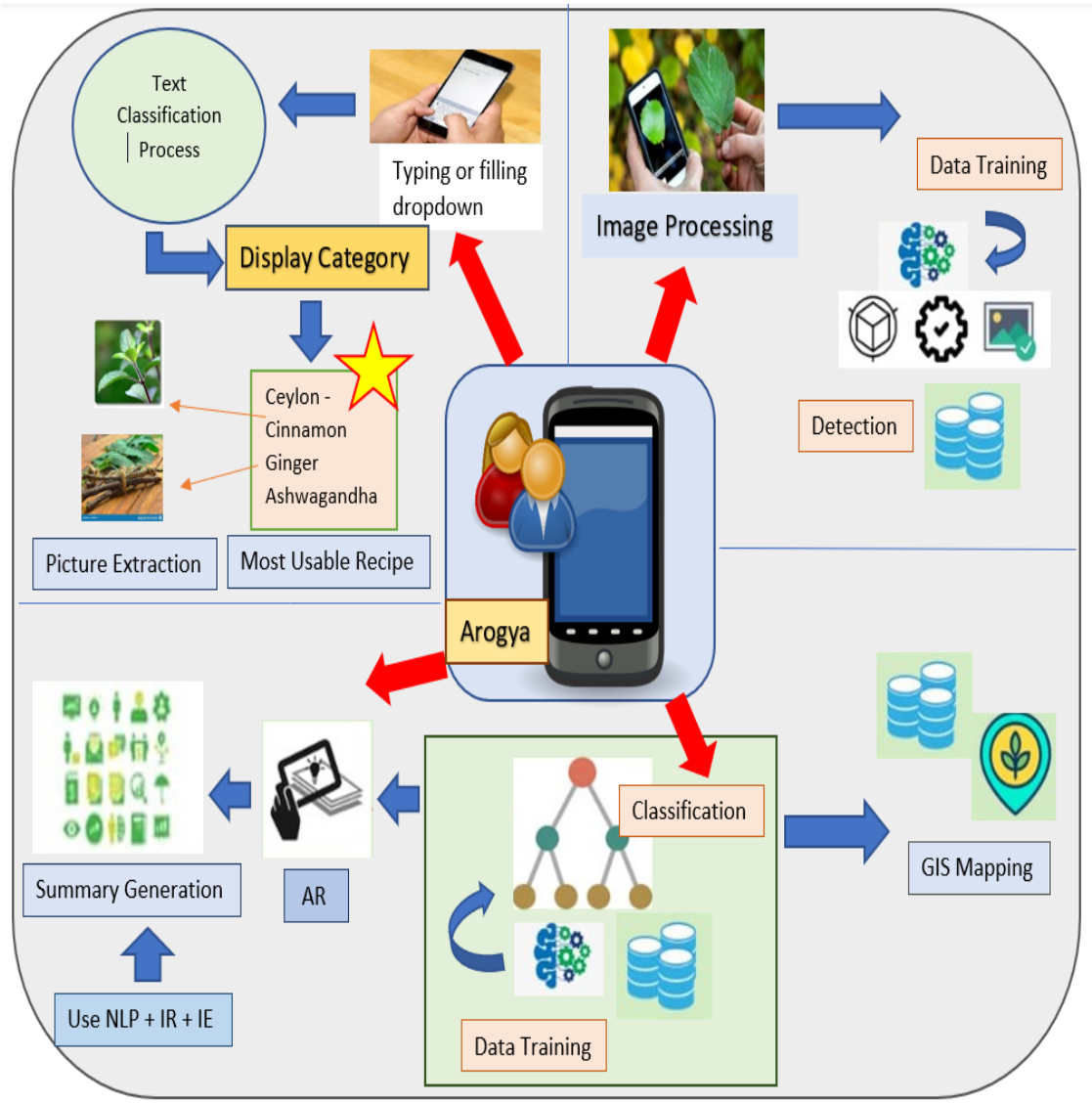


Figure 3.4.1: Software Architectural diagram

4. DESCRIPTION OF PERSONAL AND FACILITIES

Facilitators: -

Mrs. Lokesha Weerasinghe - Sri Lanka Institute of Information Technology (SLIIT)

Dr. Dharshana Kasthurirathna - Sri Lanka Institute of Information Technology (SLIIT)

Facilities: -

Navinna Ayurveda Medical Hospital

Institute of Ayurveda

Plant nursery of Navinna Ayurveda Medical Hospital

4.1 Gantt chart

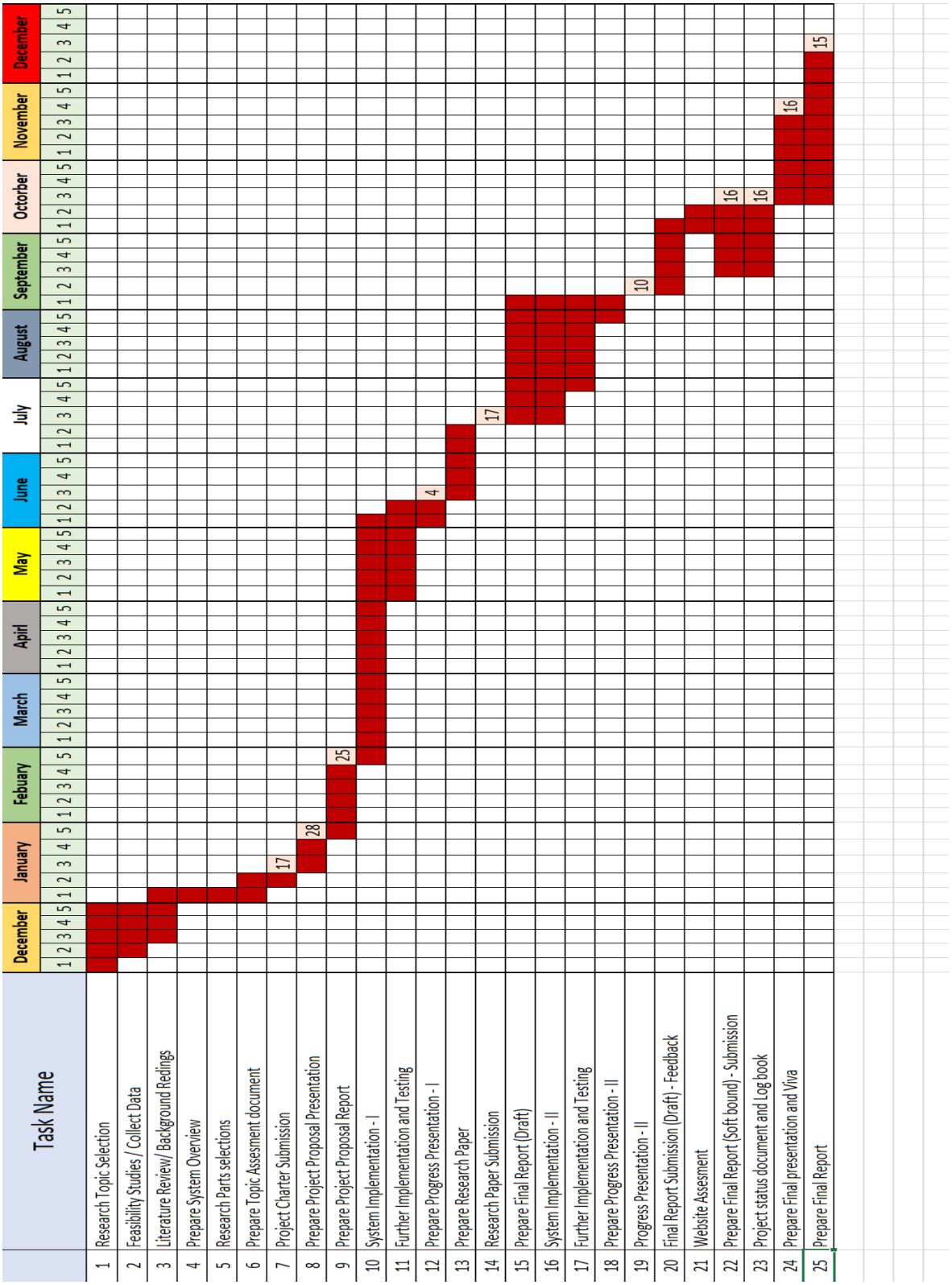


Figure 4.1.1: Gantt chart

4.2Workload Allocation

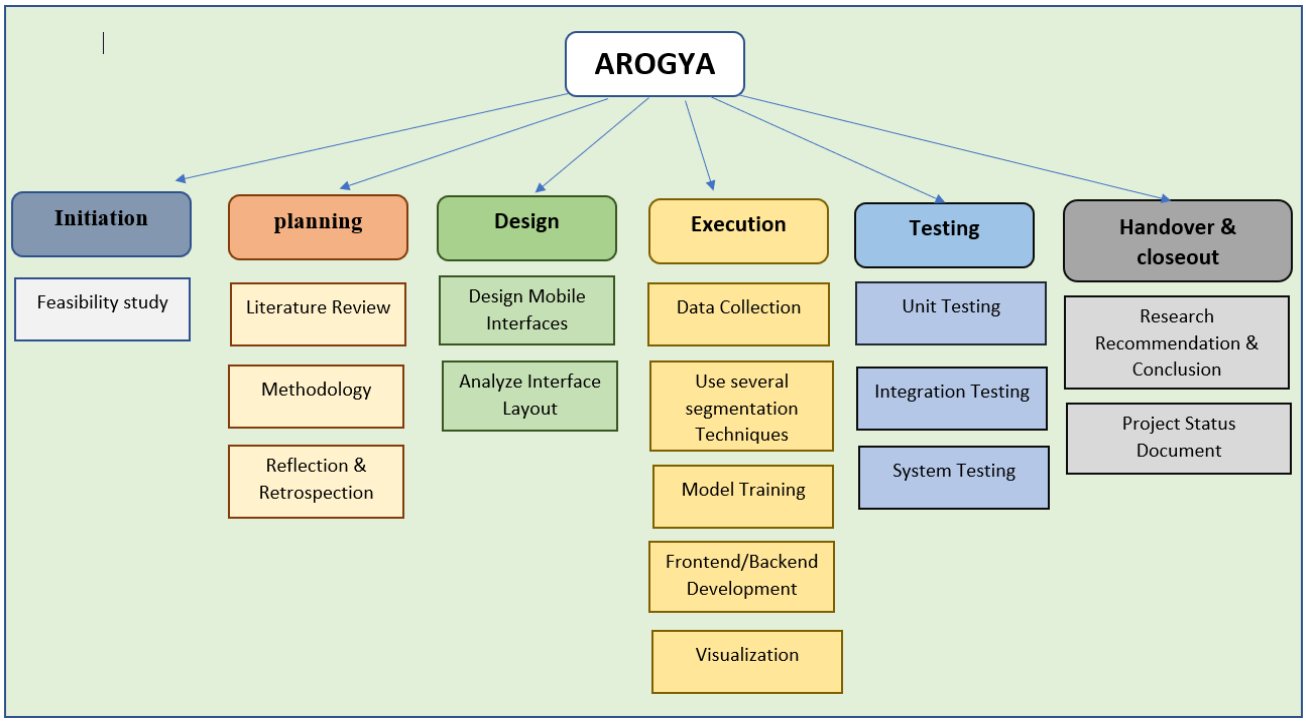


Figure 4.2:1 Work Breakdown Structure

4.3 Requirements

4.3.1 Functional Requirements

1. Select the best segmentation technique with comparing several segmentation techniques.
2. Classify a set of selected ayurvedic herbs in an offline environment.
3. Compare accuracies of the selected segmentation and classification techniques
4. Request to re-train the existing model with new datasets by premium uses.
5. Show classification history on the mobile device
6. Manage Ayurveda herb details in mobile device.
7. Display information on the top of the plant species.
8. Summary should be generated regarding the medical plant.
9. Application should be able to track the location of users.
10. Locations should be able to be viewed as a map view.

4.3.2 Non-Functional Requirements

1. The application should be able to give the results as fast as possible.
2. Accuracy should be high, and results should be efficient.
3. Feasible in order to understand the functionalities of the app
4. User friendly Interfaces should be provided
5. User Experiences should be properly managed in order to achieve a specific functionality

5. TECHNOLOGIES

1. Python
2. Keras
3. TensorFlow
4. TensorFlow Lite
5. Google Colab
6. Flutter
7. Android
8. Firebase – for the database
9. Google Information System
10. SpaCy
11. NLTK
12. Genism
13. Sumy

Research Area – Deep Learning Algorithms (deep neural network architectures based on transfer learning) and Image Processing Techniques, Natural Language Processing, Augmented Reality

6. BUDGET AND BUDGET JUSTIFICATION

Requirements	Cost
Internet Cost (per month)	12*4*2000.00
Documentation printout cost	3000.00
Data collection travelling charges	3000.00
Mobile phone as a resource	40000.00
Total Target Cost (per year)	**142000.00**

Table 6.1: Budget & Cost Estimation

7.COMMERCIALIZATION

7.1 Target Audience

- People who use ayurvedic treatment

- Researchers in the field of botany, medicine, chemical structure analysis, agriculture, ayurvedic medicinal practitioners, forest department officials, those who are involved in the preparation of ayurvedic medicines and others who are concerned with plant studies
- Doctors, Students, locals and foreigners
- Ayurvedic plant sellers

7.2 Market Space

- No age limitations for the users
- No need of advance computer literacy
- No need of advance knowledge in Ayurveda field

7.3 Revenue Earning

- Through subscription fee
- Revenue via additional services

8.REFERENCES

[1] Pathirage Kamal Perera, "Current scenario of herbal medicine in Sri Lanka", ASSOCHAM, 4th annual Herbal International Summit cum Exhibition on Medicinal & Aromatic Products, Spices and finished products(hi-MAPS), pp. 1-3, April 2012.

[2] Lakpura LLC, Ayurveda. Accessed on: February 15,2020. [Online]

Available: https://lanka.com/about/interests/ayurveda/

[3] Y. R. Azeez, R. A. C. P. Rajapakse, "An application of image processing techniques in identifying herbal plants in Sri Lanka", Proceedings of the 3rd. International Research Symposium on Pure and Applied Sciences, pp.1, 26th October 2018.

[4] Ayurveda Bansko, Blog, Benefits of Ayurveda and the difference to the traditional western medicine, 2015. Accessed on: February 15,2020. [Online]

Available: https://www.ayurvedabansko.com/benefits-ayurveda-and-difference-to-western-traditional-medicine/.

[5] Basavaraj S. Anami, Suvarna S. Nandyal, A. Govardhan, "A Combined Color, Texture and Edge Features Based Approach for Identification and Classification of Indian Medicinal Plants", International Journal of Computer Applications (0975 – 8887), Volume 6– No.12, pp.1-3, September 2010.

[6] Robert G. de Luna, Renann G. Baldovino, Ezekiel A. Cotoco, Anton Louise P. de Ocampo, Ira C. Valenzuela, Alvin B. Culaba, Elmer P. Dadios, "Identification of Philippine Herbal Medicine Plant Leaf Using Artificial Neural Network", IEEE, 978-1-5386-0912-5/17/$31.00,PP.1-2, 2017.

[7] Rademaker C. A. 2000, "The classification of plants in the United States Patent Classification system", World Patent Information 22: 301–307.

[8] Neeraj Kumar, Peter N Belhumeur, Arijit Biswas, David W Jacobs, W John Kress, Ida C Lopez, João VB Soares, "Leafsnap: A computer vision system for automatic plant species identification," ECCV, pp. 502–516. Springer, 2012.

[9] Cem Kalyoncu, Önsen Toygar, "Geometric leaf classification," Computer Vision and Image Understanding", vol. 133, pp. 102–109, 2014.

[10] Abdul Kadir, Lukito Edi Nugroho, Adhi Susanto, Paulus Insap Santosa, "Leaf classification using shape, color, and texture features," arXiv preprint arXiv:1401.4447, 2013.

[11] Thibaut Beghin, James S Cope, Paolo Remagnino, SarahBarman, "Shape and texture based plant leaf classification," Advanced Concepts for Intelligent Vision Systems, 2010, pp. 345–353.

[12] James Charters, Zhiyong Wang, Zheru Chi, Ah Chung Tsoi, David Dagan Feng, “Eagle: A novel descriptor for identifying plant species using leaf lamina vascular features,” ICME-Workshop, 2014, pp. 1–6.

[13] Heysem KAYA, Ilhan KEKLIK, Tolga ENSARI, Fatih ALKAN, Yagmur BIRICIK, “Oak Leaf Classification: An Analysis of Features and Classifiers”, 978-1-7281-1013-4/19/$31.00, pp.1-4, 2019.

[14] James S Cope, Paolo Remagnino, Sarah Barman, Paul Wilkin, “The extraction of venation from leaf images by evolved vein classifiers and ant colony algorithms,” Advanced Concepts for Intelligent Vision Systems, pp. 135–144, 2010.

[15] Muammer TORKOGLU, Davut HANBAY, “Combination of Deep Features and KNN Algorithm for Classification of Leaf-Based Plant Species”.

[16] Jing Hu, Zhibo Chen, Meng Yang, Rongguo Zhang, Yaji Cui, “A Multiscale Fusion Convolutional Neural Network for Plant Leaf Recognition”, 1070-9908, IEEE SIGNAL PROCESSING LETTERS, VOL. 25, NO. 6, JUNE 2018.

[17] Agnieszka Mikołajczyk, Michał Grochowski, “Data augmentation for improving deep learning in image classification problem”, 17859931, ISBN: 978-1-5386-6143-7, May 2018.

[18] Will Koehrsen, Medium, towards data science, Transfer Learning with Convolutional Neural Networks in PyTorch, November 27, 2018. Accessed on: February 15,2020. [Online]

Available: https://towardsdatascience.com/transfer-learning-with-convolutional-neural-networks-in-pytorch-dd09190245ce.

[19] Inside.Techlabs, Leaf Classification Project, October 23, 2018, Accessed on: February 15,2020. [Online]

Available:https://medium.com/@inside.techlabs/leaf-classification-project-89bb5c7577f3.

[20] Dileep M.R, Pournami P.N, “AyurLeaf: A Deep Learning Approach for

Classification of Medicinal Plants”, IEEE, 978-1-7281-1895-6/19/$31.00, pp.1-4, 2019.

[21] Jun Woo Lee, Yeo Chan Yoon, "Fine-Grained Plant Identification using wide and deep learning model", Institute for Information & communications Technology Promotion (IITP), No. 2016-0-00010-003, Digital Content In-House R&D.

[22] Benjaphan Sommana, Thanaruk Theeramunkong, “Improving Plant Recognition using Hybrid features from Connectionist and Knowledge-Based Approaches”, IEEE, 978-1-7281-0161-3/18/$31.00, 2018.

[23] Hulya Yalcin, Salar Razavi, “Plant Classification using Convolutional Neural Networks”, Visual Intelligence Laboratory Electronics & Telecommunication Engineering Department Istanbul Technical University, 2017.

[24] Jyotismita Chaki, Ranjan Parekh, Samar Bhattacharya, “Recognition of Whole and Deformed Plant Leaves using Statistical Shape Features and Neuro-Fuzzy Classifier”, IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS), 978-1-4799-8349-0/15/$31.00, 2015.

[25] Chaoyun Zhang, Pan Zhou, Chenghua Li, Lijun Liu, “A Convolutional Neural Network for Leaves Recognition Using Data Augmentation”, IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, DOI 10.1109/CIT/IUCC/DASC/PICOM.2015.318, 978-1-5090-0154-5/15 $31.00, 2015.

[26] Sue Han Lee, Paul Wilkin, Chee Seng Chan, Paolo Remagnino, “DEEP-PLANT: PLANT IDENTIFICATION WITH CONVOLUTIONAL NEURAL NETWORKS”, IEEE, 978-1-4799-8339-1/15/$31.00, 2015.

[27] ArunPriya C, Balasaravanan T, Antony Selvadoss Thanamani, “An Efficient Leaf Recognition Algorithm for Plant Classification Using Support Vector Machine”, Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, 978-1-4673-1039-0/12/$31.00, 2012, March 21-23, 2012.

[28] Shanwen Zhang, YouQian Feng, “Plant Leaf Classification Using Plant Leaves Based on Rough Set”, 2010 International Conference on Computer Application and System Modeling (ICCASM 2010), 978-1-4244-7237-6/10/$26.00, 2010.

[29] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang, Qiao-Liang Xiang, “A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network”, IEEE International Symposium on Signal Processing and Information Technology, 978-1 -4244-1 835-0/07/$25.00 , 2007.

[30] Pushpa BR, Anand C, Mithun Nambiar P, “Ayurvedic Plant Species Recognition using Statistical Parameters on Leaf Images”, International Journal of Applied Engineering Research, ISSN 0973-4562, Volume 11, Number 7 (2016) pp 5142-5147.

[31] Sethulekshmi A V, Sreekumar K, “Ayurvedic leaf recognition for Plant Classification”, Sethulekshmi A V et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6), ISSN:0975-9646, 2014, 8061-8066.

[32] A Gopal, S. Prudhveeswar Reddy, V. Gayatri, “Classification of Selected Medicinal Plants Leaf Using Image Processing”, IEEE, 978-1-4673-2322-2112/$31.00, 2012.

[33] J. W. Tan, S. Chang, S. Binti Abdul Kareem, H. J. Yap, K. Yong, “Deep learning for plant species classification using leaf vein morphometric”, pages 1–1, 2018.

[34] R. Janani, A. Gopal, “Identification of selected medicinal plant leaves using image features and ann”, 2013 International Conference on Advanced Electronic Systems (ICAES), pages 238–242, Sept 2013.

[35] Gavai, N. R., Jakhade, Y. A., Tribhuvan, S. A., & Bhattad, R. (2017, December). MobileNets for flower classification using TensorFlow. In 2017 International Conference on Big Data, IoT and Data Science (BID) (pp. 154-158). IEEE.

[36] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016). Tensorflow: A system for large-scale machine learning. In 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16) (pp. 265-283).

[37] MathWorks, Deep Learning. Accessed on: February 19,2020. [Online]

Available: https://www.mathworks.com/discovery/deep-learning.html.

[38] Machine Learning Mastery, How to Improve Performance With Transfer Learning for Deep Learning Neural Networks. Accessed on: February 19,2020. [Online]

Available: https://machinelearningmastery.com/how-to-improve-performance-with-transfer-learning-for-deep-learning-neural-networks/.

[39] TechTarget, Convolutional Neural Network. Accessed on: February 19,2020. [Online]

Available: https://searchenterpriseai.techtarget.com/definition/convolutional-neural-network

[40] Hervé Goëau, Pierre Bonnet, Alexis Joly, Julien Barbe, Souheil Selmi, Vera Baki´c, Jennifer Carré, Daniel Barthelemy, Nozha Boujemaa, “Pl@ntNet Mobile App”, Proceedings of the 21st ACM international conference on Multimedia, DOI: 10.1145/2502081.2502251, ppt. 1-2, October 2013.

[41] Neeraj kumar Kumar, Peter N. Belhumeur, Arijit Biswas, João V. B. Soares, “Leafsnap: A Computer Vision System for Automatic Plant Species Identification”, European Conference on Computer Vision, DOI:10.1007/978-3-642-33709-3_36, ppt. 1, October 2012.

[42] Desta Sandya, Yeni Herdiyeni, “MedLeaf: Mobile Application For Medicinal Plant Identification Based on Leaf Image”, International Journal on Advanced Science, Engineering and Information Technology, Vol. 3 (2013) No. 2 ISSN: 2088-5334, ppt 1-3, May 2013.

[43] Lin-Hai Ma, Zhong-Qiu Zhao, Jing Wang, “ApLeafis: An Android-Based Plant Leaf Identification System”, International Conference on Intelligent Computing, LNCS 7995, pp. 106–111, 2013

[44] H. J. Gunathilaka, P. Vitharana, L. Udayanga, and N. Gunathilaka, “Assessment of Anxiety, Depression, Stress, and Associated Psychological Morbidities among Patients Receiving Ayurvedic Treatment for Different Health Issues: First Study from Sri Lanka,” BioMed Research International, vol. 2019, pp. 1–10, Nov. 2019, doi: 10.1155/2019/2940836.

[45]J.-R. Reichert, K. L. Kristensen, R. R. Mukkamala, R. Vatrapu, “A supervised machine learning study of online discussion forums about type-2 diabetes,” IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), DOI: 10.1109/healthcom, 2017.8210815, 2017

[46] accenture, Natural Language Processing (NLP) Techniques for Extracting Information, Accessed on: February 24,2020. [Online]
Available:https://www.searchtechnologies.com/blog/natural-language-processing-techniques

[47] “Ayurveda modern medicine interface: A critical appraisal of studies of Ayurveda medicines to treat osteoarthritis and rheumatoid arthritis” [Online]
Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3087360/

[48] ResearchGate, Current scenario of herbal medicine in Sri Lanka. Accessed on: February 22,2020. [Online]
Available:https://www.researchgate.net/publication/224771841_Current_scenario_of_herbal_medicine_in_Sri_Lanka

[49]Ayurvedabansko, Benefits of ayurveda. Accessed on: February 20,2020. [Online]
Available:https://www.ayurvedabansko.com/benefits-ayurveda-and-difference-to-western-traditional-medicine/

[50] Neeraj Kumar, Peter N Belhumeur, Arijit Biswas, David W Jacobs, W John Kress, Ida C Lopez, João VB Soares, "Leafsnap: A computer vision system for automatic plant species identification" in ECCV, Springer, pp. 502-516, 2012.

[51] Analytics Vidhya, Comprehensive guide to text summarization. Accessed on: February 19,2020. [Online]
Available:https://www.analyticsvidhya.com/blog/2019/06/comprehensive-guide-text-summarization-using-deep-learning-python/

[52] Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO, (2014) “The biodiversity of species and their rates of extinction,
distribution, and protection”,pp. 987 , doi:10.1126/science.1246752, 2014.

[53]Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM,
Sexton JO, (2014) “The biodiversity of species and their rates of extinction,
distribution, and protection”,pp. 987 , doi:10.1126/science.1246752, 2014.

[54] Joly A, Müller H, Goëau H, Glotin H, Spampinato C, Rauber A, Bonnet P, Vellinga WP,
Fisher B ,"LifeCLEF: multimedia life species identification". In: International workshop on
environmental multimedia retrieval 2014, Glasgow,pp 345 -348,2014.

[55]Nursuriati Jamil, Nuril Aslina Che Hussin, Sharifalillah Nordin, Khalil Awang,
" Automatic Plant Identification: Is Shape the Key Feature?".
Accessed on: February 18,2020.
[Online]
Available:
https://www.researchgate.net/publication/290003529_Automatic_Plant_Identification_Is_Shape_the_Key_Feature

[56]AnalyticsVidhya,Introduction to image segmentations in python.Accessed on: February 18,2020.
Available on:https://www.analyticsvidhya.com/blog/2019/04/introduction-image-segmentation-techniques-python/

[57]K.Deepak, A.N.Vinoth," Leaf Detection Application For Android Operating System ",
in international conference on computation of power, energy, information and communication ,pp. 511-513,2014.

[58]Lin-Hai Ma , Zhong-Qiu Zhao, Jing Wang," ApLeafis: An Android-Based Plant Leaf
Identification System" in ICIC , Springer, pp. 106-107, 2013.

[59] ResearchGate, "MedLeaf: Mobile Application For Medicinal
Plant Identification Based on Leaf Image" . Accessed on: February 20,2020. [Online]
Available:https://www.researchgate.net/publication/237441676_MedLeaf_Mobile_Application_For_Medicinal_Plant_Identification_Based_on_Leaf_Image]

[60] Neeraj Kumar, Peter N Belhumeur, Arijit Biswas,
David W Jacobs, W John Kress, Ida C Lopez, João VB Soares, "Leafsnap: A computer vision system for
automatic plant species identification" in ECCV, Springer, pp. 502-517, 2012.

[61] Y. R. Azeez, R. A. C. P. Rajapakse, " An application of image processing techniques in identifying herbal plants in Sri Lanka" , Proceedings of the 3rd International Research Symposium on Pure and Applied Sciences, 26th October 2018 - Faculty of Science, University of Kelaniya, Sri Lanka

[62] H. P. Edmundson, "New methods in automatic extracting", Journal of the ACM, vol. 16, no. 2, pp. 264-285, 1969.

[63] Neeraj Kumar, Peter N Belhumeur, Arijit Biswas, David W Jacobs, W John Kress, Ida C Lopez, and João VB Soares, “ Leafsnap: A computer vision system for automatic plant species identification,” in ECCV, pp. 502– 516. Springer, 2012.

[64] H. P. Edmundson, "New methods in automatic extracting", Journal of the ACM, vol. 16, no. 2, pp. 264-285, 1969.

[65] J. Kupiec, J. Pedersen, F. Chen, "A trainable document summarizer", Proceedings SIGIR '95, pp. 68-73, 1995.

[66] D. R. Radev, H. Jing, M. Stys, D. Tam, "Centroid-based summarization of multiple documents", Information Processing and Management, vol. 40, pp. 919-938, 2004.