

AUTOMATIC RECOGNITION OF MEDICINAL PLANTS

Submitted in partial fulfillment of the requirements

for the award of

Bachelor of engineering degree in Computer Science and Engineering

by

Animesh Dubey (38110033)
Sumit Kumar Singh (38110572)



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI

CHENNAI-600119

APRIL 2022



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with "A" grade by NAAC

Jeppiar Nagar, Rajiv Gandhi Salai, Chennai - 600119

www.sathyabama.ac.in



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **Animesh Dubey (38110033)** and **Sumit Kumar Singh (38110572)** who carried out the project entitled "**AUTOMATIC RECOGNITION OF MEDICINAL PLANTS**" under my supervision from November 2021 to April 2022.

Internal Guide

Dr. G. Meera Gandhi M.E., Ph.D.,

Head of the Department

Dr.L.Lakshmanan.,M.E.,Ph.D.,

Dr.S.Vigneshwari.,M.E.,Ph.D.,

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

We, Animesh Dubey (38110033) and Sumit Kumar Singh (38110572), hereby declare that the Project Report entitled “**AUTOMATIC RECOGNITION OF MEDICINAL PLANTS**” done by me under the guidance of **Dr. G. Meera Gandhi M.E., Ph.D.**, is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE:

PLACE: CHENNAI

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D, Dean**, School of Computing , **Dr.S.Vigneshwari M.E., Ph.D.** and **Dr. L .Lakshmanan M.E., Ph.D.** , Heads of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. G. Meera Gandhi M.E., Ph.D.**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

A fully robotic strategy for recognizing therapeutic plants using PC vision and artificial intelligence methods has been demonstrated. The leaves of 25 distinctive therapeutic plant species were collected and captured using an advanced cell in a research environment. Countless highlights have been separated from each sheet, for example length, width, border and area, number of vertices, shading, border and frame territory. Then some highlights determined by these properties have been elaborated. The best results were obtained by a CNN classifier that uses a 10 times cross approval strategy. With an accuracy of 98.3%, the CNN classifier performed better than any other AI approach, such as the nearest k- neighbor, Naïve Bayes, KNN and neural systems. These results are enriching and future work will be equipped to use a larger data set and elite processing offices to explore the exposure of deep learning neural systems to recognize restoration plants used in essential medicinal services. As far as we know, this work is the first of its kind that has created a unique image data set for therapeutic plants accessible on the island of Mauritius. It is expected that an electronic or portable PC framework for the programmed recognition of restoration plants will help the population of the neighborhood to improve their knowledge of therapeutic plants, to help taxonomists develop recognizable test methods for increasingly competent species.

TABLE OF CONTENTS

CHAPTER No.	TITLE	PAGE No.
	ABSTRACT	V
	LIST OF FIGURES	Viii
	LIST OF ABBREVIATIONS	X
1	INTRODUCTION	1
	1.1 OVERVIEW OF THE PROJECT	1
	1.2 EXSISTING SYSTEM	3
	1.3 PROPOSED SYSTEM	4
2	LITRATURE SURVEY	5
3	METHODOLOGY	9
	3.1 AIM OF THE PROJECT	9
	3.2 SYSTEM REQUIREMENTS	9
	3.2.1 SOFTWARE REQUIREMENTS	9
	3.2.2 HARDWARE REQUIREMENTS	9
	3.3 OVERVIEW OF THE PLATFORM	9
	3.3.1 PYTHON	9
	3.3.2 CNN INTRODUCTION	13
	3.3.3 CNN ALGORITHM	15
	3.4 APPLICATIONS OF CNN	22
	3.4 SYSTEM ARCHITECTURE	26

4	MODULE DESCRIPTION	28
	SYSTEM STUDY	28
	MODULE DESCRIPTION	30
	PREPROCESSING	30
	FEATURE EXTRACTION	30
	FEATURE SELECTION	30
	PREDICTION	31
5	RESULTS AND DISCUSSIONS	32
	USESOF MEDICINAL PLANTS	32
6	CONCLUSION AND FUTURE WORK	34
	REFERENCES	35
	APPENDIX	38
	A SOURCE CODE	38
	B SCREEN SHOTS	42
	C PAPER PUBLISH	44

LIST OF FIGURES

FIGURE	FIGURE NAME	PAGE
No.		No.
3.1	STEPS IN CNN	16
3.2	FIRST NEURON LAYER	17
3.3	ORIGINAL IMAGE	18
3.4	VISUALIZATION OF DETECTOR FIELD	18
3.5	VISUALIZATION OF RESPECTIVE FIELD	19
3.6	PIXEL REPRESENTATION OF FILTER	19
3.7	MAX POOLING	22
3.8	OUTLINE OF THE PROPOSED SYSTEM	27
5.1	TRAINING ACCURACY Vs VALIDATION ACCURACY	32

LIST OF ABBREVIATIONS

CNN	CONVOLUTIONAL NEURAL NETWORK.
FCH	FUZZY LOCAL HISTOGRAM.
FLBP	FUZZY LOCAL BINARY PATTERN.
GLCM	GRAY LEVEL CO-OCCURENCE MATRIX.
PNN	PROBABLISTIC NEURAL NETWORK.
RGB	RED, GREEN, BLUE.
SVM	SUPPORT VECTOR MACHINE.

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW OF THE PROJECT

Plant ID has a significant job in current logical issues, for example, biodiversity, environment, and pharmacology among others. In Biology, plant distinguishing proof includes breaking down numerous organs, for example, blossoms, seeds, leaves and woody parts. This methodology renders the errand troublesome as blossoms and seeds, which are occasional and subject to the plant's age and condition, are difficult to discover. In exceptional circumstances, for example, discovering fossils or uncommon plants, the material accessible to distinguish a plant is only the leaves. To comprehend these circumstances, a leaf morphological scientific categorization method is proposed, which considers just the leaves to play out the distinguishing proof errands [14]. This methodology joins different highlights of leaves, for example, shape, vein structure, surface, and some histological data. Recognizing plants is a troublesome and complex undertaking because of the idea of the leaves. Although that the leaves present some major highlights, they additionally present a wide example of the variety. This variety may happen in various leaves from a similar plant, where attributes, for example, development and presentation to the sun produce varieties in the size, shading, surface, and state of the leaves. These varieties are additionally present in leaves from similar species, however from various plants. Right now, are an outcome of soil impact, atmosphere or even condition when the leaf is being shaped.

Plants have been utilized as medicines for thousands of years in different countries and are a source of many potent and powerful drugs worldwide, a total of more than 35,000 plant species are used for medicinal purposes. The value of medicinal plants to human livelihoods is essentially infinite. The World Health Organization estimated that 80% of the population of de-veloping countries relies on traditional medicines, mostly plant drugs, for their primary healthcare needs. Since time immemorial man has used various parts of plants in the treatment and prevention of many ailments. From prehistoric days, plants are used for shelter, food and medicine. The use of plants for medicinal purposes is as old as our

civilization. The first known written record of curative plants was of Sumerian herbal of 2200 BC. In the 5th century BC, The Greek doctor Hippocrates list out some 400 herbs in common use (Lakshmi V et al., 2006). Dioscorides, in the 1st century AD, wrote an herbal by using 600 plants which ultimately became the base for many later works. The World Health Organization estimated that 80% of the population of developing countries religion traditional medicines, mostly plant drugs, for their primary health care needs. Time immemorial man has used various parts of plants in the treatment and prevention of many ailments

The frames have grown so far using programmed order procedures, however the procedures are very comparative. These means include the configuration of the collected leaves, the realization of some previous managements to distinguish their particular characteristics, the arrangement of the leaves, the compilation of the database, the preparation for the recognition and evaluation of the results. The world has a more number of plant species, a significant amount of which have remedial qualities, close to elimination and others are destructive to humans. The recognition of dark plants depends very much on the intrinsic information of a specialized botanist. The best technique for distinguishing plants effectively and effectively is a manual processing strategy based on morphological qualities. In this way, a large number of procedures relating to the organization of these plant species "depend on the accumulation of information and the skills of individuals" [1]. Be that as it may, this procedure of manual acknowledgment is regularly relentless and time consuming. Henceforth numerous scientists have directed examinations to help the programmed grouping of plants dependent on their physical qualities .

The frames have grown so far using programmed order procedures, however the procedures are very comparative. These means include the configuration of the collected leaves, the realization of some previous managements to distinguish their particular characteristics, the arrangement of the leaves, the compilation of the database, the preparation for the recognition and evaluation of the results. A computerized plant distinguishing proof framework can be utilized by non-botanical specialists to rapidly recognize plant species easily.

Normally customary individuals are doled out with the activity of gathering the plants from the timberlands. Once in a while they couldn't perceive the uncommon and significant plants due to human blunder. These uncommon sorts

of plants are critical to spare the life of a patient. Additionally, here and there these individuals could get off base species which might be hurtful plants. In such cases, it is important to utilize the programmed plant acknowledgment framework. This framework helps a conventional people or any layman to perceive the diverse plant species. These sorts of frameworks are likewise extremely accommodating for the trekking individuals if they are intrigued to gather the plant species while trekking the mountains.

EXISTING METHOD

This research proposed a new mobile application based on Android operating system for identifying Indonesian medicinal plant images based on texture and color features of digital leaf images. In the experiments we used 51 species of Indonesian medicinal plants and each species consists of 48 images, so the total images used in this research are 2,448 images. This research investigates effectiveness of the fusion between the Fuzzy Local Binary Pattern (FLBP) and the Fuzzy Color Histogram (FCH) in order to identify medicinal plants. The FLBP method is used for extracting leaf image texture. The FCH method is used for extracting leaf image color. The fusion of FLBP and FCH is done by using Product Decision Rules (PDR) method. This research used Probabilistic Neural Network (PNN) classifier for classifying medicinal plant species. The experimental results show that the fusion between FLBP and FCH can improve the average accuracy of medicinal plants identification. The accuracy of identification using fusion of FLBP and FCH is 74.51%. This application is very important to help people identifying and finding information about Indonesian medicinal plant. Historically all medicinal preparations were derived from plants, whether in the simple form of plant parts or in the more complex form of crude extracts, mixtures, etc. Today a substantial number of drugs are developed from plants (Fabricant and Farnsworth, 2001) which are active against a number of diseases. The majority of these involve the isolation of the active ingredients found in a particular medicinal plant and its subsequent modification. In the developed countries, 25 percent of the medical drugs are based on plants and their derivatives and the use of medicinal plants is well known among the indigenous people in rural areas of many developing countries

Disadvantages

- The most of existing methods has ignored the poor quality images like images with noise or poor brightness.
- Less accuracy

PROPOSED MEHOD

The proposed technique was tested on a dataset of 55 medicinal plants from Vietnam and a very high accuracy of 98.3% was obtained with a CNN classifier. The size of each image was 256*256 pixels. Proposed an approach based on fractal dimension features based on leaf shape and vein patterns for the recognition and classification plant leaves. Using a volumetric fractal dimension approach to generate a texture signature for a leaf and the GLCM (Gray level co occurrence matrix) algorithm.

Advantages

- High accuracy is obtained and time consumption for detecting the shape.
- More datasets are included.

CHAPTER 2

LITERATURE SURVEY

Some tests have been performed to create tools for distinctive tests on plants in the past 10 years. Wu et al. They completed one of the most definitive works in the field of plant disposal. [2] From five fundamental geometric aspects, twelve morphological aspects are deduced and, therefore, the analysis of the main components (PCA) is used to reduce the measurement in order to send fewer sources of information to a probabilistic neural system (PNN). They achieved normal accuracy of 90.3% with the Flavia dataset, which is its creation. Using an alternative dataset but a similar classifier, Hossain et al [4] have achieved a degree of accuracy comparable to the comparative highlights [4].

However, using comparable evidence, an alternative dataset with only 20 species, Du et al [5] reached 93% with the closest k classifier [5]. Using another separation measure called "ISO map", we obtain 92.3% accuracy in a dataset of 2000 images containing 20 unique types of sheets [6]. Herdiyenet et al. They used -9a combination of an example of an almost spongy couplet and a spongy shading histogram and a probabilistic neural system classifier (PNN) in a data set of 2448 leaf images (270 * 240 pixels) acquired from therapeutic plants in the forest Indonesia will achieve a grouping accuracy of 74.5% [7]. Prasvita et al [7] have created a versatile comparative application that depends on previous research [8]. Using the parts descriptor (KDES) as another component extraction strategy, Le et al [9] built a recognizable test framework for a fully mechanized implant. The proposed procedure was tested on a dataset of 55 therapeutic plants in Vietnam and a high accuracy of 98.3% was achieved with the help vector machine classifier (SVM). In addition, its calculation achieved 98.5% accuracy in the Flavia dataset, which is the best result distributed so far in this dataset. Using discrete wavelet modification to eliminate the interpretation of invariant reflections from a variety of 8 distinctive elaborate plants in Indonesia, Arai et al [10] achieved 95.8% accuracy using a classifier of help vector machines (SVM) [10].

The size of each image was 256 * 256 pixels. Du et al [11] proposed a methodology that depends on the measurement of the fractal highlighted on the basis of the shape of the leaf and the designs of the veins for the recognition and order of the leaves of the plant [11]. Using a closer k classifier with 20 highlights,

they had the opportunity to achieve a high recognition rate of 87.1%. Using a form of volumetric fractal measurement to try to create a surface mark for a sheet and calculating the linear discriminant analysis (LDA), Backes et al [12] had the opportunity to overcome the conventional methodologies that depended on the Gabor channels and the Fourier test achieved 87.3% accuracy in a data set of 640 leaves from 32 distinctive plant species [13]. They used to model and combine data as it were. The images were obtained using a camera for cell phones with a 1980 * 1024 lens. Hernández-Serna et al [14] achieved an accuracy level of 92.9% using the Flavia dataset [14]. Sixteen sources of information (6 geometric aspects, 8 surfaces and 2 morphological highlights) were sent to a counterfeit neuronal system (ANN) with 60 centers in the covered layer and a learning frequency of over 0.100 pages. Using the equivalent data set, Chaki et al [15] achieved an overall accuracy of 97.6% using a Neuro-Fuzzy (NFC) classifier with a 44-component surface vector and a component shape vector [15]. Use of the module only includes the Flavia and Pattern Net data set (a type of neural system). His neuronal advancement system had two hidden layers with 26 neurons each and was prepared for more than 100 pages.

Fascinating work has been done. They performed the calculation of the scale curvature histogram (HCoS) to extract the shape data and the modification of the almost corresponding example (LBPV) to eliminate the surface data. In the best case, the clean dataset exceeded the 7.3% dataset. This recommends that legitimately acquired images using a mobile phone can create good levels of proven accuracy and images that are physically prepared in a laboratory and then grouped together. Previously, Amin et al [18] used an appropriate neuronal level diagram (DHGN) to acquire fold data using 64 element vectors and the closest k classifier with Canberra separation to acquire 71.5% accuracy [18]. A study of the 25 best-selling pharmaceutical drugs in 1997 found that 11 of them (42%) were either biological, natural products or entities derived from natural products, with a total value of 17.5 billion US \$. The total sales value of drugs (such as Taxol) derived from just one plant species (*Taxus baccata*) was 2.3 billion US\$ in 2000. The world market for herbal remedies in 1999 was calculated to be worth 19.4 billion US\$, with Europe in the lead (US\$ 6.7 bil-lion), followed by Asia (US\$ 5.1 billion), North America (US\$ 4.0 billion), Japan (US\$ 2.2 bil-lion), and then the rest of the world (US\$ 1.4 billion) (Hamilton, 2004).

According to a survey by NCI, USA, 61% of the 877 small-molecules as new chemical entities introduced as drugs worldwide during 1981-2002 were inspired by natural products (Newman et al., 2000). Plant species still serve as a rich source of many novel biologically active compounds; very few plant species have been thoroughly investigated for their medicinal properties (Heinrich and Gibbons, 2001). Thus, there is a renewed interest in phytomedicine during last decade and now a day many medicinal plant species are being screened for pharmacological activities. The use of plants as therapeutic tools, especially those used to relieve Sub-Acute pathologies, have had a remarkable role in the popular medicine of different countries. Many indigenous drugs have been used by practitioners for the treatment of DM throughout the world (Bailey & Day, 1989). However, only a few have received scientific or medical scrutiny and the World Health Organization has recommended accordingly that those traditional plant treatments for diabetes warrant further evaluation (WHO, 1980). Several hundred plants are known to have antidiabetic properties and a large number of compounds from plant extracts have been reported to have beneficial effects for treatment of diabetes. But, the use of medicinal plants in modern medicine suffers from the fact that though hundreds of plants are used in the world to prevent or cure diseases, scientific evidence in terms of modern medicine is lacking in most cases. Thus, identification of potential antidiabetic agents using mechanism-based studies holds great promise for elucidating mechanisms and devising more specific and effective treatments for diabetes related diseases (Izzo & Ernst, 2001). The plant families, including the species (sp), most studied for their confirmed hypoglycemic effects include: Leguminosae (11 sp), Lamiaceae (7 sp), Liliaceae (8 sp), Cucurbitaceae (7 sp), Asteraceae (6 sp), Moraceae (6 sp), Rosaceae (6 sp), Euphorbiaceae (5 sp) and Araliaceae (5 sp). The most studied species are: *Citrullus colocynthis* (*Opuntia streptacantha* Lem. (Cactaceae), *Trigonella foenum greacum* L. (Leguminosae), *Momordica charantia* L. (Cucurbitaceae), *Ficus bengalensis* L. (Moraceae), *Polygala senega* L. (Polygalaceae), and *Gymnema sylvestre* R. (Asclepiadaceae). Many studies have confirmed the benefits of medicinal plants with hypoglycemic effects in the management of diabetes mellitus. The effects of these plants may delay the development of diabetic complications and correct the metabolic abnormalities. The native people of Canada use 16 plants for the treatment of diabetes from very ancient period. In

Mexico, there are 306 plants recorded as anti-diabetic plants and the local traditional herbal healers in south eastern Morocco (Tafilalet region), use 92 plants for the treatment of diabetes mellitus and their complications. In the Chinese traditional medical treatment of diabetes, there are hundreds of prescriptions to aim directly at different symptoms of diabetes, most of which come from plants. Parag Bhandarkar, Rizwan Ahmed et al decomposed the morphology of leaf edges using predefined structural elements and extracted a structural signature which quantifies the leaf shape feature. They used the root mean square error between the feature vectors of the input image and the image in the database for computing the identity. The database created by the authors consists of 40 leaf samples of 10 different species. They achieved an overall classification rate of 67.5%, which is independent of leaf size and orientation. The identification rate is comparatively low to be of use in practical implementations.

CHAPTER 3

METHODOLOGY

AIM OF THE PROJECT

The main aim of this project is to automatically recognize the medicinal plants. Using the base features which are extracted directly from the image, a number of derived features are calculated. Ratios are more suitable for comparison as they are independent of the actual size of the image in pixels.

SYSTEM REQUIREMENTS

Software Requirements

- Operating System :WINDOWS 7,8,10
- Software : PYTHON 3.7
- Simulation Tool :OPENCV PYTHON

Hardware Requirements

- CPU type :Intel Pentium 4
- Clock speed :3.0 GHz
- Ram size :512 MB
- CD -drive type :52xmax

OVERVIEW OF THE PLATFORM

Python

Python is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code.

Python is a programming language that lets you work quickly and integrate systems more efficiently.

It is used for:

- web development (server-side),
- software development,

- mathematics,
- System scripting.

What can Python do?

- Python can be used on a server to create web applications.
- Python can be used alongside software to create workflows.
- Python can connect to database systems. It can also read and modify files.
- Python can be used to handle big data and perform complex mathematics.
- Python can be used for rapid prototyping, or for production-ready software development.

Why Python?

- Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc).
- Python has a simple syntax similar to the English language.
- Python has syntax that allows developers to write programs with fewer lines than some other programming languages.
- Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.
- Python can be treated in a procedural way, an object-orientated way or a functional way.

Good to know

- The most recent major version of Python is Python 3, which we shall be using in this tutorial. However, Python 2, although not being updated with anything other than security updates, is still quite popular.
- Python 2.0 was released in 2000, and the 2.x versions were the prevalent releases until December 2008. At that time, the development team made the decision to release version 3.0, which contained a few relatively small but significant changes that were not backward compatible with the 2.x versions. Python 2 and 3 are very similar, and some features of Python 3

have been backported to Python 2. But in general, they remain not quite compatible.

- Both Python 2 and 3 have continued to be maintained and developed, with periodic release updates for both. As of this writing, the most recent versions available are 2.7.15 and 3.6.5. However, an official End of Life date of January 1, 2020 has been established for Python 2, after which time it will no longer be maintained.
- Python is still maintained by a core development team at the Institute, and Guido is still in charge, having been given the title of BDFL (Benevolent Dictator For Life) by the Python community. The name Python, by the way, derives not from the snake, but from the British comedy troupe Monty Python's Flying Circus, of which Guido was, and presumably still is, a fan. It is common to find references to Monty Python sketches and movies scattered throughout the Python documentation.
- It is possible to write Python in an Integrated Development Environment, such as Thonny, Pycharm, Netbeans or Eclipse which are particularly useful when managing larger collections of Python files.

Python Syntax compared to other programming languages

- Python was designed to for readability, and has some similarities to the English language with influence from mathematics.
- Python uses new lines to complete a command, as opposed to other programming languages which often use semicolons or parentheses.
- Python relies on indentation, using whitespace, to define scope; such as the scope of loops, functions and classes. Other programming languages often use curly-brackets for this purpose.

Python is Interpreted

- Many languages are compiled, meaning the source code you create needs to be translated into machine code, the language of your computer's processor, before it can be run. Programs written in an interpreted language are passed straight to an interpreter that runs them directly.

- This makes for a quicker development cycle because you just type in your code and run it, without the intermediate compilation step.
- One potential downside to interpreted languages is execution speed. Programs that are compiled into the native language of the computer processor tend to run more quickly than interpreted programs. For some applications that are particularly computationally intensive, like graphics processing or intense number crunching, this can be limiting.
- In practice, however, for most programs, the difference in execution speed is measured in milliseconds, or seconds at most, and not appreciably noticeable to a human user. The expediency of coding in an interpreted language is typically worth it for most applications.
- For all its syntactical simplicity, Python supports most constructs that would be expected in a very high-level language, including complex dynamic data types, structured and functional programming, and object-oriented programming.
- Additionally, a very extensive library of classes and functions is available that provides capability well beyond what is built into the language, such as database manipulation or GUI programming.
- Python accomplishes what many programming languages don't: the language itself is simply designed, but it is very versatile in terms of what you can accomplish with it.

Introduction to Machine Learning using Python

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data. In this article, we'll see basics of Machine Learning, and implementation of a simple machine learning algorithm using python.

Convolutional Neural Networks (CNN) Introduction

Convolutional neural networks (CNN) sounds like a weird combination of biology and math with a little CS sprinkled in, but these networks have been some of the most influential innovations in the field of computer vision. 2012 was the first year that neural nets grew to prominence as Alex Krizhevsky used them to win that year's ImageNet competition (basically, the annual Olympics of computer vision), dropping the classification error record from 26% to 15%, an astounding improvement at the time. Ever since then, a host of companies have been using deep learning at the core of their services. Facebook uses neural nets for their automatic tagging algorithms, Google for their photo search, Amazon for their product recommendations, Pinterest for their home feed personalization, and Instagram for their search infrastructure.

Inputs and Outputs

When a computer sees an image (takes an image as input), it will see an array of pixel values. Depending on the resolution and size of the image, it will see a $32 \times 32 \times 3$ array of numbers (The 3 refers to RGB values). Just to drive home the point, let's say we have a color image in JPG form and its size is 480×480 . The representative array will be $480 \times 480 \times 3$. Each of these numbers is given a value from 0 to 255 which describe the pixel intensity at that point. These numbers, while meaningless to us when we perform image classification, are the only inputs available to the computer. The idea is that you give the computer this array of numbers and it will output numbers that describe the probability of the image being a certain class (.80 for cat, .15 for dog, .05 for bird, etc).

What We Want the Computer to do

Now that we know the problem as well as the inputs and outputs, let's think about how to approach this. What we want the computer to do is to be able to differentiate between all the images it's given and figure out the unique features that make a dog a dog or that make a cat a cat. This is the process that goes on in our minds subconsciously as well. When we look at a picture of a

dog, we can classify it as such if the picture has identifiable features such as paws or 4 legs. In a similar way, the computer is able to perform image classification by looking for low level features such as edges and curves, and then building up to more abstract concepts through a series of convolutional layers. This is a general overview of what a CNN does. Let's get into the specifics.

Biological Connection

But first, a little background. When you first heard of the term convolutional neural networks, you may have thought of something related to neuroscience or biology, and you would be right. Sort of. CNNs do take a biological inspiration from the visual cortex. The visual cortex has small regions of cells that are sensitive to specific regions of the visual field. This idea was expanded upon by a fascinating experiment by Hubel and Wiesel in 1962 (Video) where they showed that some individual neuronal cells in the brain responded (or fired) only in the presence of edges of a certain orientation. For example, some neurons fired when exposed to vertical edges and some when shown horizontal or diagonal edges. Hubel and Wiesel found out that all of these neurons were organized in a columnar architecture and that together, they were able to produce visual perception. This idea of specialized components inside of a system having specific tasks (the neuronal cells in the visual cortex looking for specific characteristics) is one that machines use as well, and is the basis behind CNNs.

The Problem Space

Image classification is the task of taking an input image and outputting a class (a cat, dog, etc) or a probability of classes that best describes the image. For humans, this task of recognition is one of the first skills we learn from the moment we are born and is one that comes naturally and effortlessly as adults. Without even thinking twice, we're able to quickly and seamlessly identify the environment we are in as well as the objects that surround us. When we see an image or just when we look at the world around us, most of the time we are able to immediately characterize the scene and give each object a label, all without

even consciously noticing. These skills of being able to quickly recognize patterns, generalize from prior knowledge, and adapt to different image environments are ones that we do not share with our fellow machines.

Convolution Neural Network Algorithm

Convolution neural network algorithm is a multilayer perceptron that is the special design for identification of two-dimensional image information. Always has more layers: input layer, convolution layer, sample layer and output layer. In addition, in a deep network architecture the convolution layer and sample layer can have multiple. CNN is not as restricted boltzmann machine, need to be before and after the layer of neurons in the adjacent layer for all connections, convolution neural network algorithms, each neuron don't need to do feel global image, just feel the local area of the image. In addition, each neuron parameter is set to the same, namely, the sharing of weights, namely each neuron with the same convolution kernels to deconvolution image. Fully connection vs partial connection CNN algorithm has two main processes: convolution and sampling . Convolution process: use a trainable filter F_x , deconvolution of the input image (the first stage is the input image, the input of the after convolution is the feature image of each layer, namely Feature Map), then add a bias b_x , we can get convolution layer C_x . A sampling process: n pixels of each neighborhood through pooling steps, become a pixel, and then by scalar weighting $W_x + 1$ weighted, add bias $b_x + 1$, and then by an activation function, produce a narrow n times feature map $S_x + 1$. Main process of CNN The key technology of CNN is the local receptive field, sharing of weights , sub sampling by time or space, so as to extract feature and reduce the size of the training parameters. The advantage of CNN algorithm is that to avoid the explicit feature extraction, and implicitly to learn from the training data; The same neuron weights on the surface of the feature mapping, thus network can learn parallelly , reduce the complexity of the network; Adopting sub sampling structure by time or space, can achieve some degree of robustness, scale and deformation displacement; Input information and network topology can be a very good match, It has unique advantages in speech recognition and image processing

A convolutional neural network (CNN) is a deep friendly neural network designed from biologically controlled modules. They are applied most frequently used for image processing. CNN's work differently in that they treat data as spatial. Instead of being connected to all the neurons in the previous layer, the neurons are connected only to neighbouring neurons and all have the same weight. This simplification in connections means that the network maintains the spatial aspect of the data set. The word convolution refers to the filtering process that occurs in this type of network. If an image is complex, the convolutional neural network simplifies it, so it can be better elaborated and understood. Like a normal neural network, CNN is also composed of several levels. There are a couple of levels that make the convolutional level and the grouping level unique. However, like other neural networks, it will also have a rectified linear unit layer and a fully connected layer. The ReLU level acts as an activation function, ensuring non-linearity as the data moves through each level of the network.

Without it, the data entered in each level would lose the dimensionality that we want to maintain. Meanwhile, the fully connected layer allows you to classify the data set.

However, the classic, and arguably most popular, use case of these networks is for image processing. Within image processing, let's take a look at how to use these CNNs for image classification.

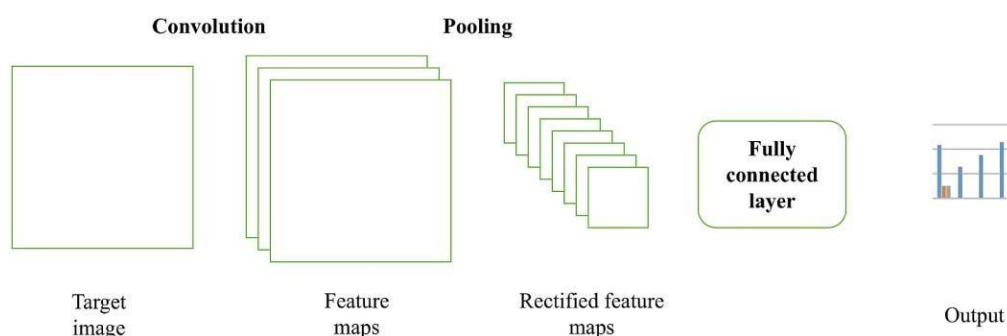


Fig 3.1: Steps in CNN

First Layer – Math Part

The first layer in a CNN is always a Convolutional Layer. First thing to make sure you remember is what the input to this conv (I'll be using that

abbreviation a lot) layer is. Like we mentioned before, the input is a $32 \times 32 \times 3$ array of pixel values. Now, the best way to explain a conv layer is to imagine a flashlight that is shining over the top left of the image. Let's say that the light this flashlight shines covers a 5×5 area. And now, let's imagine this flashlight sliding across all the areas of the input image. In machine learning terms, this flashlight is called a filter (or sometimes referred to as a neuron or a kernel) and the region that it is shining over is called the receptive field. Now this filter is also an array of numbers (the numbers are called weights or parameters). A very important note is that the depth of this filter has to be the same as the depth of the input (this makes sure that the math works out), so the dimension of this filter is $5 \times 5 \times 3$. Now, let's take the first position the filter is in for example. It would be the top left corner. As the filter is sliding, or convolving, around the input image, it is multiplying the values in the filter with the original pixel values of the image (aka computing element wise multiplications). These multiplications are all summed up (mathematically speaking, this would be 75 multiplications in total). So now you have a single number. Remember, this number is just representative of when the filter is at the top left of the image. Now, we repeat this process for every location on the input volume. (Next step would be moving the filter to the right by 1 unit, then right again by 1, and so on). Every unique location on the input volume produces a number. After sliding the filter over all the locations, you will find out that what you're left with is a $28 \times 28 \times 1$ array of numbers, which we call an activation map or feature map. The reason you get a 28×28 array is that there are 784 different locations that a 5×5 filter can fit on a 32×32 input image. These 784 numbers are mapped to a 28×28 array.

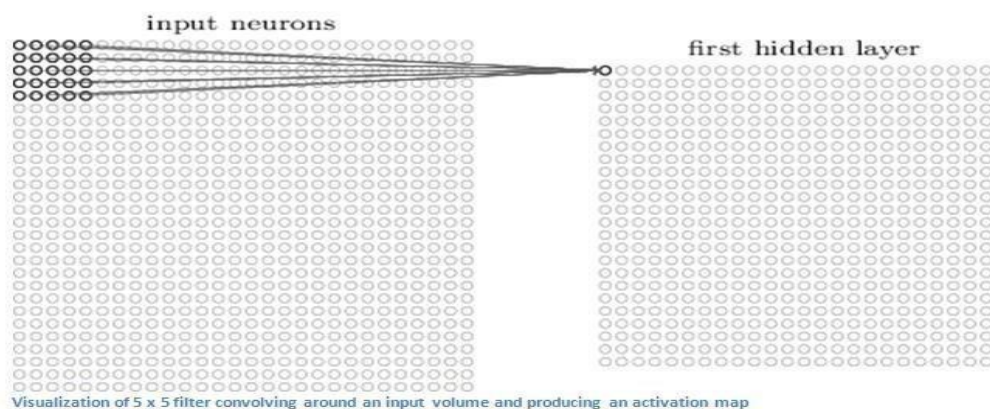


Fig 3.2: First Neuron Layer

Let's say now we use two $5 \times 5 \times 3$ filters instead of one. Then our output volume would be $28 \times 28 \times 2$. By using more filters, we are able to preserve the spatial dimensions better. Mathematically, this is what's going on in a convolutional layer.

First Layer – High Level Perspective

However, let's talk about what this convolution is actually doing from a high level. Each of these filters can be thought of as **feature identifiers**. When I say features, I'm talking about things like straight edges, simple colors, and curves. Think about the simplest characteristics that all images have in common with each other. Let's say our first filter is $7 \times 7 \times 3$ and is going to be a curve detector. (In this section, let's ignore the fact that the filter is 3 units deep and only consider the top depth slice of the filter and the image, for simplicity.) As a curve detector, the filter will have a pixel structure in which there will be higher numerical values along the area that is a shape of a curve (Remember, these filters that we're talking about as just numbers!).

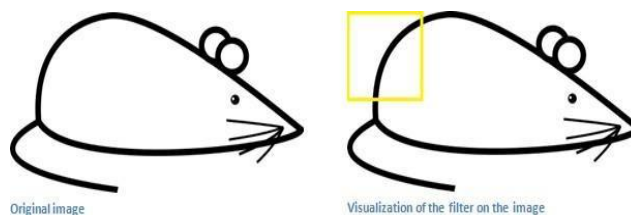


Fig 3.3: Original Image

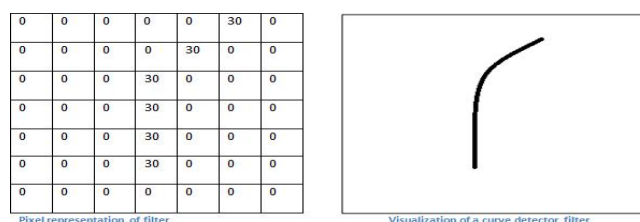


Fig 3.4: Visualization of a curve detector filter

Now, let's go back to visualizing this mathematically. When we have this filter at the top left corner of the input volume, it is computing multiplications between the filter and pixel values at that region. Now let's take an example of an image that we want to classify, and let's put our filter at the top left corner.

Remember, what we have to do is multiply the values in the filter with the original pixel values of the image.

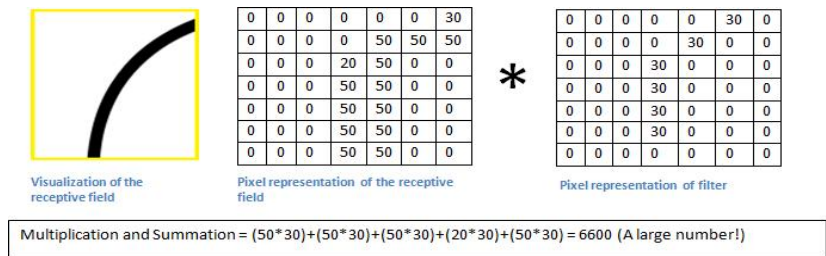


Fig 3.5: Visualization of the receptive field

Basically, in the input image, if there is a shape that generally resembles the curve that this filter is representing, then all of the multiplications summed together will result in a large value! Now let's see what happens when we move our filter.

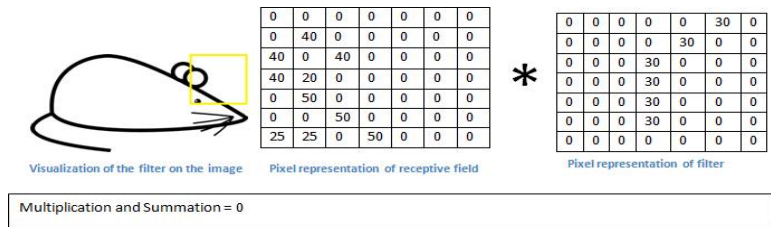


Fig 3.6: Pixel Representation of Filter

The value is much lower! This is because there wasn't anything in the image section that responded to the curve detector filter. Remember, the output of this conv layer is an activation map. So, in the simple case of a one filter convolution (and if that filter is a curve detector), the activation map will show the areas in which there at mostly likely to be curves in the picture. In this example, the top left value of our 26 x 26 x 1 activation map (26 because of the 7x7 filter instead of 5x5) will be 6600. This high value means that it is likely that there is some sort of curve in the input volume that caused the filter to activate. The top right value in our activation map will be 0 because there wasn't anything in the input volume that caused the filter to activate (or more simply said, there wasn't a curve in that region of the original image). Remember, this is just for one filter. This is just a

filter that is going to detect lines that curve outward and to the right. We can have other filters for lines that curve to the left or for straight edges. The more filters, the greater the depth of the activation map, and the more information we have about the input volume.

Going Deeper Through the Network

Now in traditional convolutional neural network architecture, there are other layers that are interspersed between these conv layers. I'd strongly encourage those interested to read up on them and understand their function and effects, but in a general sense, they provide nonlinearities and preservation of dimension that help to improve the robustness of the network and control over fitting. A classic CNN architecture would look like this.

Input -> Conv -> ReLU -> Conv -> ReLU -> Pool -> ReLU -> Conv -> ReLU -> Pool -> Fully Connected

The last layer, however, is an important one and one that we will go into later on. Let's just take a step back and review what we've learned so far. We talked about what the filters in the first conv layer are designed to detect. They detect low level features such as edges and curves. As one would imagine, in order predicting whether an image is a type of object, we need the network to be able to recognize higher level features such as hands or paws or ears. So let's think about what the output of the network is after the first conv layer. It would be a $28 \times 28 \times 3$ volume (assuming we use three $5 \times 5 \times 3$ filters). When we go through another conv layer, the output of the first conv layer becomes the input of the 2nd conv layer. Now, this is a little bit harder to visualize. When we were talking about the first layer, the input was just the original image. However, when we're talking about the 2nd conv layer, the input is the activation map(s) that result from the first layer. So each layer of the input is basically describing the locations in the original image for where certain low level features appear. Now when you apply a set of filters on top of that (pass it through the 2nd conv layer), the output will be activations that represent higher level features. Types of these features could be semicircles (combination of a curve and straight edge) or squares (combination of several straight edges). As you go through the network and go through more conv

layers, you get activation maps that represent more and more complex features. By the end of the network, you may have some filters that activate when there is handwriting in the image, filters that activate when they see pink objects, etc. If you want more information about visualizing filters in ConvNets, Matt Zeiler and Rob Fergus had an excellent research paper discussing the topic. Jason Yosinski also has a video on YouTube that provides a great visual representation. Another interesting thing to note is that as you go deeper into the network, the filters begin to have a larger and larger receptive field, which means that they are able to consider information from a larger area of the original input volume (another way of putting it is that they are more responsive to a larger region of pixel space).

The convolution layer computes the output of neurons that are connected to local regions or receptive fields in the input, each computing a dot product between their weights and a small receptive field to which they are connected to in the input volume. Each computation leads to extraction of a feature map from the input image. In other words, imagine you have an image represented as a 5x5 matrix of values, and you take a 3x3 matrix and slide that 3x3 window or kernel around the image. At each position of that matrix, you multiply the values of your 3x3 window by the values in the image that are currently being covered by the window. As a result, you'll get a single number that represents all the values in that window of the images. You use this layer to filtering: as the window moves over the image, you check for patterns in that section of the image. This works because of filters, which are multiplied by the values outputted by the convolution.

The objective of subsampling is to get an input representation by reducing its dimensions, which helps in reducing overfitting. One of the techniques of subsampling is max pooling. With this technique, you select the highest pixel value from a region depending on its size. In other words, max pooling takes the largest value from the window of the image currently covered by the kernel. For example, you can have a max-pooling layer of size 2 x 2 will select the maximum pixel intensity value from 2 x 2 region. You're right to think that the pooling layer then works a lot like the convolution layer! You also take a kernel or a window and move it over the image; The only difference is the function that is applied to the kernel and the image window isn't linear.

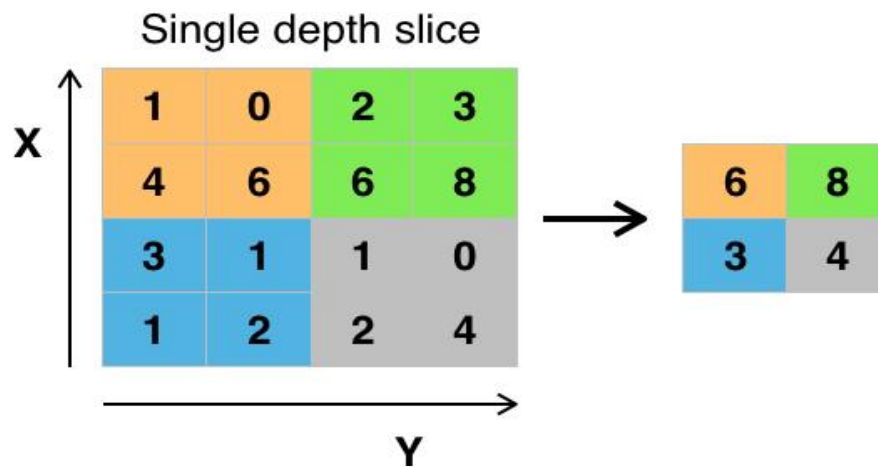


Fig 3.7: Max Pooling

The objective of the fully connected layer is to flatten the high-level features that are learned by convolutional layers and combining all the features. It passes the flattened output to the output layer where you use a softmax classifier or a sigmoid to predict the input class label.

How Companies Use CNNs

Data, data, data. The companies that have lots of this magic 4 letter word are the ones that have an inherent advantage over the rest of the competition. The more training data that you can give to a network, the more training iterations you can make, the more weight updates you can make, and the better tuned to the network is when it goes to production. Facebook (and Instagram) can use all the photos of the billion users it currently has, Pinterest can use information of the 50 billion pins that are on its site, Google can use search data, and Amazon can use data from the millions of products that are bought every day. And now you know the magic behind how they use it.

APPLICATIONS OF CNN

Image recognition

CNNs are often used in image recognition systems. In 2012 an error rate of 0.23 percent on the MNIST database was reported. Another paper on using CNN for image classification reported that the learning process was "surprisingly fast"; in the same paper, the best published results as of 2011 were achieved in the

MNIST database and the NORB database. Subsequently, a similar CNN called AlexNet[74] won the ImageNet Large Scale Visual Recognition Challenge 2012.

When applied to facial recognition, CNNs achieved a large decrease in error rate. Another paper reported a 97.6 percent recognition rate on "5,600 still images of more than 10 subjects". CNNs were used to assess video quality in an objective way after manual training; the resulting system had a very low root mean square error.

The ImageNet Large Scale Visual Recognition Challenge is a benchmark in object classification and detection, with millions of images and hundreds of object classes. In the ILSVRC 2014,[76] a large-scale visual recognition challenge, almost every highly ranked team used CNN as their basic framework. The winner GoogLeNet(the foundation of DeepDream) increased the mean average precision of object detection to 0.439329, and reduced classification error to 0.06656, the best result to date. Its network applied more than 30 layers. That performance of convolutional neural networks on the ImageNet tests was close to that of humans. The best algorithms still struggle with objects that are small or thin, such as a small ant on a stem of a flower or a person holding a quill in their hand. They also have trouble with images that have been distorted with filters, an increasingly common phenomenon with modern digital cameras. By contrast, those kinds of images rarely trouble humans. Humans, however, tend to have trouble with other issues. For example, they are not good at classifying objects into fine-grained categories such as the particular breed of dog or species of bird, whereas convolutional neural networks handle this.[citation needed]

In 2015 a many-layered CNN demonstrated the ability to spot faces from a wide range of angles, including upside down, even when partially occluded, with competitive performance. The network was trained on a database of 200,000 images that included faces at various angles and orientations and a further 20 million images without faces. They used batches of 128 images over 50,000 iterations.

Video analysis

Compared to image data domains, there is relatively little work on applying CNNs to video classification. Video is more complex than images since it has another (temporal) dimension. However, some extensions of CNNs into the video

domain have been explored. One approach is to treat space and time as equivalent dimensions of the input and perform convolutions in both time and space. Another way is to fuse the features of two convolutional neural networks, one for the spatial and one for the temporal stream. Long short-term memory (LSTM) recurrent units are typically incorporated after the CNN to account for inter-frame or inter-clip dependencies. Unsupervised learning schemes for training spatio-temporal features have been introduced, based on Convolutional Gated Restricted Boltzmann Machines and Independent Subspace Analysis.

Natural language processing

CNNs have also been explored for natural language processing. CNN models are effective for various NLP problems and achieved excellent results in semantic parsing, search query retrieval, sentence modeling, classification, prediction and other traditional NLP tasks.

Anomaly Detection

A CNN with 1-D convolutions was used on time series in the frequency domain (spectral residual) by an unsupervised model to detect anomalies in the time domain.

Drug discovery

CNNs have been used in drug discovery. Predicting the interaction between molecules and biological proteins can identify potential treatments. In 2015, Atomwise introduced AtomNet, the first deep learning neural network for structure-based rational drug design. The system trains directly on 3-dimensional representations of chemical interactions. Similar to how image recognition networks learn to compose smaller, spatially proximate features into larger, complex structures, AtomNet discovers chemical features, such as aromaticity, sp³ carbons and hydrogen bonding. Subsequently, AtomNet was used to predict novel candidate biomolecules for multiple disease targets, most notably treatments for the Ebola virus and multiple sclerosis.

Health risk assessment and biomarkers of aging discovery

CNNs can be naturally tailored to analyze a sufficiently large collection of time series data representing one-week-long human physical activity streams augmented by the rich clinical data (including the death register, as provided by, e.g., the NHANES study). A simple CNN was combined with Cox-Gompertz proportional hazards model and used to produce a proof-of-concept example of digital biomarkers of aging in the form of all-causes-mortality predictor.

Checkers game

CNNs have been used in the game of checkers. From 1999 to 2001, Fogel and Chellapilla published papers showing how a convolutional neural network could learn to play checker using co-evolution. The learning process did not use prior human professional games, but rather focused on a minimal set of information contained in the checkerboard: the location and type of pieces, and the difference in number of pieces between the two sides. Ultimately, the program (Blondie24) was tested on 165 games against players and ranked in the highest 0.4%. It also earned a win against the program Chinook at its "expert" level of play.

Go

CNNs have been used in computer Go. In December 2014, Clark and Storkey published a paper showing that a CNN trained by supervised learning from a database of human professional games could outperform GNU Go and win some games against Monte Carlo tree search Fuego in a fraction of the time it took Fuego to play. Later it was announced that a large 12-layer convolutional neural network had correctly predicted the professional move in 55% of positions, equalling the accuracy of a 6 dan human player. When the trained convolutional network was used directly to play games of Go, without any search, it beat the traditional search program GNU Go in 97% of games, and matched the performance of the Monte Carlo tree search program Fuego simulating ten thousand playouts (about a million positions) per move.

A couple of CNNs for choosing moves to try ("policy network") and evaluating positions ("value network") driving MCTS were used by AlphaGo, the first to beat the best human player at the time.

Time series forecasting

Recurrent neural networks are generally considered the best neural network architectures for time series forecasting (and sequence modeling in general), but recent studies show that convolutional networks can perform comparably or even better. Dilated convolutions might enable one-dimensional convolutional neural networks to effectively learn time series dependences. Convolutions can be implemented more efficiently than RNN-based solutions, and they do not suffer from vanishing (or exploding) gradients. Convolutional networks can provide an improved forecasting performance when there are multiple similar time series to learn from. CNNs can also be applied to further tasks in time series analysis.

Fine-tuning

For many applications, the training data is less available. Convolutional neural networks usually require a large amount of training data in order to avoid overfitting. A common technique is to train the network on a larger data set from a related domain. Once the network parameters have converged an additional training step is performed using the in-domain data to fine-tune the network weights. This allows convolutional networks to be successfully applied to problems with small training sets.

Human interpretable explanations

End-to-end training and prediction are common practice in computer vision. However, human interpretable explanations are required for critical systems such as a self-driving cars. With recent advances in visual salience, spatial and temporal attention, the most critical spatial regions/temporal instants could be visualized to justify the CNN predictions.

3.5 SYSTEM ARCHITECTURE

The initial step is to delete the RGB channel data from the information image to distinguish the implant data from the anterior area. To do this, data is extracted from the excess green (ExG) and excess red (Exp) channels.

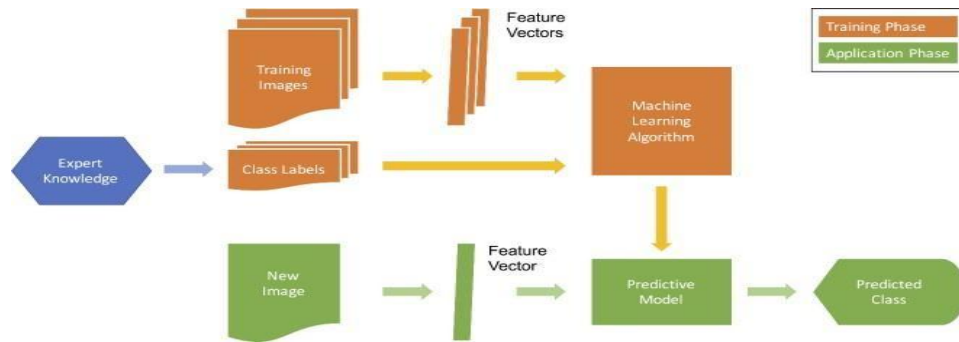


Fig 3.8: Outline of the Proposed System

Then, the red overabundance data is removed from the abundance of green data to progressively contain the green channel data. When the double limit is applied in the improved vegetative archive, the ExG ExR data, a double intrigue district is divided into plants. The next advance is to use the first double image shadow pixel as a cover that separates the leaves as secondary images. The division of the test sheet images into the dataset using the vegetative record mentioned above. The important component in the representation of an image is shading. The shading minutes are obtained through the highlighted facts, for example the average and the standard deviation evaluated in the three RGB channels (red, green, blue). The main measures for requesting an image are acquired from the standard deviation and the average.

The surface of the image can be estimated using the concurrency pane, the different shading elements or the grayscale estimates of the image. The highlights that are created using this strategy are known as Haralick highlights. A calculated regression classifier is used to distinguish the restoration plan. Since our proposed strategy consists of the characterization of numerous classes, the relapse model calculated in parallel has been extended to the calculated multinomial relapse. Through a combination of relapses calculated in pairs, the multinomial logit thinks of different encounters. This allows you to compare each brand class with a reference classification in the data set.

CHAPTER 4

MODULE DESCRIPTION

SYSTEM STUDY

FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- Economical Feasibility
- Technical Feasibility
- Social Feasibility
- Legal Feasibility
- Operational Feasibility
- Scheduling Feasibility

Economical Feasibility

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

Technical Feasibility

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the

client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

Social Feasibility

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

Legal Feasibility

This assessment investigates whether any aspect of the proposed project conflicts with legal requirements like zoning laws, data protection acts or social media laws. Let's say an organization wants to construct a new office building in a specific location. A feasibility study might reveal the organization's ideal location isn't zoned for that type of business. That organization has just saved considerable time and effort by learning that their project was not feasible right from the beginning.

Operational Feasibility

This assessment involves undertaking a study to analyze and determine whether—and how well—the organization's needs can be met by completing the project. Operational feasibility studies also examine how a project plan satisfies the requirements identified in the requirements analysis phase of system development.

Scheduling Feasibility

This assessment is the most important for project success; after all, a project will fail if not completed on time. In scheduling feasibility, an organization estimates how much time the project will take to complete.

When these areas have all been examined, the feasibility analysis helps identify any constraints the proposed project may face, including:

- Internal Project Constraints: Technical, Technology, Budget, Resource, etc.
- Internal Corporate Constraints: Financial, Marketing, Export, etc.

- External Constraints: Logistics, Environment, Laws, and Regulations.

MODULE DESCRIPTION

Preprocessing

Data pre-processing is an important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values (e.g., Income: -100), impossible data combinations missing values, etc. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis. Often, data pre-processing is the most important phase of a machine learning project, especially in computational biology. The product of data pre-processing is the final training set.

Feature Extraction

Feature extraction involves reducing the number of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power, also it may cause a classification algorithm to over fit to training samples and generalize poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Many machine learning practitioners believe that properly optimized feature extraction is the key to effective model construction.

Feature Selection

Feature selection, also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features (variables, predictors) for use in model construction. Feature selection techniques are used for several reasons:

- Simplification of models to make them easier to interpret by researchers/users.

- Shorter training times.
- To avoid the curse of dimensionality.
- Enhanced generalization by reducing over fitting by reducing over fitting.

Prediction

Predictive analytics is driven by predictive modelling. It's more of an approach than a process. Predictive analytics and machine learning go hand-in-hand, as predictive models typically include a machine learning algorithm. These models can be trained over time to respond to new data or values, delivering the results the business needs. Predictive modelling largely overlaps with the field of machine learning.

Patterned after the operation of neurons in the human brain, neural networks (also called artificial neural networks) are a variety of deep learning technologies. They're typically used to solve complex pattern recognition problems – and are incredibly useful for analysing large data sets. They are great at handling nonlinear relationships in data – and work well when certain variables are unknown. This is the last step of our project. Here the we can identify the image weather it is medicinal or non-medicinal.

CHAPTER 5

RESULTS AND DISCUSSIONS

The obtained results of training accuracy and validation accuracy values are shown in following graph.

In this diagram for where there has less number of epochs (upto 13) we can know that validation accuracy is same as the training accuracy. In case if we need more number of epochs the training accuracy and validation accuracy values are almost comparable.

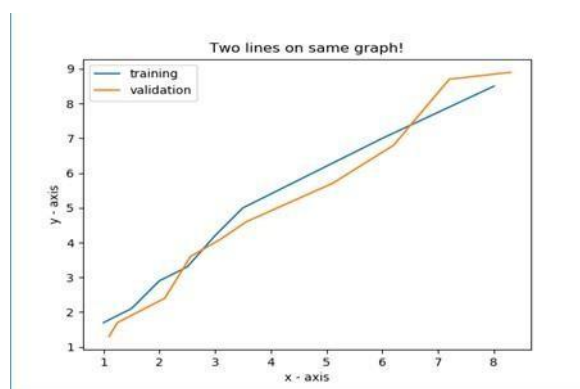


Fig 5.1: Training accuracy vs validation accuracy graph

5.1 USES OF MEDICINAL PLANTS

India is a veritable emporium of medicinal plants, because of varied physiography, soil and climatic conditions. The number of medicinal plants in India, both indigenous and exotic, has been variously put at between 3, 000 to 3,500 species. Sixteen medicinal plants of exotic origin, introduced in India years back, are now considered as a part of India's medicinal plants resources. Notable among these are, Senna, Psyllium, Belladonna, Cinchona, Eucalyptus, Ipecac, Digitalis and Mexican Dioscorea. The number of plants having confirmed therapeutic properties or yielding a useful chemical compound thus lays around 700 species. Out of these the plants providing largely or regularly used raw materials by Indian Drug and Pharmaceutical Industry are about 335, including those whose materials are imported from other countries. Some of which are liquorice, henbane, cassia bark, galangal, ephedra, long pepper and star anise are

used in large quantities. India ranks foremost after South Korea in the supply of medicinal plants to the industrialized countries of the west, where demand for natural drug has been on the increase in recent years. Nearly three fourth of the drugs mentioned in the various pharmacopoeia are found in their natural state in India, since the climate and geographical conditions are congenial for commercial growing of indigenous and exotic medicinal plants. Globally, about 85 percent of the traditional medicines used in primary healthcare are derived from plants. In today's scenario there is lot of increasing pressure in health care costs. Government is encouraging the use of indigenous system of traditional medicine rather than expensive synthetic drugs. It is reported that around 80% of the world's population have still trust on traditional system of medicine. Herbal medicines include herbs, herbal preparations and finished herbal products that contain active principles of plant parts, or other plant materials, or combinations. Traditional use of herbal medicines has a long historical use of these medicines. Their use is well established and widely acknowledged to be safe and effective and is globally accepted. The chemical composition present in them is a part of the physiological function of living flora and hence they are believed to have better compatibility with the human body. The herbal health care have therefore, been derived from rich traditions or ancient civilizations and scientific heritage which came to the present generations through primarily oral traditions. Medicinal plant parts should be authentic and free from pesticides, heavy materials, microbial or radioactive contamination, etc. The extract should be got checked for indicated biological activity on animals. The bioactive extract should be standardized on the basis of active constituents or major compounds. The raw herbal material used for value added products must be used for:

- * Isolation of active principles for formulation into drugs.
- * Isolation of Intermediates for the production of semi-synthetic drugs.
- * Preparation of standardized galenicals (extracts, powders, tinctures, etc).

India is a gold mine of herbal medicines with well-recorded and well practiced knowledge of traditional system of herbal medicines. But India has not been able to capitalize on this herbal wealth like China by promoting its use in the developed world despite their renewed interest in herbal medicines. This can be achieved by judicious product identification based on diseases found in the developed world for which no medicine or therapy is available.

CHAPTER 6

CONCLUSION AND FUTURE WORK

In this work, we addressed the problem of identifying the medicinal plant species by the analysis of leaf images obtained directly from their habitat and irrespective of lighting conditions. The fixed zero threshold, vegetative index is successfully tested for image dataset. The result shows that the algorithm can adequately segment the leaf region. This method worked well in images with reflection. The feature extraction based on the color and texture features is done. The classification of medicinal plant species is done by using Weka and the accuracy of 98.3% is measured. In future we have planned to design and develop a system which automatically identifies plant species through the analysis of not only the leaf images also the other parts of the plant acquired directly in their habitat irrespective of complex backgrounds and various lighting condition. This excellent performance indicates the viability of such computer-aided approaches in the classification of biological specimens and its potential applicability in combatting the 'taxonomic crists'. A web-based or mobile computer system for automatic recognition of medicinal plants will help the local population to improve their knowledge on medicinal plants, help taxonomists to develop more efficient species identification techniques and also contribute significantly in the protection of endangered species. It is necessary to study about medicinal plants that find place in folklore. As a result, such plants should be investigated to understand their physical and chemical properties as well as safety and efficacy. The medicinal plants find applications in many industries. The use of medicinal plants for curing diseases has been increased. In order to identify the medicinal or non-medicinal it will take more time. So that we can create an application by using this project. Then, it will be easy to identify the medicinal plants. It is easy to use and also takes less time.

REFERENCES

- [1] Amin, A.H.M. and Khan, A.I., 2013. One- Shot Classification of 2-D Leaf Shapes using Distributed Hierarchical Graph Neuron (DHGN) Scheme with a k-NN Classifier. *Procedia Computer Science*, 24, 84-96.
- [2] Arai, K., Abdullah, I.N. and Okumura, H., 2013. Identification of Ornamental Plant Functioned as Medicinal Plant-Based on Redundant Discrete Wavelet Transformation. *International Journal of Advanced Research in Artificial Intelligence*, 2(3), 60-64.
- [3] Babatunde, A., Armstrong, L., Diepeveen,D. and Leng, J., 2015. A survey of computer- based vision systems for automatic identification of plant species. *Journal of Agricultural Informatics*, 6(1), 61-71.
- [4] Backes, A.R., Casanova, D. and Bruno, O.M., 2009. Plant Leaf Identification based on Volumetric Fractal Dimension. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(6), 145- 1160.
- [5] Carranza-Rojas, J. and Mata-Montero, E., 2016. Combining Leaf Shape and Texture for Costa Rican Plant Species Identification. *CLEI Electronic Journal*, 19(1), Paper 7.
- [6] Chaki, J., Parekh, R. and Bhattacharya, S, 2015. Plant leaf recognition using texture and shape features with neural classifiers. *Pattern Recognition Letters*, 58, 61-68.
- [7] Du, J.X., Wang, X.F. and Zhang, G.J., 2007. Leaf shape based plant species recognition. *Applied Mathematics and Computation*, 185, 883- 893.
- [8] Du, J.X., Zhai, C.M. and Wang, Q.P., 2013. Recognition of plant leaf image based on fractal dimension features. *Neurocomputing*, 116, 150- 156.
- [9] Du, M., Zhang, S. and Wang, H., 2009. Supervised Isomap for Plant Leaf Image Classification. *5th International Conference on Emerging Intelligent Computing Technology and Applications*, Ulsan, South Korea, 627-634.

- [10] Gao, W. and Lin, W., 2012. Frontal Parietal Control Network Regulates the Anti-Correlated Default and Dorsal Attention Networks. *Human Brain Mapping*, 33(1), 192-202.
- [11] Herdiyeni, Y. and Wahyuni, N.K.S., 2012. Mobile Application for Indonesian Medicinal Plants Identification using Fuzzy Local Binary Pattern and Fuzzy Color Histogram. *International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, West Java, Indonesia, 301-306.
- [12] Hernandez-Serna, A. and Jimenez-Segura, L.F., 2014. Automatic Identification of species with neural networks. *PeerJ* 2:e563; DOI:10.7717/peerj.563.
- [13] Hossain, J. and Amin, M.A., 2010. Leaf Shape Identification Based Plant Biometrics. *13th International Conference on Computer and Information Technology*, Dhaka, Bangladesh, 458-463.
- [14] Le, T.L., Tran, D.T. and Hoang, V.N., 2014. Fully Automatic leaf-based plant identification, application for Vietnamese medicinal plant search. *Fifth Symposium on Information and Communication Technology*, Hanoi, Vietnam, 146-154.
- [15] Mata-Montero, E. and Carranza-Rojas, J., 2016. Automated Plant Species Identification: Challenges and Opportunities. *IFIP Advances in Information and Communication Technology*, 481, 26-36.
- [16] Munisami, T., Ramsurn, M., Krishnan, S. and Pudaruth, S., 2015. Plant leaf recognition using shape features and color histogram with k- nearest neighbor classifiers. *Procedia Computer Science*, 58, 740-747.
- [17] Prasvta, D.S. and Herdiyeni, Y., 2013. Red Leaf: Mobile Application for Medicinal Plant Identification Based on Leaf Image. *International Journal of Advanced Science, Engineering and Information Technology*, 3, 5- 8.
- [18] Siravenha, A.C.Q. and Carvalho, S.R., 2015. Exploring the use of Leaf Shape Frequencies for Plant Classification. *28th SIBGRAPI Conference on Graphics, Patterns, and Images*, Salvador, Brazil, 297-304.

[19] Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F. and Xiang, Q.L., 2007. A Leaf Recognition Algorithm for Plant Classification using Probabilistic Neural Network. 7th IEEE International Symposium on Signal Processing and Information Technology, Giza, Egypt, 11- 16.

[20] Zhang X., Liu Y., Lin H., Liu Y. (2016) Research on SVM Plant Leaf Identification Method Based on CSA. In: Che W. et al. (eds) Social Computing. ICYCSEE 2016. Communications in Computer and Information Science, Vol 624, Springer, Singapore.

APPENDICES

A.SOURCE CODE

```
import cv2
from PIL import Image
import os, sys
import os, glob
from os import listdir, makedirs
from os.path import isfile, join
path = 'D:/codes/medical and non medical plant/dataset/test/non'
# Source Folder
dstpath = 'D:/codes/medical and non medical plant/dataset/test/graynon' #
Destination Folder
try:
    makedirs(dstpath)
except:
    print ("Directory already exist, images will be written in same folder")
# Folder won't used
files = [f for f in listdir(path) if isfile(join(path,f))]
for image in files:
    try:
        img = cv2.imread(os.path.join(path,image))
        gray = cv2.cvtColor(img,cv2.COLOR_BGR2GRAY)
        dstPath = join(dstpath,image)
        cv2.imwrite(dstPath,gray)
    except:
        print ("{} is not converted".format(image))
for fil in glob.glob("*.jpg"):
    try:
        image = cv2.imread(fil)
        image = cv2.resize(image,(256,256))
        gray_image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY) # convert
to greyscale
```

```

        cv2.imwrite(os.path.join(dstpath,fil),gray_image)
except:
    print('{} is not converted')
    import numpy as np
    x=np.mean(img)
    print ('MEAN=',x)
    y=np.std(img)
    print('STANDARD DEVIATION=', y)
    z=np.var(img)
    print('VARIANCE=',z)
    break

from keras.models import Sequential
from keras.layers import Convolution2D
from keras.layers import MaxPooling2D
from keras.layers import Flatten
from keras.layers import Dense
# Initializing the CNN
classifier = Sequential()
# Step 1 - Convolution
classifier.add(Convolution2D(32, 3, 3, input_shape=(64, 64, 3),
activation='relu'))
# Step 2 - Pooling
classifier.add(MaxPooling2D(pool_size=(2, 2)))
# To increase Efficiency, add another Convolutional layer
classifier.add(Convolution2D(32, 3, 3, activation='relu'))
classifier.add(MaxPooling2D(pool_size=(2, 2)))
# Step 3 - Flattening
classifier.add(Flatten())
# Step 4 - Full Connection
classifier.add(Dense(output_dim=128, activation='relu'))
# Output layer
classifier.add(Dense(output_dim=2, activation='softmax'))
# Compiling the CNN
classifier.compile(optimizer='adam',loss='categorical_crossentropy',

```

```

metrics=['accuracy'])
# Part 2 : Fitting the CNN to the images
from keras.preprocessing.image import ImageDataGenerator
train_datagen = ImageDataGenerator(
    rescale=1. / 255,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True)
test_datagen = ImageDataGenerator(rescale=1. / 255)
training_set=train_datagen.flow_from_directory('C:/Users/Kallyani/New
folder/medical and non medical plant/dataset/train/',
                                              target_size=(64, 64),
                                              batch_size=5,
                                              class_mode='categorical')
test_set = test_datagen.flow_from_directory('C:/Users/Kallyani/New
folder/medical and non medical plant/dataset/test/',
                                              target_size=(64, 64),
                                              batch_size=5,
                                              class_mode='categorical')
classifier.fit_generator(training_set,
                        steps_per_epoch=(210/4),
                        epochs=30,
                        validation_data=test_set,
                        validation_steps=(90 /4))
# Making New Predictions
import numpy as np
from keras.preprocessing import image
import cv2
test_image = image.load_img('C:/Users/Kallyani/New folder/medical and non
medical plant/non2.png'/target_size=(64,64))
test_image = image.img_to_array(test_image)
test_image = np.expand_dims(test_image, axis=0)
result = classifier.predict(test_image)
training_set.class_indices

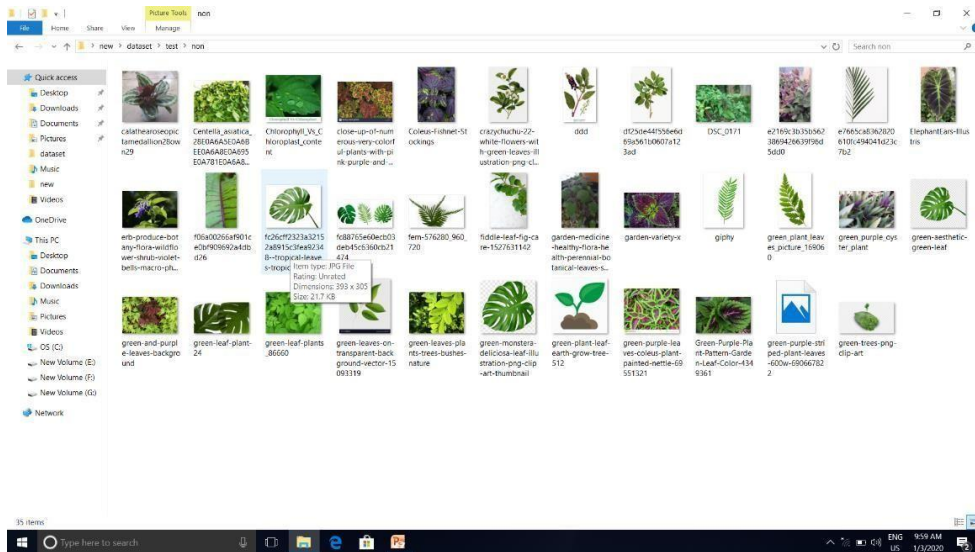
```

```

path= r'C:/Users/Kallyani/New folder/medical and non medical plant/non2.png'
image=cv2.imread(path)
window_name="detection"
if result[0][0] == 1:
    prediction = 'medical'
    color = (255, 0, 0)
    font = cv2.FONT_HERSHEY_SIMPLEX
    org = (50, 50)
    fontScale = 1
    thickness = 2
    image = cv2.putText(image, 'Medical', org, font,
        fontScale, color, thickness, cv2.LINE_AA)
    cv2.imshow(window_name,image)
else:
    prediction="non-medical"
    color = (255, 0, 0)
    font = cv2.FONT_HERSHEY_SIMPLEX
    org = (50, 50)
    fontScale = 1
    thickness = 2
    image = cv2.putText(image, 'not-medical', org, font,
        fontScale, color, thickness, cv2.LINE_AA)
    cv2.imshow(window_name,image)
print(prediction)

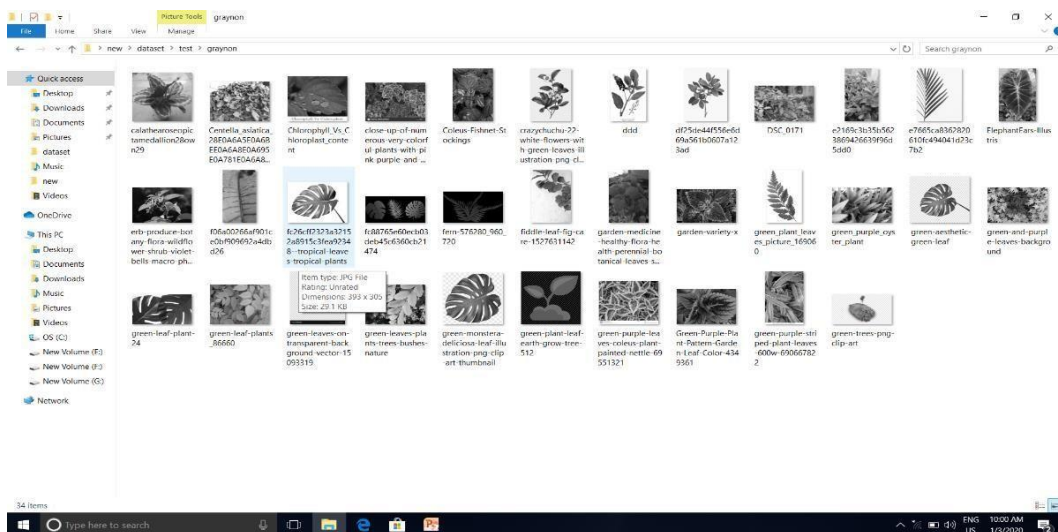
```

B.SCREEN SHOTS



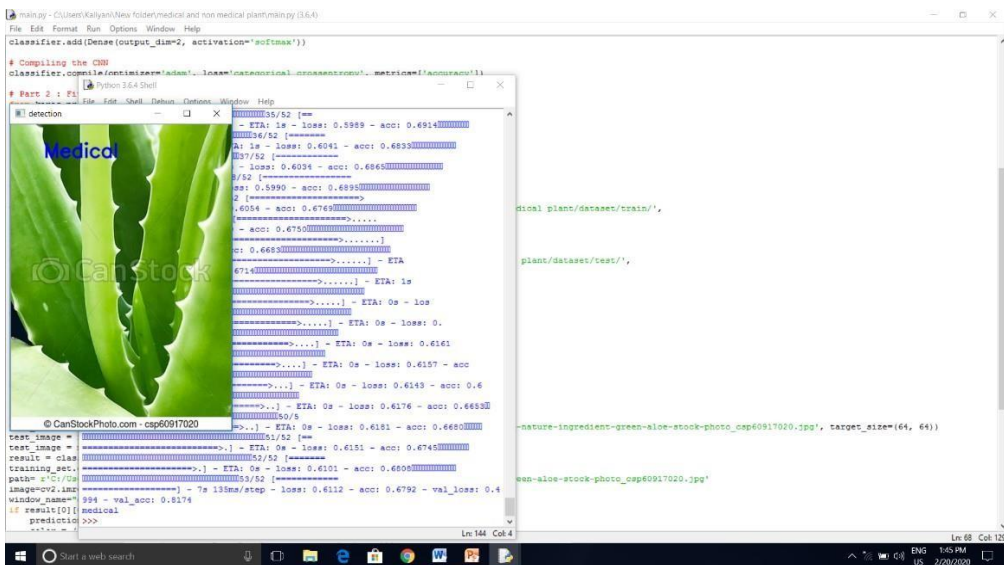
Images before conversion into gray scale

In the above image we have collected the group of medicinal and non-medicinal images.



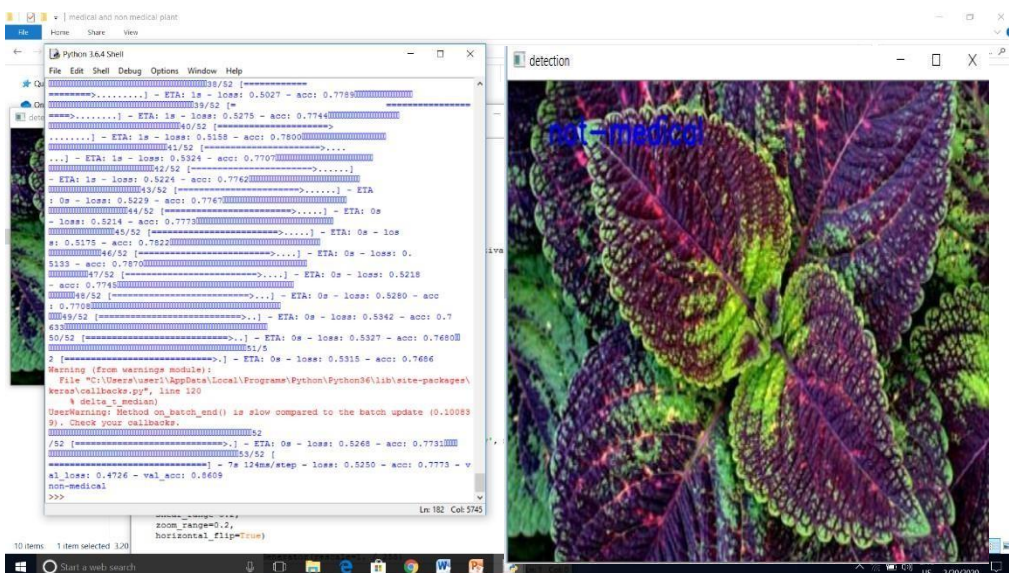
Images after gray scale conversion

In the above image the collected medicinal and non-medicinal plants are converted into gray scale images.



Recognition of medicinal plants

In above image shows that the given plant is medicinal.



Recognition of non-medicinal plants

In the above image shows that the given plant is non-medicinal

C.PAPER PUBLISH

AUTOMATIC RECOGNITION OF MEDICINAL PLANTS

KRISHNA VINESHA P^{[1]*}, LAKSHMI PRIYANKA P^{[2]*}, LAKSHMANAN L^{[3] *}

¹*, UG Student, Dept. of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India.

³Associate Professor, Department of Computer Science and Engineering,

Sathyabama Institute of Science and Technology, Chennai, India

pabbithivineesha@gmail.com¹, Lakshmi.priyanka565@gmail.com², lakshmanan.cse@sathyabama.ac.in³

Abstract– A fully robotic strategy for recognizing therapeutic plants using PC vision and artificial intelligence methods has been demonstrated. The leaves of 25 distinctive therapeutic plant types were gathered and captured using an advanced cell in a research environment. Countless highlights have been separated from each sheet, for example length, width, border and area, number of vertices, shading, border and frame territory. Then some highlights determined by these properties have been elaborated. The best results were obtained by a CNN algorithm that uses a 10 times cross approval strategy. By an accuracy of 98.3%, the CNN algorithm accomplish superior than any other AI approach, such as the nearest K-Neighbor, Naïve Bayes, KNN and Neural systems. Those results are enriching and later work will be equipped to use a huge data set and elite processing offices to explore the exposure of deep learning neural systems to recognize restoration plants used in essential medicinal services. It is expected that an electronic or portable PC framework for the programmed recognition of restoration plants will guide the population of the neighborhood to enhance their knowledge of therapeutic plants, to help taxonomists develop recognizable test methods for increasingly competent species.

Keywords:- Medicinal Plant Recognition; CNN ; Image recognition;

Introduction

Plant ID has a significant job in current logical issues, for example, biodiversity, environment, and pharmacology among others. In Biology, plant distinguishing proof includes breaking down numerous organs, for example, blossoms, seeds, leaves and woody parts [18, 20]. This methodology renders the errand troublesome as blossoms and seeds, which are occasional and subject to the plant's age and condition, are difficult to discover. In exceptional circumstances, for example, discovering fossils or uncommon plants, the material accessible to distinguish a plant is only the leaves. To comprehend these circumstances, a leaf morphological scientific categorization method is proposed, which considers just the leaves to play out the distinguishing proof errands [14]. This methodology joins different highlights of leaves, for example, shape, vein structure, surface, and some histological data. Recognizing plants is a troublesome and complex undertaking because of the idea of the leaves. Although that the leaves present some major highlights, they additionally present a wide example of the variety. This variety may happen in various leaves from a similar plant, where attributes, for example, development and presentation to the sun produce varieties in the size, shading, surface, and state of the leaves. These varieties are additionally present in leaves from similar species, however from various plants. Right now, are an outcome of soil impact, atmosphere or even condition when the leaf is being shaped.

The frames have grown so far using programmed order procedures, however the procedures are very comparative. These means include the configuration of the collected leaves, the realization of some previous managements to distinguish their particular characteristics, the arrangement of the leaves, the compilation of the database, the preparation for the recognition and evaluation of the results. The world has a more number of plant species, a significant amount of which have remedial qualities, close to elimination and others are destructive to humans. The recognition of dark plants depends very much on the intrinsic information of a specialized botanist. The best technique for

distinguishing plants effectively and effectively is a manual processing strategy based on morphological qualities. In this way, a large number of procedures relating to the organization of these plant species "depend on the accumulation of information and the skills of individuals" [1]. Be that as it may, this procedure of manual acknowledgment is regularly relentless and time consuming. Henceforth numerous scientists have directed examinations to help the programmed grouping of plants dependent on their physical qualities.

The frames have grown so far using programmed order procedures, however the procedures are very comparative. These means include the configuration of the collected leaves, the realization of some previous managements to distinguish their particular characteristics, the arrangement of the leaves, the compilation of the database, the preparation for the recognition and evaluation of the results. A computerized plant distinguishing proof framework can be utilized by non-botanical specialists to rapidly recognize plant species easily.

Normally customary individuals are doled out with the activity of gathering the plants from the timberlands. Once in a while they couldn't perceive the uncommon and significant plants due to human blunder. These uncommon sorts of plants are critical to spare the life of a patient. Additionally, here and there these individuals could get off base species which might be hurtful plants. In such cases, it is important to utilize the programmed plant acknowledgment framework. This framework helps a conventional people or any layman to perceive the diverse plant species. These sorts of frameworks are likewise extremely accommodating for the trekking individuals if they are intrigued to gather the plant species while trekking the mountains.

Related Work

Some tests have been performed to create tools for distinctive tests on plants in the past 10 years. Wu et al. They completed the most definitive works in the plant field disposal. [2] From five fundamental graphic aspects, twelve structural aspects are deduced and, therefore, the analysis of the main components (PCA) is used to reduce the measurement in order to send fewer sources of information to a feasibility neural system (PNN). They achieved normal accuracy of 90.3% with the Flavia set of data, which is its creation. Using an alternative dataset but a similar algorithm, Hossain et al [4] have achieved a degree of accuracy comparable to the comparative highlights [4].

The image was in the size of 256 * 256 pixels. Du et al [11] proposed a methodology that depends on the measurement of the fractal highlighted on the basis of the shape of the leaf and the designs of the veins for the recognition and order of the leaves of the plant [11]. By using a closer k-classifier with 20 highlights, they achieved the high identification rate of 85.1%. Using a form of volumetric fractal measurement to try to create a surface mark by a sheet and calculating the linear discriminant analysis (LDA), Backes et al [12] had the opportunity to overcome the conventional methodologies that depended on the Gabor channels and the Fourier test achieved 87.3% accuracy in a data set of 640 leaves from 32 distinctive plant species [13]. They used to model and combine data as it were. Sixteen sources of information (6 geometric aspects, 8 surfaces and 2 morphological highlights) were sent to a counterfeit neuronal system (ANN) with 60 centers in the covered layer and a learning frequency of over 0.100 pages. Using the equivalent data set, Chaki et al [15] obtained an accuracy of 97.6% using a Neuro-Fuzzy (NFC) algorithm with a 44-component surface vector and a 3-component shape vector [15]. Use of the module only includes the Flavia and Pattern Net data set (a type of neural system). His neuronal advancement system consists of 2 hidden layers with twenty six neurons each and was prepared for more than 100 pages.

Fascinating work has been done. They performed the calculation of the scale curvature histogram (HCoS) to extract the shape data and the modification of the almost corresponding example (LBPV) to eliminate the surface data. In the best case, the clean dataset exceeded the 7.3% dataset. This recommends that legitimately acquired images using a mobile phone can create good levels of proven accuracy and images that are physically prepared in a laboratory and then grouped together. Previously,

Amin et al [18] used an appropriate neuronal level diagram (DHGN) to acquire fold data using 64 element vectors and the closest k classifier with Canberra separation to acquire 71.5% accuracy [18]

SYSTEM ARCHITECTURE

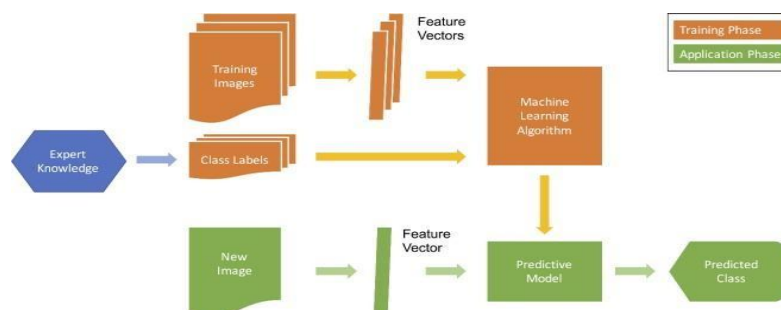


Fig. 1 Overall System Design

The initial step is to delete the RGB channel data from the information image to distinguish the implant data from the anterior area. To do this, data is extracted from the excess green (ExG) and excess red (Exp) channels. Then, the red overabundance data is removed from the abundance of green data to progressively contain the green channel data. When the double limit is applied in the improved vegetative archive, the ExGExR data, a double intrigue district is divided into plants. The next advance is to use the first double image shadow pixel as a cover that separates the leaves as secondary images. The division of the test sheet images into the dataset using the vegetative record mentioned above. The important component in the representation of an image is shading. The shading minutes are obtained through the highlighted facts, for example the average and the standard deviation evaluated in the three RGB channels (red, green, blue). The main measures for requesting an image are acquired from the standard deviation and the average.

The surface of the image can be estimated using the concurrency pane, the different shading elements or the grayscale estimates of the image. The highlights that are created using this strategy are known as Haralick highlights. A calculated regression classifier is used to distinguish the restoration plan. Since our proposed strategy consists of the characterization of numerous classes, the relapse model calculated in parallel has been extended to the calculated multinomial relapse. Through a combination of relapses calculated in pairs, the multinomial logit thinks of different encounters. This allows you to compare each brand class with a reference classification in the data set.

Algorithm Used

A convolutional neural network (CNN) is a deep friendly neural network designed from biologically controlled modules. They are applied most frequently used for image processing. CNN's work differently in that they treat data as spatial. Instead of being connected to all the neurons in the previous layer, the neurons are connected only to neighbouring neurons and all have the same weight. This simplification in connections means that the network maintains the spatial aspect of the data set. The word convolution refers to the filtering process that occurs in this type of network. If an image is complex, the convolutional neural network simplifies it, so it can be better elaborated and understood. Like a normal neural network, CNN is also composed of several levels. There are a couple of levels that make the convolutional level and the grouping level unique. However, like other neural networks, it will also have a rectified linear unit layer and a fully connected layer. The ReLU level acts as an activation function, ensuring non-linearity as the data moves through each level of the network. Without it, the data entered in each level would lose the dimensionality that we want to maintain. Meanwhile, the fully connected layer allows you to classify the data set.

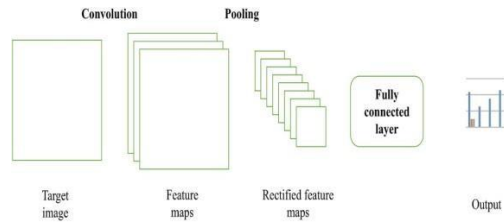


Fig 2 Steps in CNN

Results and Discussions

The obtained results of training accuracy and validation accuracy values are shown in following graph.

In this diagram for where there has less number of epochs (upto 13) we can know that validation accuracy is same as the training accuracy. In case if we need more number of epochs the training accuracy and validation accuracy values are almost comparable.

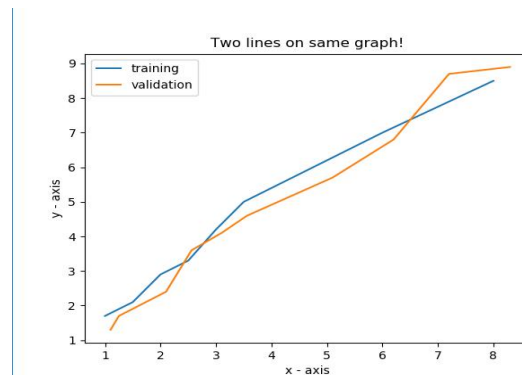


Fig 3 Training accuracy vs validation accuracy graph

Conclusion

Another set of data on therapeutic plants in Mauritius has been made openly accessible at the entrance of the AI store. At this time, vision strategies have been used to extract some reflexes based on the shape of the leaves of the therapeutic plants. AI calculations were used to group the leaves of 25 distinctive plant types into their suitable classifications. The most remarkable precision of 98.3% was obtained by the irregular forest classifier. Normally conventional individuals are doled out with the activity of gathering the plants from the timberlands. Once in a while, they couldn't perceive the uncommon and significant plants in light of human blunder. These uncommon sorts of plants are essential to spare the life of a patient. Likewise here and there these individuals could get erroneous species which might be hurtful plants. In such cases, it is important to utilize the programmed plant acknowledgment framework. These sorts of frameworks are likewise helpful for the trekking individuals.

Future Work

It is necessary to study about medicinal plants that find place in folklore. As a result, such plants should be investigated to understand their physical and chemical properties as well as safety and efficacy. The medicinal plants find applications in many industries. The use of medicinal plants for curing diseases has been increased. In order to identify the medicinal or non-medicinal it will take more time. So that we can create an application by using this project. Then, it will be easy to identify the medicinal plants. It is easy to use and also takes less time.

Program Code:

```
import cv2
from PIL import Image
import os, sys
import os, glob
from os import listdir, makedirs
from os.path import isfile, join
path = 'D:/codes/medical and non medical plant/dataset/test/non'
# Source Folder
dstpath = 'D:/codes/medical and non medical plant/dataset/test/graynon' # Destination Folder
try:
    makedirs(dstpath)
except:
    print ("Directory already exist, images will be written in same folder")
# Folder won't used
files = [f for f in listdir(path) if isfile(join(path, f))]
for image in files:
    try:
        img = cv2.imread(os.path.join(path, image))
        gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
        dstPath = join(dstpath, image)
        cv2.imwrite(dstPath, gray)
    except:
        print("{} is not converted".format(image))
for fil in glob.glob("*.jpg"):
    try:
        image = cv2.imread(fil)
        image = cv2.resize(image, (256, 256))
        gray_image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY) # convert to greyscale
        cv2.imwrite(os.path.join(dstpath, fil), gray_image)
    except:
        print('{} is not converted')
        import numpy as np
        x = np.mean(img)
        print('MEAN=', x)
        y = np.std(img)
        print('STANDARD DEVIATION=', y)
        z = np.var(img)
        print('VARIANCE=', z)
        break
from keras.models import Sequential
from keras.layers import Convolution2D
from keras.layers import MaxPooling2D
from keras.layers import Flatten
from keras.layers import Dense
# Initializing the CNN
classifier = Sequential()
# Step 1 - Convolution
classifier.add(Convolution2D(32, 3, 3, input_shape=(64, 64, 3), activation='relu'))
# Step 2 - Pooling
classifier.add(MaxPooling2D(pool_size=(2, 2)))
```

References:

- [1] Amin, A.H.M. and Khan, A.I., 2013. One- Shot Classification of 2-D Leaf Shapes using Distributed Hierarchical Graph Neuron (DHGN) Scheme with a k-NN Classifier. *Procedia Computer Science*, 24, 84-96.
- [2] Arai, K., Abdullah, I.N. and Okumura, H., 2013. Identification of Ornamental Plant Functioned as Medicinal Plant-Based on Redundant Discrete Wavelet Transformation. *International Journal of Advanced Research in Artificial Intelligence*, 2(3), 60-64.
- [3] Babatunde, A., Armstrong, L., Diepeveen, D. and Leng, J., 2015. A survey of computer- based vision systems for automatic identification of plant species. *Journal of Agricultural Informatics*, 6(1), 61-71.
- [4] Backes, A.R., Casanova, D. and Bruno, O.M., 2009. Plant Leaf Identification based on Volumetric Fractal Dimension. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(6), 145- 1160.
- [5] Carranza-Rojas, J. and Mata-Montero, E., 2016. Combining Leaf Shape and Texture for Costa Rican Plant Species Identification. *CLEI Electronic Journal*, 19(1), Paper 7.
- [6] Chaki, J., Parekh, R. and Bhattacharya, S., 2015. Plant leaf recognition using texture and shape features with neural classifiers. *Pattern Recognition Letters*, 58, 61-68.
- [7] Du, J.X., Wang, X.F. and Zhang, G.J., 2007. Leaf shape based plant species recognition. *Applied Mathematics and Computation*, 185, 883- 893.
- [8] Du, J.X., Zhai, C.M. and Wang, Q.P., 2013. Recognition of plant leaf image based on fractal dimension features. *Neurocomputing*, 116, 150- 156.
- [9] Du, M., Zhang, S. and Wang, H., 2009. Supervised Isomap for Plant Leaf Image Classification. 5th International Conference on Emerging Intelligent Computing Technology and Applications, Ulsan, South Korea, 627-634.
- [10] Gao, W. and Lin, W., 2012. Frontal Parietal Control Network Regulates the Anti-Correlated Default and Dorsal Attention Networks. *Human Brain Mapping*, 33(1), 192–202.
- [11] Herdiyeni, Y. and Wahyuni, N.K.S., 2012. Mobile Application for Indonesian Medicinal Plants Identification using Fuzzy Local Binary Pattern and Fuzzy Color Histogram. *International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, West Java, Indonesia, 301-306.
- [12] Hernandez-Serna, A. and Jimenez-Segura, L.F., 2014. Automatic Identification of species with neural networks. *PeerJ* 2:e563; DOI:10.7717/peerj.563.
- [13] Hossain, J. and Amin, M.A., 2010. Leaf Shape Identification Based Plant Biometrics. 13th International Conference on Computer and Information Technology, Dhaka, Bangladesh, 458-463.
- [14] Le, T.L., Tran, D.T. and Hoang, V.N., 2014. Fully Automatic leaf-based plant identification, application for Vietnamese medicinal plant search. Fifth Symposium on Information and Communication Technology, Hanoi, Vietnam, 146-154.
- [15] Mata-Montero, E. and Carranza-Rojas, J., 2016. Automated Plant Species Identification: Challenges and Opportunities. *IFIP Advances in Information and Communication Technology*, 481, 26-36.
- [16] Munisami, T., Ramsurn, M., Krishnan, S. and Pudaruth, S., 2015. Plant leaf recognition using shape features and color histogram with k- nearest neighbor classifiers. *Procedia Computer Science*, 58, 740-747.
- [17] Prasvta, D.S. and Herdiyeni, Y., 2013. Red Leaf: Mobile Application for Medicinal Plant Identification Based on Leaf Image. *International Journal of Advanced Science, Engineering and Information Technology*, 3, 5– 8.
- [18] Siravenha, A.C.Q. and Carvalho, S.R., 2015. Exploring the use of Leaf Shape Frequencies for Plant Classification. 28th SIBGRAPI Conference on Graphics, Patterns, and Images, Salvador, Brazil, 297-304.
- [19] Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F. and Xiang, Q.L., 2007. A Leaf Recognition Algorithm for Plant Classification using Probabilistic Neural Network. 7th IEEE International Symposium on Signal Processing and Information Technology, Giza, Egypt, 11- 16.
- [20] Zhang X., Liu Y., Lin H., Liu Y. (2016) Research on SVM Plant Leaf Identification Method Based on CSA. In: Che W. et al. (eds) *Social Computing. ICYCSEE 2016. Communications in Computer and Information Science*, Vol 624, Springer, Singapore.

PLAGIARISM REPORT

WhatsApp x Automatic Recognition of medic x +

File | C:/Users/user1/Downloads/Automatic%20Recognition%20of%20medicinal%20plants%20(6)%20(1).pdf

Automatic Recognition of medicinal plants (6) (1).pdf 11 / 11

Automatic Recognition of medicinal plants

ORIGINALITY REPORT

1%	1%	1%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	P. Saikumar, P. Bharadwaja, J. Jabez. "Android and Bluetooth Low Energy Device Based Safety System", 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), 2019	1%
	Publication	

Type here to search

ENG 5:19 PM
US 4/24/2020

