BAHIR DAR UNIVERSITY

BAHIR DAR INSTITUTE OF TECHNOLOGY

SCHOOL OF RESEARCH AND POSTGRADUATE STUDIES

FACULTY OF COMPUTING

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

MSc Thesis Proposal

DEVELOP A MODEL FOR RECOGNITION, CLASSIFICATION AND DESCRIPTION OF THE HERBS PLANT FOR PATIENT CARE USING DEEP LEARNING

SURAFEL AMSALU TADESSE

BDU1402966

A thesis proposal submitted to the school of graduate studies of Bahir Dar Institute of Technology, BDU in partial fulfillment of the degree of Masters in Artificial Intelligence and Data Science in the Faculty of Computing

ADVISOR: GEBEYHU BELAY(PHD, ASSOCIATE PROFESSOR)

DECEMBER 2022

BAHIR DAR, ETHIOPIA

Contents

LIST OF FIGURES

LIST OF Tables

Abbreviations

AI - Artificial Intelligence

BDU - Bahir Dar University

BiT - Bahir Dar Institute of Technology

CNN - Convolutional Neural Network

ICT - Information and Communication Technologies)

ML - Machine learning

NLP - Natural Language Processing

RNN - Recurrent Neural Network

INTRODUCTION

Medicinal plants have been used in traditional medicine practices for a long time because of their nutrients and medicinal properties. (Azadnia et al., 2022) The world's primary means of treating diseases and fighting infections have been based on the use of medicinal plants. Due to their critical role in maintaining human health, medicinal plants have been researched and taken into consideration. Around 85% of the entire population uses herbal medicines for disease treatment, and it's also increasing significantly. The World Health Organization (WHO) estimates that almost 3.5 billion people in underdeveloped nations, including Ethiopia, frequently utilize plant treatments and consider them to be effective. However, it takes a skilled expert a lot of time and effort to identify medicinal plants. (Azadnia et al., 2022)

The human body is complex and organic, while chemical medicines contain chemicals that are inorganic and pure. Therefore, chemical medicines are considered not very suitable for consumption by the human body and, if consumed continuously, can even be bad for human health. However, some chemical drugs are symptomatic (temporary), so they must be taken for life by patients with certain diseases. (Putri et al., 2021)

Ethiopia is located in the Horn of Africa between 3 and 15° north latitude and 33 and 48° east longitude, and also comprises ten national, two regional, and two administrative states with varied agro-ecological zones. Since the country is characterized by a wide range of ecological, edaphic, and climatic conditions, Ethiopia is also very diverse in its flora composition. The flora of Ethiopia is estimated to contain close to 6500–7000 species, including medicinal plants; of those, 12–19% are endemic to the country. Medicinal plants have been used for various types of human and animal treatments in the country. In Ethiopia, about 80% of the human population and 90% of livestock rely on traditional medicine. The medicinal plants have shown very effective medicinal values for some diseases of humans and livestock. (Moges & Moges, 2020). During this study, we focus on different types of plants, especially those plants that we found in Ethiopia. And also this research is not only focusing on plant leaf images but also the entire plant image or the fruit images. Medicinal plants and herbs have continued to be used for the

traditional management of various illnesses in many societies since time immemorial. With advancements in artificial intelligence (AI) and different information and communication technologies (ICTs), the need to automatically identify medicinal plants from the thousands of plant species can only continue to grow (Malik et al., 2022).

Automated identification systems based on computer vision and machine learning techniques provide an alternative and assistive approach for this task. These systems are useful, but their accuracy varies due to the diversity of the species. The identification of the medicinal plant parts can be done using image processing and chemical ingredient extraction. However, using the medicinal plant's chemical ingredients consumes time and requires high expenditures for laboratory equipment. So, to reduce this problem, image processing is the preferred approach. Deep Learning is proposed for handling this approach. Deep learning can be used in speech recognition, natural language processing, machine translation, audio recognition, bioinformatics, drug design, medical image identification, and medicinal plant identification and classification.

The convolutional neural network (CNN) has achieved tremendous success in computer vision applications with impressive accuracy during the past several years. A convolutional neural network (CNN or ConvNet) is a network architecture for deep learning that learns directly from data. CNNs are particularly useful for finding patterns in images to recognize objects, classes, and categories.

A convolutional neural network can have tens or hundreds of layers that each learn to detect different features of an image. Filters are applied to each training image at different resolutions, and the output of each convolved image is used as the input to the next layer. The filters can start with very simple features, such as brightness and edges, and increase in complexity to include features that uniquely define the object. They can also be quite effective for classifying audio, time series, and signal data. But in this case, we used convolutional neural networks (CNN) to identify or find a pattern for the plant images.

Among the deep learning methods that will be applied and recommended for use for image classification and identification are VGG-19, Mobile Net, VGG-16, Inception V3, and so on. One of the main points of this study is to increase the ability to cure disease using plant parts and

the knowledge of those indigenous plants in modern medical science. There are reportedly 450,000 different types of plants, but one-third of those are facing extinction. Through this study, we could also save thousands of plants, especially herbal plants, by identifying and recognizing them and ensuring that those who harvest them give them proper care.

RELATED WORKS

There are several types of research conducted for identifying and recognizing a plant image, but most of them are working on only leaf image detection and recognition, and they only detect the image and write only the name of those plants; other than this, we did not see anything. Of course, they wrote about how the model was developed or what types of algorithms they used to detect the plant images, but these things only help or are used by academics and are not useful to end users.

Plant image detection and recognition

With the development of image processing and pattern recognition, it is possible to apply them to recognize plants automatically. Many studies in the past decades have shown that leaf contains rich information (e.g. color, shape, texture) for recognition. The shape is the general feature of a leaf. As the color of a leaf may vary with the climatic and seasonal conditions, and most plants have similar colors (e.g., green), color is not commonly used in classification.

There are many agricultural problems currently addressed by classical machine vision techniques that may benefit from a deep learning approach. We consider this paper a successful example of this behavior by applying deep learning to automatic plant identification. Automatic plant identification constitutes a challenging problem that has received increasing attention in recent years, in particular for identification based on leaf image analysis. Much of this work makes use of leaf features that humans can perceive. The goal of automatization in this case is to avoid the use of human experts handling huge catalogs of plant species and to reduce classification time. Compared with the flowers or vein structures of plants, leaves can be easily captured throughout the year. More importantly, they can also provide enough discriminative features for plant recognition. Recently, automatic leaf recognition has become a research focus in the computer

vision and intelligent systems areas. Several representative methods were proposed to recognize plants automatically. relying on shape features and a hypersphere classifier (Derso, 2021; Grinblat et al., 2016; Lee et al., 2017; Moges & Moges, 2020; Tafese Awulachew, 2021; Z. Wang et al., 2014). Computerized plant classification systems are mostly based on two-dimensional images. This makes plant classification based on leaves an appropriate choice compared to the use of shapes of flowers, seedlings, and morphs of plants, which are three-dimensionally complex in structure. Plant classification based on leaves involves leaf feature extraction. Leaf feature extraction is a process of identifying features that can discriminate different kinds of leaves.

Compared with other methods, such as cell and molecule biology methods, classification based on leaf image is the first choice for leaf plant classification. Sampling leaves and photographing them is low-cost and convenient. One can easily transfer the leaf image to a computer, and a computer can extract features automatically through image processing techniques.

Canopy structure and leaf shape have been key features for plant species identification by weed specialists. Machine vision and shape feature analysis are usually the final steps of a process that includes identifying green plant regions of interest and then isolating individual plant canopy crowns.(Abera, 2014; Aman et al., 2020; Amuthalingeswaran et al., 2019; Neto et al., 2006; B. Wang et al., 2013; Wu et al., 2007)

PROBLEM STATEMENT

In the introduction part, we deal with and identify medicinal plants that need skills such as a manual or scientific process and characterization for proper recognition and classification. However, the manual process is laborious and time-consuming. Recently, technological development has called for more efficient methods to meet species' identification requirements, such as developing digital image processing and pattern recognition techniques (Malik et al., 2022). Most of the systems developed for identifying the herbal plants are not useful for the majority of the Ethiopian people because, according to a study conducted by "macrotrends," almost 100 million people live in the rural part of Ethiopia, and the literacy rate is also half of the

entire Ethiopian population, which means over 50 million people cannot read or write. The system or model developed by the previous generation did not consider this.
On the medication side, a research team in “transformational treatment” researched how many Americans died each year because of medication. The number shows us more than 20,000 people died because of medication. The important note that we should see in this number is that all the medicine they took was not expired or had any other problem, but their bodies did not get along with that medication anymore, which is what makes it fatal. On the other hand, people who choose herbal plants for medicine tend to be relieved quickly and with no side effects as well.
This study will answer the following questions:

1. How to characterize medicinal plants for proper recognition and classification in a consideration of their medicinal values?
2. How to design and develop a recognition and classification model using deep learning techniques??
3. What factors can be influential for this model Performance to identify and describe those herbaceous plants
4. How can we use this model as an image and text search engine?

OBJECTIVE OF THE STUDY

General objective

The general objective of this study is to design and develop a model for medicinal plant recognition, the identification, description, and administration of the herbal plant.

Specific Objectives

1. To characterize medicinal plants for proper recognition and classification.
2. To design and develop a recognition and classification model using deep learning techniques.
3. To perform the performance of this model.
4. To use this model as an image and text search engine.

METHODOLOGY

Research design

In this study, we will use CNN, which is the backbone of deep learning algorithms and can automatically extract multiple unique features of fruits and leaves as well as the entire plant image.

According to a study (Widneh et al., 2021), to identify a leaf, it's a good idea to get an image of the leaf from the backside. It has (leaf) has more recognized pattern than on the front. For the other part of the plant, we use the regular image.

In this research six procedures were followed such as

1. Collecting plant image
2. Labeling plant image
3. Process the plant image
4. Extracting features for the plant image
5. Description of medicinal plant part (Purpose of the herbal plant)
6. Administration of the Herbal Plant(How to use them)

Data collection and preparation

The data was collected using a high-resolution camera (Samsung Note8 with 24MP and Some websites). The data collection might take some time because there are a lot of plants out there, and choosing a plant and labeling it for Amharic and English is not only for the description but also for the administration purpose as well.

Implementation and algorithm

The fine-tuned models of CNN, such as Mobile Net, VGG16, and Inception_V3, are adopted to extract useful leaves, fruits, and herbal plant features and feed them to the classifier. These feature extractions are processed in the feature learning layer part of CNN(Widneh et al., 2021)

After we collect the data set, we will rename it, writing the description for each plant and also the administration after the model identifies or recognizes those are the steps it followed. Although we will choose the perfect algorithm and state-of-the-art algorithm, we may look for a more perfect and precise algorithm because we are attempting to classify and recognize not only the leaves but also the fruit and the entire herbal plant image.

Evaluation method

The model is designed for the recognition, classification, and description of the herbal plant. We use

it to evaluate the model's accuracy, precision, recall, and F-score.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FPFN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1\_Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Figure 1 model's Accuracy formulas

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively.

SCOPE AND LIMITATION OF THE STUDY

The main goal of this study is to classify recognized descriptions and administration of the herbal plant both in English and Amharic. Due To the lack of computational power to train so many herbal plant images this research might not include all the herbal plants but if the result is promising we will buy GPUs from Google and train so many plant images

SIGNIFICANT OF THE STUDY

There are some studies conducted about herbal plants including in Ethiopia but none of them use the Amharic Language as the output most of the farmers or even 50% of the population can not understand the English language

The significance of the study is summarized as follows:

1. Reduce the language barrier between the herbal plant harvester and the end user.
2. To reduce the side effects caused by chemical medicine.
3. To understand herbal plants, especially in Ethiopia.
4. To share herbal medicine preparation and administration wisdom.
5. To include fruit images in the training
6. To save the extinct plant by identifying the use of those plants
7. The data set collected and prepared by this research will help future researchers in the area

WORK PLAN

Table 1 Work Plan FOR RECOGNITION AND DESCRIPTION OF THE HERBS PLANT FOR PATIENT CARE

Time Schedule	Activities
15 October -30 December	Proposal preparation
1 January- 1 March	Literature review
22 January -18 February	Data collection
19 February -30 February	Data clarification and splitting
1 March - 15 April	Experimentation of a model
16 April -30 April	Make analytic by using the model
1 May – 10 May	Evaluation & result clarification
11 May -15 May	Documentation & finalizing the thesis
16 May - 30 May	Presentation

BUDGET PLAN

Table 2 Budget Plan FOR RECOGNITION AND DESCRIPTION OF THE HERBS PLANT FOR PATIENT CARE

No	Items	Unit	Quantity	Unite price (ETB)	Total Price (ETB)
1	A4 paper	Pack	3	500	1500
2	Pen	Pack	1	500	500
3	Printing	Page	2	1,500	3000
4	Binding	Module	12	50	600
5	Transportation	-	-	2,000	2000
6	Data Collection	person	15	1,000	15000
7	Flash Disk	32 GB	1	400	400
8	Mobile card	-	-	2,000	2000
				Total	25000

REFERENCES

Abera, B. (2014). Medicinal plants used in traditional medicine by Oromo people, Ghimbi District, Southwest Ethiopia. *Journal of Ethnobiology and Ethnomedicine*, *10*(1), 40. https://doi.org/10.1186/1746-4269-10-40

Aman, M., Asfaw, Z., & Dalle, G. (2020). *Medicinal Plants Used for Treating Human and Livestock Ailments in Tiyo District, Arsi Zone of Oromia, Ethiopia* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-20640/v1

Amuthalingeswaran, C., Sivakumar, M., Renuga, P., Alexpandi, S., Elamathi, J., & Hari, S. S. (2019). Identification of Medicinal Plant's and Their Usage by Using Deep Learning. *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 886–890. https://doi.org/10.1109/ICOEI.2019.8862765

Azadnia, R., Al-Amidi, M. M., Mohammadi, H., Cifci, M. A., Daryab, A., & Cavallo, E. (2022). An AI Based Approach for Medicinal Plant Identification Using Deep CNN Based on Global Average Pooling. *Agronomy*, *12*(11), 2723. https://doi.org/10.3390/agronomy12112723

Derso, M. (2021). *RECOGNITION OF ETHIOPIAN HERBACEOUS PLANTS USING A MACHINE LEARNING APPROACH* [Thesis]. http://ir.bdu.edu.et//handle/123456789/12399

Grinblat, G. L., Uzal, L. C., Larese, M. G., & Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture*, *127*, 418–424. https://doi.org/10.1016/j.compag.2016.07.003

Lee, S. H., Chan, C. S., Mayo, S. J., & Remagnino, P. (2017). How deep learning extracts and learns leaf features for plant classification. *Pattern Recognition*, *71*, 1–13. https://doi.org/10.1016/j.patcog.2017.05.015

Malik, O. A., Ismail, N., Hussein, B. R., & Yahya, U. (2022). Automated Real-Time Identification of Medicinal Plants Species in Natural Environment Using Deep Learning Models—A

Case Study from Borneo Region *Plants*, *11*(15), Article 15. https://doi.org/10.3390/plants11151952

Moges, A., & Moges, Y. (2020). Ethiopian Common Medicinal Plants: Their Parts and Uses in Traditional Medicine - Ecology and Quality Control. In A. Gonzalez, M. Rodriguez, & N. Gören Sağlam (Eds.), *Plant Science—Structure, Anatomy and Physiology in Plants Cultured in Vivo and in Vitro*. IntechOpen. https://doi.org/10.5772/intechopen.86202

Neto, J. C., Meyer, G. E., Jones, D. D., & Samal, A. K. (2006). Plant species identification using Elliptic Fourier leaf shape analysis. *Computers and Electronics in Agriculture*, *50*(2), 121–134. https://doi.org/10.1016/j.compag.2005.09.004

Putri, Y. A., Djamal, E. C., & Ilyas, R. (2021). Identification of Medicinal Plant Leaves Using Convolutional Neural Network. *Journal of Physics: Conference Series*, *1845*(1), 012026. https://doi.org/10.1088/1742-6596/1845/1/012026

Tafese Awulachew, M. (2021). Hand Book of Common Ethiopian Traditional Medicinal Plants: Their Parts and Uses for Human and Animal Treatments. *Journal of Diseases and Medicinal Plants*, *7*(3), 48. https://doi.org/10.11648/j.jdmp.20210703.11

Wang, B., Brown, D., Gao, Y., & Salle, J. L. (2013). Mobile plant leaf identification using smart-phones. *2013 IEEE International Conference on Image Processing*, 4417–4421. https://doi.org/10.1109/ICIP.2013.6738910

Wang, Z., Sun, X., Ma, Y., Zhang, H., Ma, Y., Xie, W., Wang, Z., & Zhang, Y. (2014). Plant recognition based on intersecting cortical model. *2014 International Joint Conference on Neural Networks (IJCNN)*, 975–980. https://doi.org/10.1109/IJCNN.2014.6889656

Widneh, M. A., Workneh, A. T., & Alemu, A. A. (2021). Medicinal Plant Parts Identification and Classification using Deep Learning Based on Multi Label Categories. *Ethiopian Journal of Sciences and Sustainable Development*, 96-108 Pages. https://doi.org/10.20372/EJSSDASTU:V8.I2.2021.380

Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y.-X., Chang, Y.-F., & Xiang, Q.-L. (2007). A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network. *2007 IEEE International Symposium on Signal Processing and Information Technology*, 11–16. https://doi.org/10.1109/ISSPIT.2007.4458016