INTERPRETABLE DEEP LEARNING APPROACHES FOR IDENTIFICATION AND CLASSIFICATION OF ETHIOPIAN INDIGENOUS MEDICINAL PLANT SPECIES

Mulugeta Adibaru Kiflie

A Dissertation Submitted to the department of Computer Science and Engineering, School of Electrical Engineering and Computing

Presented in Partial Fulfillment of the requirements for the Degree of Doctor of Philosophy in Computer Science and Engineering.

Office of Graduate Studies

Adama Science and Technology University

June 2024

Adama, Ethiopia

INTERPRETABLE DEEP LEARNING APPROACHES FOR IDENTIFICATION AND CLASSIFICATION OF ETHIOPIAN INDIGENOUS MEDICINAL PLANT SPECIES

Name of Candidate: Mulugeta Adibaru Kiflie

Major Supervisor: Prof. DP. Sharma

Co-Supervisor: Dr. Mesfin Abebe

Dissertation Submitted to the department of Computer Science and Engineering, School of Electrical Engineering and Computing

Presented in Partial Fulfillment of the requirements for the Degree of Doctor of Philosophy in Computer Science and Engineering.

Office of Graduate Studies

Adama Science and Technology University

June 2024

Adama, Ethiopia

Declaration

I declare that this dissertation entitled **"*Interpretable deep learning approaches for identification and classification of Ethiopian indigenous medicinal plants species*"** is my original work. That is, it has not been submitted for the award of any academic degree, diploma or certificate in any other university. All sources of materials used for this thesis have been duly acknowledged through appropriate citations.

Mulugeta Adibaru	_____	_____
Name of student	Signature	Date

Recommendation

We, the supervisors of this dissertation, hereby certify that I/we have read and revised the dissertation entitled "***Interpretable deep learning approaches for identification and classification of Ethiopian indigenous medicinal plants species" by*** Mulugeta Adibaru submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Science and Engineering. Therefore, we recommend the submission of the dissertation to the department for further review and defense.

Prof DP. Sharma	_____	_____
Major-advisor/Supervisor	Signature	Date
Dr. Mesfin Abebe	_____	_____
Co-advisor/Co-supervisor	Signature	Date

Approval Page

We, the undersigned, members of the Board of Examiners of the dissertation open defense By **Mulugeta Adibaru Kiflie** have read and evaluated the dissertation entitled "***Identification and Classification of Ethiopian Indigenous Medicinal Plants using Deep Learning and Interpretability***" And examined the candidate during open defense. This is, therefore, to certify that the Dissertation is accepted for partial fulfillment of the requirement of the degree of Doctor of Philosophy in Computer Science and Engineering

Chairperson	Signature	Date
Internal Examiner	Signature	Date
External Examiner 1	Signature	Date
External Examiner 2	Signature	Date

Finally, approval and acceptance of the dissertation is contingent upon submission of its final copy to the Office of Postgraduate Studies (OPGS) through the candidate's Department Graduate Council (DGC) and School Graduate Committee (SGC).

Department Head	Signature	Date
School Dean	Signature	Date
Office of Postgraduate Studies, Dean	Signature	Date

Dedication

1. **To my father, Adibaru Kiflie**

   I only say thank you for everything you did for me. Long live my father!

2. **To my mother, Asrebeb Shume**

   Emaye! Your tender love fuels my soul eternally. May your spirit flourish forever!

3. **To my beloved wife, Yordanos Asnake:**

   Jordi! Your unwavering compassion, endless patience, and constant encouragement propel me onward with boundless love. May our journey together be everlasting!!

4. **To my cherished daughters, Delina, and my beloved sons, Dagmawi and Nolawi:**

   You continually infuse my life with vitality, rejuvenating my spirit and liberating me from the monotony of everyday routines. My love for you is eternal and unwavering.

Acknowledgements

First and foremost, I offer my praises to the almighty God, Jesus Christ, and to his mother, Saint Virgin Mary.

I extend my heartfelt appreciation to my supervisor, Prof. DP Sharma for his unwavering support and insightful guidance throughout this research endeavor. His exceptional suggestions and constructive feedback have been instrumental in the development and refinement of this work. I am indebted to him for consistently pushing me to strive for excellence and for imparting invaluable practical research skills that will continue to shape my future endeavors.

I would like to express my sincere gratitude to my co-supervisor, Dr. Mesfin Abebe, Associate Professor of Computer Science and Engineering at Adama Science and Technology University, for his invaluable guidance and mentorship. His expertise and dedication have been indispensable in keeping my research up-to-date and equipping me with the necessary skills to navigate through its complexities. I am truly appreciative of his friendly and professional support.

Special thanks are also due to the Gullele Botanical Garden Institute for their unwavering support during the data collection phase, with heartfelt appreciation extended to PhD candidate Solomon for his assistance.

I express profound gratitude to all my friends and family for their unwavering motivation and support throughout this journey. The depth of my appreciation is sincere, recognizing that the connections we share extend far beyond mere friendship.

I extend my sincere gratitude to the entire ASTU community, particularly the department members of Computer Science and Engineering, for their continued support and encouragement during my academic pursuit.

Finally, but certainly not least, I extend my heartfelt appreciation to my wife, Yordanos Asnake, and my kids, Delina, Dagmawi, and Nolawi, for their unwavering understanding, love, and encouragement. Their unwavering support has been the driving force behind my determination, fueling my perseverance towards the success of this dissertation.

Mulugeta Adibaru Kiflie

Tables of Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

ADAM: Adaptive Movement Estimation

BIER: Bosting Independent Embedding Robustly

CAMS: Class Activation Map

CNN: Convolutional Neural Network

DBN: Deep Belief Network

DNN: Deep Neural Network

DRN: Deep Reinforcement Learning

DSN: Deep Stacking Network

DSRP: Design Science Research Process

EBI: Ethiopian Biodiversity Institute

EPHI: Ethiopian Public Health Institute

GBG: Gullele Botanical Garden

GPU: Graphical Processing Unit

GRAD-CAMS: Gradient-Weighted Class Activation Map

GRAD-CAMS++: Gradient-Weighted Class Activation Map Plus Plus

IUCN: International Union for Conservation of Nature

KD: Knowledge Distillation

LIME: Local Interpretable Model Agnostic Explanation

mAP: Mean Average Precision

MLP-BP: multi-layer perceptron back-propagation

PTL: Progressive Transfer Learning

Relu: Rectified Linear Unit

RFC: Random Forest Classifier

RMSEProp: Root Mean Squared Propagation

SGD: Stochastic Gradient Decent

SGD: Stochastic Gradient Descent (SGD)

SHAP: Shapley Additive Explanation

VGG: Visual Geometry Group

WHO: World Health Organization

WWF: World Wide Web

Abstract

Ethiopia, known for its rich biodiversity, holds significant therapeutic potential in its diverse array of medicinal plants. Traditional medicines serve as cost-effective and culturally accepted healthcare solutions, used by the population in regions with limited healthcare infrastructure. However, identifying and classifying these Ethiopian indigenous medicinal plants species is a complex and time-intensive task requiring specialized scientific expertise. The main objectives of this research work is to identify and classify Ethiopian indigenous medicinal plants using deep learning and interpretability. The study started using a systematic literature review aimed at investigating deep learning approaches to identifying and classifying medicinal plants. Subsequently, various deep learning approaches were employed to develop an efficient model through transfer learning and ensemble learning for identifying and classifying medicinal plants species. To tackle the interpretability issues of deep learning, interpretable deep learning models were designed using a multiple teacher-student approach with knowledge distillation concepts. It was done to present an integrated framework for identifying and classifying indigenous medicinal plants species using interpretable deep learning approaches. In the experimental phase, a dataset containing 12,438 labeled leaf images was prepared. Employing efficient pretrained models such as VGG19, VGG16, Xception, and InceptionNetV3, was adopted to enhance the model's performance. In addition an ensemble EfficientNetB0, EfficientNetB2, and EfficientNetB4 are applied for the identification of parts and uses of Ethiopian indigenous medicinal plants. In the interpretable deep learning approach, a novel, distilled student model was designed using a collaborative teacher-student framework. The systematic review revealed disparities in global research due to resource and dataset variations, with most researchers uses private datasets and employing leaf shapes, transfer learning, and pre-trained models. The study effectively addressed the stated challenges and achieving a commendable accuracy of 95% through fine-tuning. The distilled student model attained an exceptional accuracy of 99.83%, facilitated by knowledge transfer metrics like cosine similarity and MSE. Integrating interpretability techniques such as LIME enhances model transparency and reliability, bridging traditional and modern medicine realms. Addressing the lack of globally accessible datasets for medicinal plants is essential to mitigating disparities in the field.

Keywords: deep learning, ensemble Learning, transfer learning, interpretable deep learning, distilled model, knowledge distillation.

CHAPTER ONE

INTRODUCTION

1.1. Background of the Study

Medicinal plants are crucial for traditional medicine and new drugs, especially in Ethiopia, where conservation preserves vital knowledge. In the past decade, deep learning has emerged as a dominant force in various fields, including speech recognition, face recognition, identification and classification in the various sectors such as the traditional health care sector, and self-driving cars. This study and the dissertation investigate the exploration of deep learning's potential applications within the domain of medicinal plants species. As technology continues to advance, the incorporation of deep learning techniques holds promise for revolutionizing the understanding and exploitation of Ethiopian medicinal plants. By leveraging the power of artificial intelligence, the research aims to contribute to the identification and classification of medicinal plants species. The widespread applications of deep learning in various scientific domains underscore its transformative capabilities, prompting an investigation into its adaptability and effectiveness in addressing challenges related to Ethiopian medicinal plants. This study uses an interpretable deep learning approach for the identification and classification of Ethiopian indigenous medicinal plants species.

1.2. Overview of the study

Plants are undeniably valuable sources of medicines, foods, spices, clothing, shelters, fertilizers, and, most importantly, elements in climate-change-regulating mechanisms (Abera, 2014). From ancient times to the present, plants have been employed as a source of medicine by all nations. The usage of medicinal plants is expanding quickly around the world due to the rising demand for herbal medicines, natural health products, and secondary metabolites of medicinal plants. (Chen *et al.*, 2016). Individuals who exclusively depend on allopathic medicine are probably somewhat reliant on medicinal plants, considering that 20-25% of prescribed medications are derived from plants (Smith-Hall *et al.*, 2012).

Traditional herbal medicines are also becoming incredibly popular in developed countries. For instance, traditional herbal medicine in China accounts for 30–50% of drug consumption (Aziz *et al.*, 2018). Meanwhile, traditional herbal medicines are the first choice for 60% of children in

countries like Nigeria, Ghana, Zambia, and Mali who are suffering from severe malaria (Aziz *et al.*, 2018).

According to a conservative estimate, the current loss of plants species is 100 to 1,000 times greater than the expected natural extinction rate, resulting in the loss of at least one potential major drug every two years (Bhujun *et al.*, 2017). Globally, between 50,000 and 80,000 flowering plants species are used for medicinal purposes, according to the International Union for Conservation of Nature (IUCN) and the World Wildlife Fund (WWF). Approximately 15,000 of these species are threatened with extinction due to overharvesting and habitat destruction(Kumar *et al.*, 2021). Additionally, with the growing human population and increased plants consumption, 20% of wild plants resources have already been depleted (Chen *et al.*, 2016).

Ethiopia is estimated to have 6500-7000 flora species, with 12-19% of indigenous medicinal plants (Admasu &Yohannes, 2019). Approximately 80% of the Ethiopian population uses traditional medicines because healers and local pharmacopeias are culturally acceptable. Traditional medicines are relatively inexpensive, and modern drug supply and approaches have numerous challenges, especially in rural areas where diverse populations reside(Nigussie Amsalu et al., 2018). Two of the world's 34 biodiversity hotspots, the Horn of Africa and the eastern afro-tropical sub-region, are located in Ethiopia, one of the top 25 biodiversity-rich nations on earth (Amenu et al., 2016). Scientific recognition and usage of any medicinal plants needs globalization and this is also truly anticipated for Ethiopian indigenous medicinal plants. Research and community-based conservation techniques are required to be discovered for mitigating the issues associated with the loss of Ethiopian traditional medicinal plants and their habitats, as well as for conserving and helping to maintain their survival. Therefore, immediate research efforts are necessary to identify and classify Ethiopian indigenous medicinal plants to meet these pressing requirements.

Botanists have long used traditional and experience-based methods to identify various species of medicinal plants. However, visually and manually distinguishing medicinal plants from other similar plants can be extremely difficult and time-consuming for inexperienced people(Hridoy *et al.*, 2022). Automatic Ethiopian medicinal plants identification is critical for making people aware of the benefits of these Ethiopian medicinal plants for long-term conservation before they become extinct. As a result, before proposing any practical control schemes, accurate identification and classification of Ethiopian indigenous medicinal plants is required.

The precision and accuracy of the identification of unknown indigenous medicinal plants are heavily reliant on the inherent knowledge of a skilled botanist or traditional practitioner. But it is a time-consuming task and difficult and the human expert cannot be replicated or cloned easily like a scientific system for identification with the scalability of the coverage area. The non-experts who have little or no knowledge of common botanical terms can't be efficient identifiers. Hence, an efficient way of identifying and classifying of Ethiopian indigenous medicinal plants is essential for supporting both botanist experts and people with non-botanical knowledge to easily identify these Ethiopian indigenous medicinal plants for future use.

Advances in technology-enabled support systems and techniques such as computer vision, image processing, and deep learning are required to be explored and examined in identification and classification of indigenous medicinal plants so as to help bridge the gap between the lack of expert taxonomists, traditional practitioner and non-botanical experts for proper conservation of endangered Ethiopian indigenous medicinal plants species in appropriate techniques to prevent and reduce the risks of their extinction.

In this work, deep learning approach is proposed to be used for the identification and classifications problems of Ethiopian indigenous medicinal plants. Currently, deep learning is a popular approach for classification, recognition, detection and identification problems with a reasonable performance in terms of its speed and accuracy. For extracting useful features deep learning algorithm make use of extraordinary architectures like Convolutional Neural Network, Deep Belief Network, Deep Neural Network and etc. (Tan *et al.*, 2020). Deep learning makes use of massive neural networks with interconnected neurons that can adjust their hyper-parameters whenever fresh new data is received. When using a deep learning approach, a computer or other device is capable of learning things on its own without explicit programming. This technology enables computer systems to learn new things on their own without direct programming from humans. This study makes use of the concept of transfer learning, which is the enhancement of deep learning in new classification or prediction tasks through the transferable knowledge that has already been learned in one or more tasks and uses it to enhance learning in a target task that is related to the original task(Duong Trung *et al.*, 2019).

In recent years, deep learning models have exhibited remarkable success in both industrial and academic domains, spanning a wide spectrum of applications, including computer

vision(Alzubaidi *et al.*, 2021) and natural language processing (Landolt *et al.*, 2021). Insufficient training data is a common challenge in effectively training deep learning models for many applications(Alzubaidi *et al.*, 2021). Deep learning also faces challenges such as the need for precisely collected labelled data, the absence of standardized dataset protocols, and a very limited availability of country-specific datasets. Deep learning also grapples with the inherent black-box nature and interpretability issues(Ekanayake *et al.*, 2022). Hence, there is a pressing need to create compact networks that exhibit strong generalization capabilities without an overwhelming reliance on extensive datasets. Moreover, addressing the black-box nature requires the integration of interpretable deep learning techniques. Additionally, it's crucial to acknowledge that most deep learning models come with high computational demands, rendering them unsuitable for resource-constrained devices like mobile phones and embedded systems(Polson &Sokolov, 2020).

To address the challenges posed the black-box nature of deep learning, interpretable deep learning techniques is employed. Interpretable deep learning refers to the practice of designing and training deep neural networks in a way that allows humans to understand, explain, and interpret the model's decisions and predictions(Wang, 2020) . Interpretable deep learning is essential in applications where model transparency and insight into decision-making are crucial, such as healthcare, finance, and legal systems (Preuer *et al.*, 2019).

This research work introduces a light weight interpretable deep learning model to address the identification and classification challenges of Ethiopian indigenous medicinal plants by employing EIMPS dataset.

1.3. Statement of the Problem

Numerous studies reveal that traditional indigenous medicinal plants and related knowledge in Ethiopia face a significant threat, primarily due to the exclusive interest and involvement of older generations in their utilization and conservation(Giday *et al.*, 2003). The persistent challenges of deforestation, environmental degradation, and acculturation pose serious risks to these valuable resources. As a consequence, the reliance on plants-based indigenous medical practices, crucial for primary healthcare services in Ethiopia, may weaken. The continuous loss of indigenous medicinal plants and associated knowledge could jeopardize the potential future development of modern herbal drugs. Addressing this issue requires urgent research initiatives focusing on the identification and classification of Ethiopian indigenous medicinal plants. Scientifically,

challenges arise in the form of time-consuming and labor-intensive efforts by botanist experts with inherent knowledge in botany and plants systematics. Additionally, natural, ecological, and cultural challenges must be considered.

Rigorous literature review results reveal that most researchers have created custom datasets for medicinal plants species tailored to their specific study areas or countries(Mulugeta *et al.*, 2024). Consequently, these solutions cannot be directly applied. This emphasizes the significance of feature-based intelligent customization and contextualization to attain greater precision and accuracy in the classification of medicinal plants species. Additionally, a contextualized customization of existing solutions with different datasets may reveal new knowledge for the identification and classification of medicinal plants species that are endemic to a specific country. The lack of well-structured availability of Ethiopian indigenous medicinal plants species datasets is also a research gap in the study domain. Furthermore, it is apparent that different deep learning approaches have limited scope to experimentally explore and address the challenges and issues related to classifying and identifying medicinal plants species. This is because deep learning approaches provide classification and recognition without sufficient explanation. Also, they are considered untrustworthy due to their black-box nature. Developing interpretable deep learning models that can explain their predictions is an essential area of research, particularly in fields such as medicinal plants, where model decisions can have significant consequences (Wang, 2020).

Hence, it is imperative to tackle the above-mentioned challenges through the development of interpretable deep learning models. This dissertation aims to address these challenges by developing interpretable deep learning models, preparing an Ethiopian indigenous medicinal plants dataset.

1.4.Research Questions

The following research questions are formulated to investigate and resolve the stated problems:

1. What are the existing state-of-the-art efforts available for identification and classification indigenous medicinal plants?
2. Which deep learning approaches are effective for feature extraction to achieve precise identification and classification of Ethiopian indigenous medicinal plants?
3. How can a lightweight interpretable deep learning model be designed to identify and classify Ethiopian indigenous medicinal plants?

1.5.Objectives of the Study

1.4.1. General Objective

The main objective of this research is to identify and classify Ethiopian indigenous medicinal plants using deep learning and interpretability.

1.4.2. Specific Objectives

The following are the specific objectives to achieve the general objective of the study

- To analyze the existing state-of-the-art solutions available for identification and classification indigenous medicinal plants.
- To select appropriate deep learning models for extracting features from the Ethiopian medicinal plants species dataset to effectively identify and classify Ethiopian indigenous medicinal plants species.
- To build an interpretable deep learning model to accurately identify and classify Ethiopian indigenous medicinal plants species on resource-constrained devices

1.5. Contributions

The key contributions of this research are as follows: (i) conducting a detailed review to comprehensively identify critical gaps and challenges within the domain, (ii) preparing the Ethiopian indigenous medicinal plants dataset and making it accessible to researchers through publicly accessible repositories with DOI, (iii) performing a comparative analysis of pretrained model to assess their applicability for the Ethiopian medicinal plants dataset (iv) identifying the parts and uses of Ethiopian indigenous medicinal plants, and (v) building interpretable deep learning model for Ethiopian indigenous medicinal plants identification and classification.

1.6. Significance of the Study

People believe in their community and indigenous practices. Due to their closeness to medicinal plants and inaccessible health facilities, people still rely on indigenous traditional knowledge of plants. In Ethiopia, like many other developing countries, a high percentage of the population depends on traditional medicine for primary healthcare. However, medicinal plants species are on the verge of extinction due to the increase in population, overexploitation, overharvesting, increasing market demand, deforestation, industrialization, road and other constructions, agricultural farm expansions, and biodiversity losses.

Identification and classification of medicinal plants is one of the strategies required for the conservation and sustainable utilization of indigenous medicinal plants species. Medicinal plants

species identification and classification has significant benefits for a wide range of stakeholders, including forestry services, botanists, taxonomists, physicians, pharmaceutical laboratories, traditional medicine practitioners, farmers, environmentalists, educators, government organizations, chemical engineers, chemists, and the general public. Because medicinal plants species identification and classification can speed up the identification and classification process and reduce the time required for identifying botanists, all the stated stakeholders can participate in the recognition strategies. A modern medicinal plants identification and classification strategy is also used as a baseline strategy for involving both the communities and government to preserve indigenous plants species. The technology intervention and judicious application in designing a better model the identification and classification of indigenous knowledge is the need of the hour.

Interpretable deep learning is also valuable for transparent AI, especially in the context of Ethiopian medicinal plants species identification. It ensures transparency in decision-making, aligning with ethical practices and regulations for the preservation of Ethiopian indigenous medicinal plants species. It fosters fairness, accountability, and user acceptance among traditional medicine practitioners, pharmaceutical industries, and botanists. Interpreted models generally aid error diagnosis and raise model performance, introducing confidence for the conservation of endangered Ethiopian medicinal plants species.

This work also integrates knowledge distillation, a vital aspect of deep learning, enabling knowledge transfer from complex teacher models to simpler student models. Knowledge distillation wrappings deep learning models for resource efficiency, improves generalization on limited data, and enables transfer learning. It enhances interpretability, supports ensemble learning for accuracy, and ensures adaptability in related domains. Knowledge distillation facilitates accelerated inference, stabilizes training, and contributes to lifelong learning, offering versatility in model development across various applications including Ethiopian medicinal plants species identification and classification.

In General, An accurate identification and classification of Ethiopian indigenous medicinal plants species provides significant benefits to both industry, society and indigenous community and practitioners Traditional medicine healers will benefit from the accurate identification and classification of Ethiopian indigenous medicinal plants species for collection and preservation of the herbal medicines, while the modern pharmaceutical industry can obtain insights into Ethiopian

indigenous plants species required for numerous drug discoveries. Furthermore, this work is crucial for the conservation of endangered medicinal plants species and contributes to current ethnobotanical research, expanding scientific knowledge in the field. Advances in machine learning, computer vision, and deep learning can present an opportunity to broaden and improve the practice of precise medicinal plants protection, conservation, and market expansion of deep learning applications in precision agriculture and other allied fields.

1.6. Delimitations and Limitations of the Study

This dissertation seeks to explore, apply, and advance the deep learning approaches by addressing fundamental weaknesses, and deficiencies as research gaps in the existing state-of-the-art solutions and techniques using a systematic literature review. The main focus of this research lies in addressing the issues in the existing solutions with special focus on interpretable deep learning approaches. The dissertation's scope is confined to the exploration and application of interpretable deep learning methods, emphasizing the design of a lightweight interpretable deep learning models for the identification and classification of Ethiopian indigenous medicinal plants species on resource constrained devices involving the preparation of a custom dataset.

For this work, the data is collected from Gullele Botanical Garden, which may not represent all the indigenous medicinal plants in Ethiopia. The dataset's coverage may be limited due to natural, ecological, and cultural variations across the country, potentially excluding some diverse plants species. Moreover, accurately localizing medicinal plants for specific ethnic groups necessitates understanding cultural and ethnicity-specific uses of indigenous medicinal plants parts. The variability in traditional healing practices across different societal groups further complicates accurate identification and classification within cultural contexts, aspects that are not addressed in this research.

Operational Definitions

Indigenous Medicinal Plants Species: Native plants found in multiple regions, traditionally used for medicine in Ethiopia. These species naturally occur in multiple regions or countries.

Endemic Medicinal Plants Species: Unique to Ethiopia, these plants are found nowhere else and used medicinally.

1.4. Organization of the Dissertation

The remaining sections of this dissertation are structured as follows:

Chapter 1- Introduction: This section serves as an introduction to the research topic, outlining its significance and relevance. It presents the statement of the problem, research questions, objectives, significance and scope providing a foundation roadmap for the dissertation.

Chapter 2- Review of Literature and Related Works: This chapter emphasizes on the comprehensive review of existing literature related to medicinal plants species, deep learning, knowledge distillation, and interpretable deep learning. Finally, it synthesizes the relevant research efforts in the domain focusing on the identification and classification of medicinal plants species, to identify and compile the research gaps.

Chapter 3- Research Methodology: This chapter specifies the identification of the suitable research design, Methods and tools employed for modeling and experimentation in the study. The data collection, and preprocessing processes are also decided in this chapter.

Chapter 4- Experimental Results and Discussion: This chapter conducted a systematic review of literature confined to the deep learning approaches applied to the identification and classification of medicinal plants species. It evaluated and summarized the strengths and limitations (gaps) of existing approaches, identifying challenges and opportunities for advancement in the field. It assesses the performance of various pre-trained deep learning models on the custom dataset, presenting and interpreting the experimental results. Additionally, this chapter tries to investigate ensemble learning methods where it examines their effectiveness in identifying different parts and uses of Ethiopian indigenous medicinal plants species. Further the rationale behind ensemble learning and presented in experimental findings to analyze its efficacy in the identification of parts and uses of Ethiopian indigenous medicinal plants species are discussed. Furthermore, this chapter introduces a lightweight interpretable deep learning using knowledge distillation, dedicated to the development of a model tailored to identify and classify the Ethiopian indigenous medicinal plants species.

Chapter 5- Conclusions, and Future Works: As a Final chapter the key findings, contributions and future works are covered in this chapter.

CHAPTER TWO

LITERATURE REVIEW AND RELATED WORKS

2.1. General Overview

In ancient times plants have long been used as a source of both preventive and therapeutic traditional medicine preparations for livestock and humans. Numerous therapeutic plants have been used by Chinese people in the era of 5000 to 4000 BC, and by Babylonians, Hebrews, Egyptians, and Syrians before 1600 BC(El Sheikha, 2017). Most of the traditional medicines used by different countries are linked with traditional medicinal healers and they use an indigenous knowledge system that is mostly passed orally from one generation to the next generation. Unlike scientific medicines, traditional medicines are practiced by traditional healers using their culture, religion, or indigenous knowledge(Yaniv, 2014).

In Ethiopia, over 80% of the population continues to rely on medicinal plants to address various health challenges(WHO, 2019). In the long history of using indigenous knowledge of traditional medicine the alignments and remedies of medicinal plants are confirmed by referencing the cultures and medical manuscripts related to religion in the country(Kibebew, 2001). Currently, Ethiopian medicinal plants are under threat of total miss due to natural and other human made factors. Promoting indigenous knowledge associated with plants is also required. So, identifying and recognizing such useful but threatened plants are required urgently for conserving and utilizing medicinal plants in Ethiopia(Lulekal *et al.*, 2008). Flowers, leaves shape and seeds are typically used for identifying plants and their species in botany. In order to study the plants, botanists use change in leaf characteristics as a relative tools due to leaf characteristics of the deciduous trees, annual plants and or availability of leaves throughout the year for observation and analysis(Cope *et al.*, 2012).

Due to this reason, images of leaves have been used by computer vision scientists and researchers for recognizing, classifying and identifying medicinal and other plants species (Hall *et al.*, 2015;Kalyoncu *et al.*, 2015;Kumar *et al.*, 2012). Most of the plants can't be recognized or identified by farmers, traditional healers, chemists, pharmacists and even by experts with the given scientific names(Ferentinos, 2018). Hence, Artificial Engineering, Machine learning and

Computer vision algorithms and techniques are currently employed and used to solve this complex problems(Chan *et al.*, 2015;Kamilaris *et al.*, 2018). Computer Vision algorithm mainly uses venation(Charters *et al.*, 2014;Larese *et al.*, 2014), textures(Cope *et al.*, 2010;Naresh &HS, 2016;Tang *et al.*, 2015), and shape(Mouine *et al.*, 2012;Neto *et al.*, 2006;Xiao *et al.*, 2010) characters to identify leaves of various plants species.

This section presents concepts and literature pertaining to the identification and classification of medicinal plants, specifically exploring deep learning approaches. Various approaches within deep learning, such as conventional deep learning, transfer learning, ensemble learners, and interpretable deep learning, are examined. The review encompasses key concepts and literature relevant to the application of deep learning methods in the identification and classification of medicinal plants.

2.2. Ethiopian Indigenous Medicinal Plants

Medicinal plants play a crucial role in the healthcare systems for both humans and animals in numerous countries. This is especially true for Ethiopia, one of the ancient nations in the Horn of Africa. The country boasts a rich floral diversity with over 6,500 to 7,500 vascular plants species. Consequently, there has been a growing and sustained interest in ethnobotanical research in the region (Yirgu *et al.*, 2019). Ethno medicinal knowledge, which arises from the interaction between a culture and its local biophysical environment or available plants, is diverse and can be specific to certain ecosystems and ethnic communities(Teka *et al.*, 2020). Factors such as social, ecological, and cultural backgrounds including religious and linguistic elements as well as ancestral inheritance, shape the traditional herbal knowledge within a community. As a result, herbal knowledge significantly varies across different communities, geographic settings, and ethnic groups(Teka *et al.*, 2020). The impact of cultural background and ancestral inheritance is evident in the differing perceptions and plants use preferences among people living in the same geographical or ecological area and encountering similar environmental conditions(Junsongduang *et al.*, 2014). In Ethiopia, the practice of using medicinal plants to treat and heal diseases has a long history, making traditional medicine a vital component of the healthcare system(Giday *et al.*, 2013). Traditional medicine is a primary healthcare system for over 80% of the Ethiopian population(Demie *et al.*, 2018). Its popularity is largely attributed to its low cost and the lack of accessible modern healthcare systems (Feyssa *et al.*, 2011). The burden of these diseases in the

rural and urban populations is different that mainly depends on their sociodemographic conditions, lifestyle, health risks, etc (Chen,Orom, *et al.*, 2019).

The high cost of drugs and modern healthcare has led the people of Ethiopia to heavily rely on traditional medicinal systems. This practice is widespread across many regions of Ethiopia, including Gullele Botanical Garden (GBG) in Addis Ababa. Addis Ababa, the capital city, is divided into ten sub-cities, with Gullele being one of them. Gullele spans an area of nearly 30 square kilometers and has a population density of approximately 9,500 people per square kilometer. It is home to various ethnic groups, with the major ones being Amhara, Oromo, Gamo, and Guragie. Botanical gardens play a crucial role in the conservation of diverse plants species. Globally, botanical gardens are known to help conserve about 41% of threatened species(Mounce *et al.*, 2017).

Gullele Botanical Garden was established in 2010 as a joint venture between the Addis Ababa government and the university. The garden spans 705 hectares and supports Sustainable Development Goals (SDGs) 6, 7, 13, and 15(Seta &Belay, 2022). Additionally, it has been accredited by Botanic Gardens Conservation International (BGCI) until 2025. The garden comprises over 90% of socioeconomic and protected forest areas. Initially, an assessment of the GBG landscape identified 223 plants species from nearly 66 families, a number that has since grown to over 1,200 species through collection and in situ management techniques. This increase in flora is primarily due to propagation and plantsing, reforestation, natural regeneration of endemic plants, and the sustainable use of medicinal plants by traditional healers. GBG enhances the greenery of Addis Ababa, contributing to carbon storage, habitat conservation, and soil erosion prevention. The widespread use of traditional medicine in Ethiopia is driven by its low cost and high cultural acceptance. In general, Gullele Botanical Garden harbors nearly 1,600 plants. Among these, there are around 64 endemic species, 189 exotic species, 900 indigenous species, and about 65 critically endangered species (Seta &Belay, 2022;Woldegerima *et al.*, 2017).

According to the garden authorities, traditional healers in Ethiopia have been providing various medical services, treatments, and remedies for a long time. Consequently, the residents around the garden and in Gullele sub-city heavily depend on the plants medicines offered by these healers. The garden authorities work closely with traditional healers or traditional physicians to collect, protect, and propagate medicinal plants species in GBG. Within GBG, there is a medicinal garden

dedicated to cultivating and propagating medicinally important plants that are overharvested. The medicinal garden and forest section of GBG contain a total of 166 medicinal plants, which are sustainably used by certified traditional healers in Gullele and Addis Ababa.

For bioprospecting and optimal use of medicinal plants, proper survey and documentation are essential (Awas *et al.*, 2010). Many traditional healers from various ethnic groups with different cultures and traditions rely on the medicinal plants of GBG to practice traditional medicine. However, there is no systematic documentation of the traditional medicinal knowledge used by these healers. Although GBG plays a vital role in transferring traditional medicinal knowledge from one generation to another, much of this indigenous knowledge is lost or diluted due to the oral transmission methods. Therefore, GBG has been selected for the current study, and the help of GBG authorities has been enlisted to identify traditional healers from different ethnic groups. Identifying the endemic plants most used by the healers will help the GBG authorities prioritize their cultivation and propagation efforts. Thus, the identification and classification of endangered indigenous medicinal plants species is a critical issue in Ethiopia due to the diverse cultural practices and the risk of knowledge loss.

2.3. Deep Learning for image Identification and Classification

Medicinal plants identification and classification is unsolved problems and challenging activities in computer vision in spite of many trails even using complicated and challenged machine vision techniques and algorithms because plants in nature can be characterized by colors and shapes resembles others(Szegedy *et al.*, 2017;Szegedy *et al.*, 2016). In recent times, a deep learning algorithm or methods have been used for extracting features of the plants automatically which is recently improved explicit feature selections tasks. Hence, the concepts for end to end learning is introduced in deep learning methods by referring trainable extractors of plants leaf features and trainable classifiers of plants species for identification and recognition problems(Szegedy *et al.*, 2017;Szegedy *et al.*, 2016).

In deep learning algorithm, for plants leaf classification and identification problems leaf features representation is a critical component. In the reviewed articles most of the existing algorithms and methods follow hand crafted feature extraction (Charters *et al.*, 2014;Naresh &HS, 2016;Neto *et al.*, 2006) and feature extraction using deep learning(Chan *et al.*, 2015;Lee *et al.*, 2017) approaches to the representations of extracted features of plants leave images for the classification problems of plants species. Practically, a machine/computer vision requires the knowledge and

abilities of experts to encode morphological features which is identified or predefined by botanists to design hand crafted feature extraction(Lee *et al.*, 2017).Yet, currently a deep learning algorithm becomes popular to extract features of plants automatically for the classification and identification problems.

Building a convolutional neural network (CNN) from scratch requires a substantial amount of data and computing resources, which can be both a pro and a con. On one hand, custom-built models allow for fine-tuning every aspect of the network architecture to suit specific tasks. However, due to the demanding nature of training CNNs from scratch, an extensive dataset is typically needed to achieve satisfactory performance. Alternatively, using pre-trained models like VGG-19, VGG-16, InceptionV3, and Xception, offers a promising approach. These models have been trained on large datasets such as ImageNet, where they have learned to extract meaningful features from images. Transfer learning allows us to take advantage of these learned features and adapt them to our specific tasks with smaller datasets. In this approach, we freeze the weights of the pre-trained model (except for the last few layers) and only train the fully connected classifier layers using our dataset. The explanation of some of the popular pre-trained model used in computer vision tasks are stated as follow.

2.3.1. VGG-16 (Visual Geometry Group-16 Layers)

VGG-16, developed by K. Simonyan and A. Zisserman at the University of Oxford(Simonyan &Zisserman, 2014), is a renowned convolutional neural network architecture widely recognized for its effectiveness in image classification tasks. The network comprises 16 layers, including 13 convolutional layers followed by ReLU activation functions and 3 fully connected layers at the end. Each convolutional layer typically employs 3x3 filters to extract features from input images, while max-pooling layers help reduce spatial dimensions and generate feature maps. With approximately 138 million parameters, VGG-16 is computationally intensive but excels in learning hierarchical representations of images. Trained on the ImageNet dataset, which includes over 1.2 million images across 1,000 categories, VGG-16 achieved notable performance gains in the ImageNet Large Scale Visual Recognition Challenge (Russakovsky *et al.*, 2015;Szegedy *et al.*, 2014). Despite newer architectures surpassing it in specific metrics, VGG-16 remains influential as a benchmark and foundation in deep learning research and applications, demonstrating its

enduring impact on the field of computer vision. The figure below provides a detailed illustration of the layers in VGG16:

Figure 1 Schematic Architectures of VGG16

2.3.2. VGG19 (Visual Geometry Group-19 Layers)

The VGG-19 model, proposed by K. Simonyan and A. Zisserman from the University of Oxford(Simonyan &Zisserman, 2014), is particularly renowned for its architecture's ability to achieve high accuracies on large-scale image processing tasks like ImageNet. VGG-19 consists of 19 layers, including convolutional layers, fully connected layers, max pooling, and dropout layers. It has approximately 143 million parameters, all learned from the ImageNet dataset, which contains 1.2 million images across 1,000 object categories. By using VGG-19 as a feature extractor in transfer learning to capture relevant features from image. This approach is efficient in scenarios where datasets are limited, as it maximizes the use of pre-existing learned features while minimizing the computational burden of training a CNN from scratch. The below figure illustrated the details of VGG19 pretrained model.

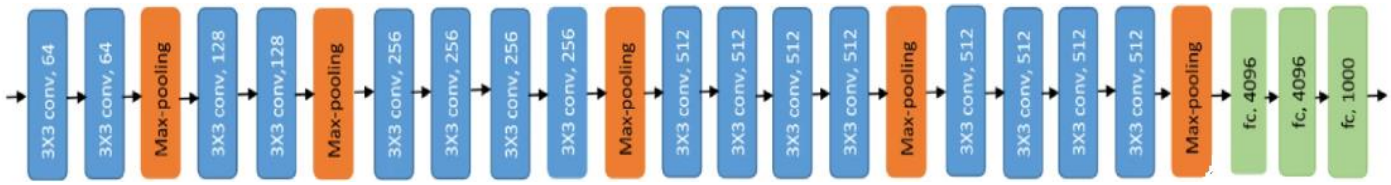


Figure 2 Schematic Architecture of VGG-19 Pretrained model

2.3.3. Inception-V3

Inception was originally proposed by Szegedy et al., featuring 42 layers(Szegedy *et al.*, 2015). The third generation of this model, Inception-V3, was developed by Google Brain and consists of 159 layers(Szegedy *et al.*, 2016). Inception-V3 is composed of three main parts: convolution layers,

Inception modules, and classifiers (as shown in Figure 3). The Inception module, based on the Network-In-Network concept, performs multiple convolution layers in parallel to expand the network's capacity, and then concatenates the results from each branch.(Lin *et al.*, 2019) . This design allows Inception-V3 to achieve superior performance in object recognition tasks compared to the original model, excelling in classifications.

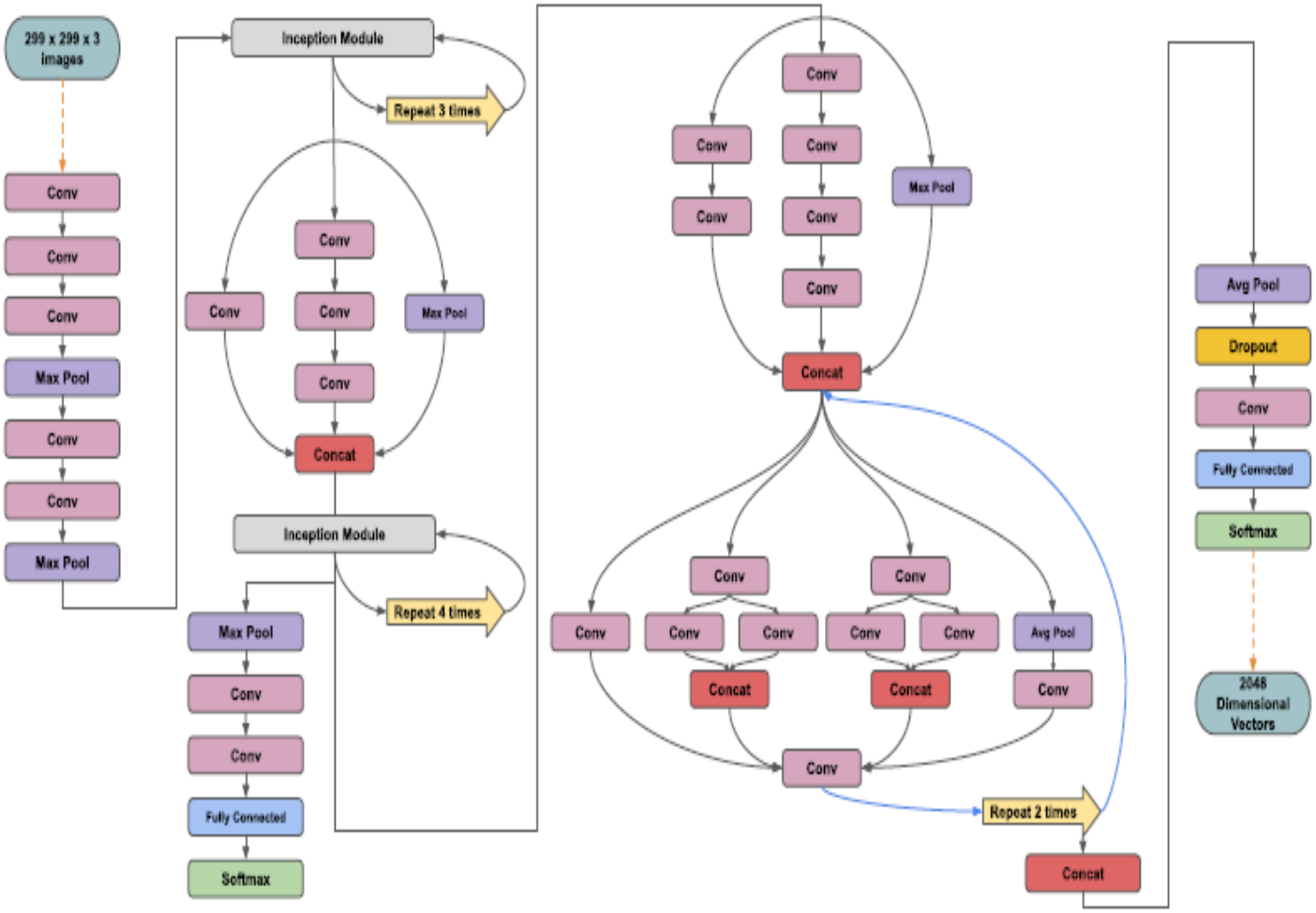


Figure 3 Schematic Architectures of Inception V3

2.3.4. Xception

Xception, derived from Inception-V3, employs a linear stack of depth-wise separable convolution layers combined with residual connections to enhance time and space efficiency, as illustrated in Figure 4 below (Chollet, 2017). In Xception, depth-wise separable convolutions decouple the

learning of channel-wise and spatial features. Additionally, the inclusion of residual connections addresses the vanishing gradient problem and improves representational learning.

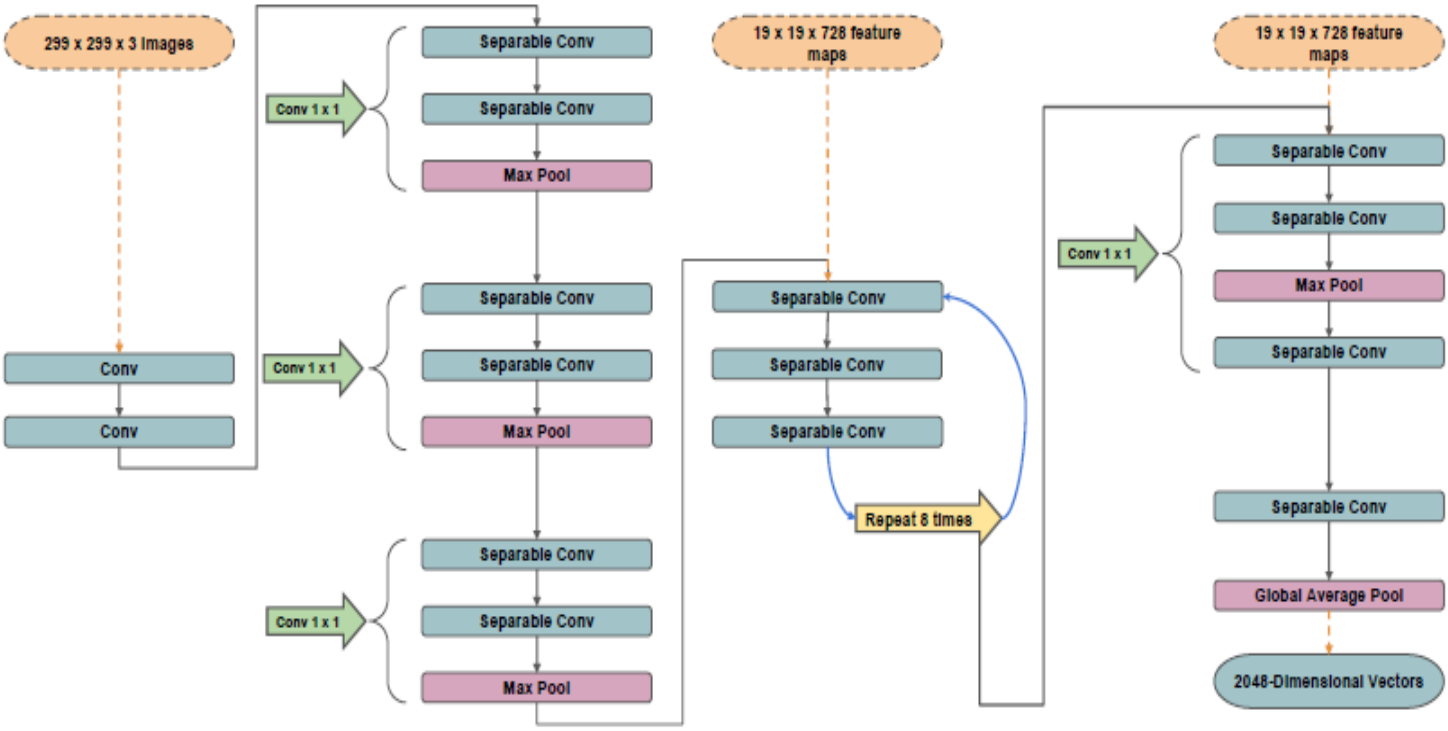


Figure 4 Schematic Architecture of Xception

2.4. Ensemble Learning

Ensemble learning methods harness the capabilities of multiple deep learning algorithms to generate predictive outcomes by leveraging features extracted through diverse predictions on data. These methods integrate results using various voting mechanisms, surpassing the performance achieved by individual algorithms in isolation. The collaboration of different models enhances the robustness and generalization of the ensemble, leading to improved overall performance compared to any single constituent algorithm(Zhou, 2012). The fundamental concept involves joining the unique strengths of numerous models to offset their respective weaknesses and enhance overall generalization for the classification and identification purposes. These methods have been pivotal in addressing challenges related to overfitting and enhancing the robustness of models across diverse datasets. Broadly categorized, these ensemble techniques encompass a range of approaches

aimed at combining multiple individual deep learning models to produce more accurate and reliable predictions.

Bagging (Breiman, 1996;Lingappa *et al.*, 2023;Nakach *et al.*, 2023;Smaida &Yaroshchak, 2020;Zhang *et al.*, 2021), also referred to as bootstrap aggregating, stands as a common technique for constructing ensemble-based algorithms, aimed at enhancing the performance of ensemble classifiers. The core concept behind bagging involves generating a sequence of independent observations that match the size and distribution of the original dataset. Through this process, an ensemble predictor is formed with superior performance compared to a single predictor trained on the original data. Bagging entails two key steps: firstly, the creation of bagging samples and their subsequent utilization by the base models, and secondly, the formulation of a strategy for aggregating the predictions generated by multiple predictors. These bagging samples can be generated with or without replacement. As for the combination of predictions from the base predictors, majority voting is commonly employed for classification tasks, whereas an averaging strategy is typically utilized in regression scenarios to produce the ensemble output.

The boosting technique(Kunapuli, 2023;Mohammed *et al.*, 2023) is utilized within ensemble models to elevate a weak learning model into one with enhanced generalization capabilities. Methods like AdaBoost (Mohammed *et al.*, 2023;Sharma *et al.*, 2023) and Gradient Boosting(Chung &Teo, 2023;Guillen *et al.*, 2023) have found applications across various domains. AdaBoost employs a greedy approach to minimize a convex surrogate function, which is upper bounded by misclassification loss through iterative augmentation. At each stage, the current model is augmented with an appropriately weighted predictor. AdaBoost effectively constructs an ensemble classifier by leveraging misclassified samples at each iteration, minimizing the exponential loss function. On the other hand, Gradient Boosting extends this framework to arbitrary differential loss functions. Originally proposed to enhance the performance of classification trees, boosting, also known as forward stage wise additive modeling, has more recently been integrated into deep learning models to further enhance their effectiveness. For instance, the Boosted Deep Belief Network (DBN)(Lasri *et al.*, 2023;Srivastav *et al.*, 2024) for facial expression recognition combines boosting techniques with multiple DBNs through an objective function, resulting in a robust classifier. This model iteratively learns complex feature representations, progressively strengthening the classifier.

Deep boosting, as outlined in (Cortes *et al.*, 2014;Ganaie *et al.*, 2022) utilizes deep decision trees within an ensemble model, offering potential enhancements to generalization performance. It can be paired with any other classifier from a rich family to further improve performance. At each stage of deep boosting, decisions regarding which classifier to include and the appropriate weights depend on the complexity of the classifier and are guided by the principle of structural risk minimization. Multiclass Deep boosting (Saberian &Vasconcelos, 2019) extends this algorithm to address theoretical, algorithmic, and empirical aspects pertaining to multiclass problems. The use of Boosting CNNs may lead to overfitting due to the limited training data in each mini-batch. In response, Incremental Boosting CNN (IBCNN) (Han *et al.*, 2016;Mosca &Magoulas, 2017) accumulates information from multiple batches to mitigate overfitting. IBCNN employs decision stumps atop single neurons as weak learners and learns weights through the AdaBoost method within each mini-batch. Unlike DBN (Lasri *et al.*, 2023;Srivastav *et al.*, 2024), which learns weak classifiers from image patches, IBCNN utilizes weak classifiers trained from the fully connected layer, allowing the entire image to be utilized. To enhance IBCNN's efficiency, weak learners' loss functions are integrated with the global loss function. Boosted CNN (Moghimi *et al.*, 2016) employs boosting for deep CNN training, incorporating least squares objective function to integrate boosting weights into CNN. Moghimi et al. (Moghimi *et al.*, 2016) demonstrated that within their boosting framework, CNNs could be replaced by network structures to enhance base classifier performance. Boosting elevates network training complexity, prompting the introduction of dense connections in a deep boosting framework to address vanishing gradient problems in image denoising (Chen,Xiong, *et al.*, 2018). The deep boosting framework extends to image restoration (Chen,Xiong, *et al.*, 2019), where the dilated dense fusion network is employed to enhance performance.

Convolutional channel features, as described in (Li,Wen, *et al.*, 2023;Yang *et al.*, 2015), employ CNNs to extract high-level features and then utilize boosted forests for final classification. Due to the higher number of hyperparameters in CNNs compared to boosted forests, this model proves to be more efficient in terms of both performance and time than end-to-end training of CNN models. Yang et al.(Yang *et al.*, 2015) showcased its usefulness in various tasks such as edge detection, object proposal generation, and pedestrian and face detection. A stagewise boosting deep CNN (Balamurugan *et al.*, 2023) trains multiple CNN models within an offline boosting framework. To address online scenarios where only a portion of data is available at any given time, Boosting

Independent Embedding Robustly (BIER) (Opitz *et al.*, 2017) was introduced. In BIER, a single CNN model is trained end-to-end using an online boosting technique, where the training set is reweighted based on the negative gradient of the loss function to map input spaces (images) into a set of independent output spaces. Hierarchical Boosted Deep Metric Learning (Waltner *et al.*, 2019) enhances BIER's robustness by incorporating hierarchical label information into the embedding ensemble, improving performance in large-scale image retrieval applications. Snapshot Boosting (Zhang,Jiang, *et al.*, 2020) combines the benefits of snapshot ensembling and boosting to enhance generalization without increasing training costs, by training each base network and combining their outputs using a meta-learner to produce more recent output combinations. The concept of boosting underlies architectures like Deep Residual Networks (He *et al.*, 2016;Siu, 2019) and AdaNet (Cortes *et al.*, 2017), with the success of Deep Residual Networks (DeepResNet) (He *et al.*, 2016) theoretically explained within the context of boosting theory (Huang *et al.*, 2018). AdaNet maps feature vectors to the classifier space and boosts weak classifiers, while BoostResNet employs multi-channel representation boosting, proving more efficient than DeepResNet in terms of computational time. The theory of boosting is extended to online boosting in (Beygelzimer *et al.*, 2015), providing theoretical convergence guarantees, which offer improved convergence for batch boosting algorithms.

Stacking ensembling involves aggregating outputs from multiple base models or selecting the most suitable base model. Stacking integrates base model outputs using a meta-learning model, known as "model blending" or simply "blending" when the final decision component is a linear model. It serves as a technique to reduce bias. The Deep Convex Net (DCN) (Deng &Yu, 2011), proposed following, comprises a deep learning architecture with a variable number of modules stacked together. Each module in DCN is convex and consists of linear input units, hidden layer non-linear units, and a second linear layer with units corresponding to the target classification classes. Modules are connected layer-wise, with lower module outputs fed as inputs to adjacent higher modules, in addition to the original data. The Deep Stacking Network (DSN) (Tang *et al.*, 2021), enabling parallel training on extensive datasets, draws its name from the concept of "stacked generalization" . To address memory requirements, the Random Fourier Feature-based Kernel Deep Convex Network (Huang *et al.*, 2013) approximates the Gaussian kernel, reducing training time and aiding evaluation over large datasets. A framework for parameter estimation and model selection in kernel deep stacking networks (Welchowski &Schmid, 2016) combines model-based

optimization and hill-climbing approaches. Welchowski and Schmid (Welchowski &Schmid, 2016) utilize a data-driven framework for parameter estimation, hyperparameter tuning, and model selection in kernel deep stacking networks. The Tensor Deep Stacking Network (T-DSN)(Hutchinson *et al.*, 2012;Zhang *et al.*, 2022) improves upon DSN by splitting large single hidden layers into two smaller ones within each stacked network block and bilinear mapping to capture higher-order feature interactions. The Sparse Deep Stacking Network (S-DSN) (Li *et al.*, 2015;Sun *et al.*, 2018) employs sparse coding for image classification and abnormal detection, utilizing sparse simplified neural network modules (SNNM) with mixed-norm regularization. In the domain of Deep Reinforcement Learning, DSN is utilized, with Zhang et al. (Zhang,Zhang, *et al.*, 2020) integrating observations from Grasp and Stacking networks via DSN to enable integrated robotic arm actions. Stacked models are employed for various tasks such as neural architecture search (Wang,Xue, *et al.*, 2020), image deblurring (Zhang *et al.*, 2019), and embedding temporal data (Palangi *et al.*, 2014). Stacked Extreme Learning Machines (Yu *et al.*, 2015;Zhou *et al.*, 2014) reduce the number of hidden nodes at each level using Principal Component Analysis (PCA) for large-scale problems. Various stacked models based on support vector machines (Hang *et al.*, 2023;Li,Yang, *et al.*, 2019;Wang *et al.*, 2017) and deep forests (Su *et al.*, 2019;Zhou &Feng, 2019) extend traditional models to deep architectures. Novel architectures such as Stacking-based Deep Neural Network (S-DNN)(Low *et al.*, 2019) and Stacking Conditional Restricted Boltzmann Machine with Deep Neural Network (Kang *et al.*, 2020) have also been proposed.

Majority voting(Davani *et al.*, 2022), akin to unweighted averaging, aggregates the outputs of base learners by considering the majority vote for predicting final labels, rather than averaging probability outcomes. While this approach reduces bias towards a specific base learner's outcome, it may lead to the dominance of certain events if favored by a majority of similar or dependent base learners. Kuncheva et al. (2003) emphasized the significance of pairwise dependence among base learners, noting that shallow networks exhibit more diverse predictions for image classification compared to deeper networks(Choromanska *et al.*, 2015) . Thus, Ju et al. (2018) suggested that majority voting may perform better with shallow ensemble models than with deep ensemble models. Voting methods have been integrated into semi-supervised deep learning as well. For instance, Li et al.(2017) proposed ensemble semi-supervised deep acoustic models for automatic speech recognition, while Wang et al. (2020) explored ensemble self-learning methods to enhance semi-supervised performance and extract adverse drug events from social media (Liu

et al., 2018). In semi-supervised classification, a deep coupled ensemble learning method combined with complementary consistency regularization achieved state-of-the-art performance (Li,Wu, *et al.*, 2019), particularly on datasets with costly annotations. Pio et al. (2014) used an ensemble method to enhance the reliability of miRNA:miRNA predicted interactions. Multi-label classification is also addressed using voting methods, such as the Random k-labelsets (RAKEL) algorithm (Tsoumakas &Vlahavas, 2007), where several single-label classifiers are trained using small random subsets of actual labels, and final output is determined by a voting scheme based on their predictions. Shi et al. (2011) proposed a solution for the multi-label ensemble learning problem by constructing accurate and diverse multi-label-based basic classifiers and employing two objective functions to evaluate their accuracy and diversity. Another approach (Li *et al.*, 2013) introduced an ensemble multi-label classification framework based on variable pairwise constraint projection. Xia et al. (2021) proposed a weighted stacked ensemble scheme that uses sparsity regularization to aid classifier selection and ensemble construction. Ensemble multi-label methods find applications in various domains, including protein subcellular localization (Guo *et al.*, 2016), protein function prediction (Yu *et al.*, 2012), gene prediction (Schietgat *et al.*, 2010), among others. Ensemble classifier chains (ECC) (Read *et al.*, 2011) is another critical algorithm for multi-label ensemble learning, involving binary classifiers linked along a chain, with the final prediction obtained through integration of predictions above a manually set threshold. Chen et al. (2017) proposed an ensemble application of convolutional and recurrent neural networks to capture both global and local textual semantics and model high-order label correlations.

2.4. Knowledge Distillation

Knowledge distillation (Allen-Zhu &Li, 2020) is a technique utilized in deep learning, where a compact model, termed the student model, is trained to emulate the behavior of a larger, more intricate model known as the teacher model. Typically, the teacher model exhibits superior performance but is computationally more demanding. During the process, the student model is trained on the same dataset as the teacher model(Alkhulaifi *et al.*, 2021). However, instead of directly learning from the dataset's labels, the student model learns from the soft targets generated by the teacher model. These soft targets comprise probability distributions across classes, reflecting the teacher model's confidence in its predictions. Through learning from these soft targets, the student model can encapsulate the nuanced decision-making process of the teacher

model, despite having fewer parameters(Ho &Gwak, 2020). Consequently, the student model attains comparable performance to the teacher model while being more computationally efficient and suitable for deployment on resource-constrained devices.

This section provides an overview of research conducted in the field of knowledge distillation. It explores various studies that have investigated the effectiveness and applications of this technique in deep learning. Researchers have examined the process of training compact student models to replicate the behavior of larger teacher models, focusing on how soft targets generated by teacher models can be utilized to enhance the learning process. Through a comprehensive review of these studies, insights into the advancements, challenges, and potential future directions of knowledge distillation in deep learning are gained, contributing to the ongoing development of more efficient and effective model compression techniques.

Large deep neural networks have demonstrated exceptional success, particularly in real-world scenarios with extensive data, as their over-parameterization enhances generalization performance when new data is encountered (Brutzkus &Globerson, 2019;Tu *et al.*, 2019;Zhang,Liu, *et al.*, 2018). Nonetheless, deploying these models on mobile devices and embedded systems poses significant challenges due to the limited computational capacity and memory resources available on such devices. To address this challenge, Bucilua et al. (2006) initially proposed model compression, aiming to transfer information from a large model or ensemble of models to train a smaller model without a significant decrease in accuracy. Additionally, knowledge transfer between a fully-supervised teacher model and a student model using unlabeled data was introduced for semi-supervised learning(Urner *et al.*, 2011). This approach, later formalized as knowledge distillation by Hinton et al. (2015), involves training a small student model under the guidance of a large teacher model (Bucilua *et al.*, 2006;Hinton *et al.*, 2015;Urban *et al.*, 2016) . The main objective is for the student model to mimic the teacher model to achieve comparable or superior performance. The key challenge lies in effectively transferring knowledge from the large teacher model to the small student model. Generally, a knowledge distillation system comprises three essential components: knowledge, distillation algorithm, and teacher-student architecture.

Despite its significant practical success, there remains a shortage of works focusing on the theoretical or empirical aspects of knowledge distillation (Cheng *et al.*, 2020;Cho &Hariharan, 2019;Phuong &Lampert, 2019;Urner *et al.*, 2011). Notably, Urner et al. (2011) demonstrated that

the transfer of knowledge from a teacher model to a student model using unlabeled data is PAC (Probably Approximately Correct) learnable.

To gain insights into the working mechanisms of knowledge distillation, Phuong and Lampert(2019) provided a theoretical justification for a generalization bound, elucidating the fast convergence of learning in distilled student networks, particularly in the context of deep linear classifiers. Their work addresses how quickly the student learns, shedding light on the factors determining the success of distillation. Cheng et al.(2020) quantified the extraction of visual concepts from intermediate layers of deep neural networks to explain knowledge distillation. Ji and Zhu (2020) offered theoretical insights into knowledge distillation on wide neural networks, focusing on risk bound, data efficiency, and imperfect teacher aspects. Empirically, Cho and Hariharan (2019) conducted a detailed analysis of the efficacy of knowledge distillation, revealing that a larger model may not always serve as a better teacher due to the model capacity gap. They also found that distillation can adversely affect student learning. However, their study did not cover the empirical evaluation of different forms of knowledge distillation regarding knowledge, distillation, and mutual interaction between teacher and student. Furthermore, Tang et al.(2020) explored knowledge distillation for label smoothing, assessing teacher accuracy, and obtaining a prior for optimal output layer geometry.

Knowledge distillation for model compression mirrors the learning process of human beings. Building on this concept, recent advancements in knowledge distillation have expanded to various forms of collaborative learning, including teacher-student learning(Hinton *et al.*, 2015) , mutual learning(Zhang,Xiang, *et al.*, 2018) assistant teaching(Mirzadeh *et al.*, 2020) , lifelong learning(Zhai *et al.*, 2019) , and self-learning(Yuan *et al.*, 2019). These extensions primarily focus on compressing deep neural networks, resulting in lightweight student networks suitable for deployment in applications such as visual recognition, speech recognition, and natural language processing (NLP). Moreover, knowledge transfer in knowledge distillation extends to other tasks like adversarial attacks (Papernot,McDaniel, *et al.*, 2016), data augmentation(Gordon &Duh, 2019;Lee *et al.*, 2020), and data privacy and security(Wang *et al.*, 2019;Wang *et al.*, 2018). Inspired by knowledge distillation for model compression, the concept of knowledge transfer has been further applied to compressing training data, known as dataset distillation. This technique

transfers knowledge from a large dataset to a smaller one, reducing the training load of deep models(Bohdal *et al.*, 2020;Wang *et al.*, 2018).

In knowledge distillation, the types of knowledge, distillation strategies, and teacher-student architectures significantly influence student learning. This section delves into various categories of knowledge utilized in knowledge distillation. Traditional knowledge distillation relies on the logits of a large deep model as the teacher's knowledge (Ba &Caruana, 2014;Hinton *et al.*, 2015;Kim *et al.*, 2018;Mirzadeh *et al.*, 2020). Alternatively, activations, neurons, or features from intermediate layers can guide student learning (Ahn *et al.*, 2019;Huang &Wang, 2017;Romero,Ballas, *et al.*, 2015;Zagoruyko &Komodakis, 2016). The relationships between different activations, neurons, or sample pairs encapsulate rich information learned by the teacher model (Lee &Song, 2019;Liu *et al.*, 2019;Yim *et al.*, 2017;Yu *et al.*, 2019). Additionally, the parameters of the teacher model (or connections between layers) represent another form of knowledge(Liu *et al.*, 2019) . We categorize these forms of knowledge into response-based, feature-based, and relation-based knowledge.

Response-based knowledge distillation typically involves mimicking the final predictions of the teacher model by focusing on the neural response of its last output layer. This method, while straightforward, is highly effective for compressing models and finds broad application across various tasks and domains. It is particularly intuitive when considering the concept of "dark knowledge." Soft targets, which are essentially the probabilities of classes, are employed akin to label smoothing or regularizes, enhancing the effectiveness of this approach (Ding *et al.*, 2019;Kim &Kim, 2017;Müller *et al.*, 2019). However, response-based knowledge distillation primarily relies on the output of the last layer, such as soft targets, neglecting the valuable intermediate-level supervision provided by the teacher model. This oversight is crucial, especially for representation learning with deep neural networks (Romero,Ballas, *et al.*, 2015). Furthermore, as soft logits represent class probability distributions, response-based knowledge distillation is inherently limited to supervised learning scenarios.

Feature-Based Knowledge Distillation revolves around leveraging the capacity of deep neural networks to acquire diverse levels of feature representations with increasing abstraction, a phenomenon known as representation learning (Bengio *et al.*, 2013). In this approach, both the outputs of the final layer and intermediate layers, referred to as feature maps, serve as supervisory

signals for training the student model. This extension of knowledge distillation proves particularly beneficial for thinner and deeper networks.

The concept of intermediate representations was initially introduced in Fitnets (Romero,Ballas, *et al.*, 2015) to provide guidance for improving student model training. The core idea involves directly aligning the feature activations of the teacher and student models. Building upon this principle, various methods have emerged to indirectly match features (Cheng *et al.*, 2021;Heo *et al.*, 2019;Kim *et al.*, 2018;Passban *et al.*, 2021;Wang,Fu, *et al.*, 2020;Zagoruyko &Komodakis, 2016). For instance, Zagoruyko and Komodakis (2016) devised an "attention map" derived from original feature maps to convey knowledge, which was further developed by Huang and Wang (2017) through neuron selectivity transfer. Liu *et al.* (2019) achieved knowledge transfer by aligning probability distributions in feature space.

To facilitate the transfer of teacher knowledge, Kim et al.(2018) introduced "factors" as more interpretable intermediate representations. JiZhu (2020) proposed route constrained hint learning to minimize performance gaps between teacher and student by supervising the student using hint layer outputs from the teacher. Additionally, Heo et al. (2019) suggested utilizing activation boundaries of hidden neurons for knowledge transfer. Notably, Zhou et al. (2019) used parameter sharing of intermediate layers along with response-based knowledge as teacher knowledge. To ensure semantic alignment between teacher and student, Chen et al. (2021) proposed cross-layer knowledge distillation, which dynamically assigns appropriate teacher layers to each student layer through attention allocation. Despite the benefits of feature-based knowledge transfer, determining optimal hint and guided layers from the teacher and student models respectively remains a subject for further investigation (Romero,Ballas, *et al.*, 2015). Moreover, addressing the significant differences in sizes between hint and guided layers, as well as effectively matching feature representations of teacher and student, requires further exploration.

Both response-based and feature-based knowledge distillation methods utilize the outputs of specific layers in the teacher model. However, relation-based knowledge distillation goes further by exploring the relationships between different layers or data samples. For instance, Yim et al. (2017) proposed the Flow of Solution Process (FSP), which relies on the Gram matrix between two layers to summarize relations between pairs of feature maps. Lee et al. (2018) introduced knowledge distillation via singular value decomposition, utilizing correlations between feature

maps to extract key information. Zhang and Peng (2018) formed graphs using the logits and features of each teacher model as nodes to capture the importance and relationships of different teachers before knowledge transfer. Lee and Song (2019) proposed multi-head graph-based knowledge distillation, focusing on intra-data relations between any two feature maps via a multi-head attention network. Additionally, Passalis et al. (2020) explored pairwise hint information by having the student model mimic mutual information flow from pairs of hint layers of the teacher model.

Traditional knowledge transfer methods often involve individual knowledge distillation, where the soft targets of a teacher are directly distilled into the student. However, this distilled knowledge contains not only feature information but also mutual relations of data samples. For instance, You et al. (2017) and Park et al. (2019) proposed methods that transfer knowledge containing instance features, relationships, and feature space transformation. Liu et al. (2019) introduced a method via instance relationship graph, while Chen et al. Chen *et al.* (2021) proposed relational knowledge distillation based on manifold learning ideas. Passalis and Tefas (2020), Passalis et al. (2020) , and Tung and Mori (Tung &Mori, 2019) proposed methods that model relations between data samples as probabilistic distributions using feature representations, matched by knowledge transfer. Peng et al.(2019) introduced a method based on correlation congruence, containing instance-level information and correlations between instances.

Distilled knowledge can be categorized from different perspectives, such as structured knowledge of the data, with contributions from Liu et al. (2019) , Chen et al. (2021), Peng et al. (2019), Tung and Mori (2019), and Tian et al. (2019), or privileged information about input features, as discussed by Lopez-Paz et al. (2015) and Vapnik and Izmailov (2015).

2.5. Teacher-Student Architecture

In the domain of knowledge distillation, the teacher-student architecture acts as a vital conduit for transferring knowledge. Essentially, the efficacy of knowledge transfer from teacher to student is heavily influenced by the design of both networks. Analogous to the way individuals seek suitable mentors for learning, the selection or design of appropriate structures for teacher and student networks poses a significant yet challenging task. Currently, the configurations of teacher and student models often remain rigid, leading to potential gaps in model capacity.

Initially conceptualized to compress an ensemble of deep neural networks Hinton *et al.* (2015). Knowledge distillation primarily deals with the complexity of deep neural networks, which stems from their depth and width. Typically, knowledge is transferred from deeper and wider networks to shallower and narrower ones Romero,Sanchis, *et al.* (2015). Student networks are typically simplified versions of teacher networks with fewer layers and channels, quantized versions preserving network structure, small networks with efficient operations, or networks with optimized global structures, or even identical to the teacher network itself.

The disparity in capacity between large deep neural networks and small student networks can hinder knowledge transfer (Mirzadeh *et al.*, 2020). To address this, various methods have been proposed to reduce model complexity. For instance, introducing teacher assistants or employing residual learning helps mitigate the training gap between teacher and student models (Mirzadeh *et al.*, 2020;Xue *et al.*, 2021). Additionally, approaches focusing on minimizing structural differences between teacher and student models, such as network quantization combined with distillation, structure compression methods, or progressive block-wise knowledge transfer, have been effective.

In online settings, teacher networks are often ensembles of student networks, ensuring shared or identical structures among them. Techniques like depth-wise separable convolution and neural architecture search (NAS) have further enhanced the efficiency and performance of small neural networks by searching for optimal global structures. Moreover, dynamically searching for a knowledge transfer regime, such as automatically removing redundant layers using reinforcement learning or finding optimal student networks given teacher networks, has gained traction in knowledge distillation Xie *et al.* (2020).

Various teacher architectures offer distinct knowledge that can benefit a student network. Leveraging multiple teacher networks, both individually and collectively, is common during student network training. Typically, teachers possess large models or ensembles thereof. One straightforward method to transfer knowledge from multiple teachers involves using the average response from all teachers as the supervision signal Hinton *et al.* (2015) .

2.6. Interpretable Deep Learning

Interpretable deep learning revolutionizes how we approach the design and utilization of deep neural networks, aiming to explain the complex inner workings of these models and reduce their decisions comprehensible to humans. It serves as a crucial bridge between the inherent opacity of traditional deep learning models and the pressing need for transparency and accountability in AI systems. By infusing transparency and explainability into model architecture and training processes, interpretable deep learning empowers users to comprehend the decision-making process, identify critical features, and understand the rationale behind specific predictions. Key techniques such as streamlined model architectures, feature visualization, attention mechanisms, layer-wise relevance propagation, and saliency maps are pivotal in achieving interpretability. These methods not only foster trust and reliability in AI systems but also facilitate critical tasks such as debugging, fairness assessment, and regulatory compliance across various domains, spanning healthcare, finance, criminal justice, and autonomous vehicles. As interpretable deep learning continues to advance, it holds the promise of fully unleashing the potential of artificial intelligence while upholding human values and ethical principles at the forefront of AI development and deployment.

The terms interpretability and explainability are often used interchangeably by researchers, but some works distinguish them. Despite numerous attempts to define these concepts and related ones such as comprehensibility, there is no concrete mathematical definition or metric (Doshi-Velez &Kim, 2017;Gilpin *et al.*, 2018;Lipton, 2018). Doshi-Velez and Kim define interpretability as the ability to explain or present in understandable terms to a human (Gilpin *et al.*, 2018), while Miller defines it as the degree to which a human can understand the cause of a decision (Miller, 2019). These definitions lack mathematical formality and rigorousness (Adadi &Berrada, 2018). Interpretability is closely linked to the intuition behind model outputs, making it easier to identify cause-and-effect relationships within inputs and outputs (Adadi &Berrada, 2018). For example, in image recognition, certain dominant patterns in the input image may lead a system to identify a specific object in the output. In contrast, explainability pertains to the internal logic and mechanics of a machine learning system, offering deeper insights into its training or decision-making processes. An interpretable model may not necessarily allow humans to understand its internal logic or processes. Thus, interpretability does not necessarily entail explainability, and vice versa. Gilpin et al. (Gilpin *et al.*, 2018) argue that interpretability alone is insufficient, and the presence

of explainability is also crucial. This aligns with the broader view of interpretability presented by Doshi-Velez and Kim (Doshi-Velez &Kim, 2017).

Interpretability in deep learning refers to understanding and explaining how models make predictions, crucial for trust and application in critical domains like healthcare and finance. LIME (Local Interpretable Model-agnostic Explanations) is a method used to explain complex model decisions by approximating locally around predictions, enhancing transparency and trustworthiness. It aids in understanding black-box models by highlighting influential features for specific predictions, fostering insights into model behavior without sacrificing predictive power. LIME, short for Local Interpretable Model-agnostic Explanations, is an explainable artificial intelligence (XAI) technique devised to clarify the predictions made by any classifier or repressor by approximating it locally with an interpretable model (Nguyen,Cao, *et al.*, 2021). The primary goal of LIME is to offer a straightforward method with local fidelity, which ensures that explanations for individual predictions accurately capture the model's behavior in the vicinity of the specific data point being predicted. Unlike global fidelity, which considers features important across the entire dataset, local fidelity focuses solely on the features relevant to the immediate context of each prediction. Consequently, it's possible that only a subset of variables may directly influence an individual prediction, even if the model incorporates a large number of variables globally. The process of LIME involves several steps: First, it generates new samples and obtains their predictions using the original model. Next, it assigns weights to these new samples based on their proximity to the instance being explained. Subsequently, using the output probabilities from a selected set of samples that cover the relevant input space, LIME constructs a linear model. The weights assigned to this surrogate model are then utilized to quantify the importance of input features. Additionally, LIME is model-agnostic, meaning it can be applied to any deep learning model type (Nguyen,Cao, *et al.*, 2021).

2.7. Related works

In this section, research related to deep learning and interpretable deep learning approaches has been thoroughly reviewed, providing a comprehensive analysis of their methodologies.

The researchers (Ganguly *et al.*, 2022) design five classification models using the ResNet-50 architecture, with each model utilizing five distinct inputs. The inputs are the five leaf grayscale variants, RGB, and three individual RGB channels (red, green, and blue). The Bonferroni mean

operator was also used by the researchers for the fusion of the five ResNet-50. The researchers proposed a two-tier training method to properly train the end-to-end model. The researchers evaluated the proposed model using the Malayakew dataset, which was collected at the Royal Botanic Gardens in New England. This dataset contains many leaves from different species that look very similar. Furthermore, the proposed method is tested by the researcher on the Leafsnap and Flavia datasets. The researchers decided that the obtained results on both datasets confirmed that the designed models provide good performance, as it compares the results obtained by many state-of-the-art models in this article.

In a study (Abdollahi, 2022) the researcher examines the use of Convolutional Neural Network (CNN)-based techniques to differentiate Indian leaf species. The study is primarily concerned with identifying medicinal plants that can be found in rural areas. They applied the Transfer Learning technique to a well-known pre-trained CNN architecture called MobileNet-V2. The medical plants dataset was created using 30 different classes of medicinal plants, totaling 3000 photos, and these models were evaluated using their pre-trained weights. The trained model had an accuracy of 98.05 % on a held-out test set, demonstrating the practicality of this approach.

According to these researchers stated in (TS &Prabalakshmi, 2021), using image processing and computer vision strategies to distinguish proof of medicinal plants is extremely important because a large number of these plants are on the verge of extinction. The researcher's primary goal is to investigate Convolutional Neural Network (CNN)-based methodologies for distinguishing Indian leaf species. The research is primarily focused on identifying therapeutic plants that are available in rural areas. The researchers use transfer learning techniques to design their model, they use the known pre-trained models using CNN architectures with ImageNet Database. They Selected ResNet101, VGG16, and InceptionV3 pre-trained models for this research. To inspect the model, the researchers used pre-trained weights for the Ayur Bharat dataset, which was created using 10 distinct classes of medicinal plants totaling 10000 images. The researchers also attempted to improve the performance of these models using the Canny edge detection method, comparing the designed three architectures to previously trained models, both without and with preprocessing techniques. They reported that the InceptionV3 architecture trained with the Canny edge detection pre-processing technique achieved the best classification performance for the Ayur Bharat dataset, with the best validation accuracy and F1-score of 0.9732 and 0.9653, respectively.

In a study (Pacifico *et al.*, 2019)a complete automatic recognition model of medicinal plants classification problems has been proposed. The author prepares a new dataset using features of color and texture extractions from the images of a medicinal plants which consists of 1148 leaf image segments that were used from 15 species of common medicinal plants found in Brazil. The paper used five machine learning algorithms which are Artificial Neural Networks of Multi-Layer Perceptron type, Decision Tree classifier, K-Nearest Neighbors classifier, Weighted K-Nearest Neighbor classifier, and Random Forest classifier which was trained with Backpropagation algorism for the proposed classification models. The performance of the selected classifiers was evaluated using a hypothesis test of type Friedman/Nemenyi test is adopted concerning the accuracy, precision-measure, and recall. As it can be compared to the other classifiers MLP-BP and RFC are achieved the best performance has been achieved as it can be pointed out from the experimental results based on the hypothesis test of Friedman/Nemeni. Considering both speed and performance for the classification models, a Random Forest classifier would be a better choice than MLP-BP. MLP-BP achieves 39.66 ranks for the testing accuracy, 39.72 ranks of testing precession, 39.66 ranks for the testing recall, and 39.65 ranks for the testing F-measure. When we see the performance of RFC, achieves 39.42 ranks for the testing accuracy, 39.32 ranks of testing precision, 39.42 ranks of testing recall, and 39.42 ranks of testing f-measure.

Another study (Paulson &Ravishankar, 2020)proposes an indigenous medicinal plants identification using Artificial Intelligence. In this work the author's uses basic convolutional neural network (CNN) and pre-trained models VGG16 and VGG19 for the identification problems of medicinal plants. To compare the identification performance of these models a dataset of 64 species of medicinal plants of kerala state college of Vaidyaratnam Ayurveda which consists of 64000 leaf images with 1000 samples per space were used. The experimental results of this work shows that VGG16 achieves better results of 97.8% of classification accuracy, then VGG19 with 97.6% of accuracy performance and basic CNN 95.79% of accuracy respectively.

The author in a study (Van Hieu &Hien, 2020) examined BJFU100 which is collected from the Beijing Forestry University campus, a collection of 100 ornamental plants species, each with 100 different 4208x3120 pixel images. For the collected dataset, the paper's author used a deep learning model with pre-trained feature extractors, for large-scale plants segmentation and recognition. The researchers done comparative research for evaluating different pre-trained models of deep learning

models. They were used ResnetV2, VGG16, MobileNetV2 and InceptionNetV2 feature extractors with a support machine learning classifier for the identification problems of medicinal plants species. According to the reports of the researchers the MobileNetV2 provides an accuracy result of 83.9 percent.

Another paper (Chang &Chung, 2020) proposes a framework that recognizes plants species with the combination of hand-crafted feature extracted and deep neural network. In order to train the deep learning classifier, the authors use image processing mechanisms using plants leaf morphology which is basically used for annotated images of plants leaf images. The data used for training is augmented using geometrical transformation methods in addition to image processing techniques. The author's collected the dataset in Taiwan, Pingtung County under various conditions of climate change like a change of season, temperature, and humidity. In this paper, the authors conducted extensive experiments on the training dataset with and without image processing for the proposed plants species recognition models to evaluate the performance of the proposed approach. The experimental results show that the proposed mixed approach achieves a total mAP of 98.3 percent for data augmentation with image processing, 98.60 percent for data augmentation without image processing, 96.22 percent for data augmentation and image processing, and 97.40 percent for data augmentation and image processing. The results demonstrated that using the image pre-processing method results in a faster average loss than using the method without image pre-processing. Furthermore, the experiment results show that using the proposed approach can improve mAP by 2%.

Author in a study (Van Hieu &Hien, 2020), the author proposes a deep learning model for identifying Vietnamese plants species. The dataset for this study was collected from Vietnam Forest environments which consist of 28,046 total images of 109 Vietnam indigenous plants species. For deep convolutional feature extraction techniques, the author uses MobileNetV2, Inception ResnetV2, ResnetV2, and VGG16 models this work. The selected models have been tested using SVM classifier to identify plants species. The experimental result shows that VGG16 has low efficiency due to the over-fitting problem. However, Inception ResnetV2 takes much more time during evaluation than other models; ResnetV2, it achieves an average accuracy of performance. At the end, using Support Vector Machine, MobilenetV2 outperforms when it compares with other models in plants recognition method. From the experimental result, it is

demonstrating that MobilenetV2 is an excellent model for mobile applications due to its compactness while running on online servers. The proposed models achieve promising recognition rates, with MobilenetV2 coming out ahead with 83.2%. This finding shows that machine learning models have the potential to identify plants species in the wild.

In(Pudaruth *et al.*, 2021), the author proposed MedicPlants for recognition of Mauritius medicinal plants using deep learning, 1000 images of 70 different medicinal plants species have been collected from the tropical islands of Mauritius which consists of 100 images for each plants species types. They used Tensor-Flow framework and convolutional neural network (CNN) to create the identification or classification models. For validation purpose 80% of the dataset is used the rest 20% is used for testing purpose. The system has more 95% recognition accuracy.

The study (Dileep &Pournami, 2019), uses a deep learning-based model designed for the classification of medicinal plants based on leaf features such as shape, size, color, and texture. This study also puts forth a private dataset for medicinal plants commonly found in various regions of Kerala, the state situated on the southwestern coast of India. The AlexNet deep learning model demonstrates an impressive classification accuracy of 96.76% on the AyurLeaf dataset.

A study (Naeem *et al.*, 2021), employed five machine learning classifiers, a multi-layer perceptron, logit-boost, bagging, random forest, and simple logistic. These classifiers are applied to an augmented dataset containing medicinal plants leaves, incorporating both multispectral and texture data. The results indicate that the multi-layer perceptron classifier stands out with an impressive accuracy of 99.01%, surpassing its counterparts. The multi-layer perceptron classifier demonstrates distinctive accuracy levels for six medicinal plants leaves: 98.40% for Catnip, 99.90% for Lemon balm, 99.10% for Tulsi, 99.80% for Peppermint, 98.40% for Bael, 98.40% for Catnip, and 99.20% for Stevia.

A very important study (Borman *et al.*, 2022), pursues to create a medicinal plants species classification model using color and texture extraction through the Radial Basis Function Neural Network (RBFNN) algorithm coupled with the Least Mean Square (LMS) algorithm. Color feature extraction involves calculating the average RGB value, while texture feature extraction employs a Gabor filter derived from mean, entropy, and variance parameters. The extracted features serve as input for the RBFNN with LMS. The RBFNN, featuring three layers with feedforward properties,

is employed to address classification and pattern recognition challenges. The LMS algorithm facilitates the learning and updating of neural network weights. The test results, evaluating precision, recall, and accuracy, reveal a precision value of 92.50%, recall at 91.74%, and accuracy at 92.08%. The manual extraction of texture and colors features is noted to be time-consuming and burdensome.

Another study (Azadnia *et al.*, 2022), introduced a vision-based system specifically designed for the identification of herb plants. The research proposes a deep learning model for the classification and identification of the mentioned medicinal plants species. The researcher assesses the model's performance by analyzing the leaves of five different medicinal plants, and the results indicate that the vision-based system achieves a remarkable accuracy surpassing 99.3% across all image definitions.

A study (Hajam *et al.*, 2023), exploited on the capabilities of three component deep learning models. Specifically, the author employs VGG16, VGG19, and DenseNet201 to extract features from input images within a medicinal plants dataset, including leaf images from 30 different classes. The combination of VGG19 and DenseNet201, with fine-tuning, reveals enhanced capabilities in identifying medicinal plants images, demonstrating a 7.43% and 5.8% improvement compared to VGG19 and VGG16, respectively. Furthermore, VGG19+DenseNet201 surpasses its individual models, achieving an inspiring accuracy of 99.12% on the test set.

In (Uddin *et al.*, 2023), the researcher employed five state of the art deep learning models to establish benchmark performance on the dataset: VGG16, ResNet50, DenseNet201, InceptionV3, and Xception. Among these models, DenseNet201 exhibited the highest accuracy at 85.28%.

In (Malik *et al.*, 2022), researchers introduced an automated real-time system for identifying medicinal plants species across the Borneo region. The system includes a computer vission system for training and testing a deep learning model. For the plants species identification task, an EfficientNet-B1-based deep learning model was used and evaluated on a combined dataset of public and private plants species. The proposed model demonstrated Top-1 accuracies of 87% and 84% on test sets for private and public datasets, respectively, marking a notable improvement of over 10% compared to the baseline model.

Another paper (Diwedi *et al.*, 2023), introduced an Enhanced Convolutional Neural Network architecture (ECNN-PTL) utilizing a modified ResNet50 with Progressive Transfer Learning. The proposed method incorporates an enhanced ResNet50 framework for feature extraction, coupled with Progressive Transfer Learning (PTL). Indian Medicinal Plants Database (IMPLAD) has been employed for the identification and classification purpose. The results indicate that the modified ResNet50 + OSVM model achieves a testing phase accuracy of 96.8% and a training phase accuracy of 98.5%.

A paper (Kavitha *et al.*, 2023), introduced a vision-based intelligent method for herb plants recognition by employing a deep learning model. The study uses the MobileNet pretrained deep learning model for the fully automatic identification of medicinal leaves. The model undergoes thorough evaluation through processes of training, validation, and testing, demonstrating an impressive accuracy rate of 98.3% in accurately identifying medicinal leaves.

In another study (Kale *et al.*, 2023) author conducted a study using a self-created dataset consisting of 4390 images representing 35 different species of medicinal leaves. The developed system is an Android platform-based application designed for the classification of medicinal leaves, employing Convolutional Neural Network (CNN) technology. The dataset was split into 80% for training and 20% for testing. Accuracy served as a performance metric, and the system achieved an accuracy rate of 94.10% in classifying these medicinal plants species.

Gilpin et al. (Gilpin *et al.*, 2018) contribute to defining interpretability in machine learning, primarily focusing on deep learning. They propose a taxonomy classifying interpretability methods for neural networks into three categories. The first involves methods simulating data processing to reveal connections between inputs and outputs. The second explains data representation within networks, while the third category comprises transparent networks that self-explain. The authors acknowledge progress in explaining deep neural networks but note a lack of combinatory approaches that merge techniques for better explanations. Recognizing the lack of formal evaluation and measurement metrics for interpretability methods, Murdoch et al. (Murdoch *et al.*, 2019) conducted a survey in 2019 to address this gap. They proposed the Predictive, Descriptive, Relevant (PDR) framework, which evaluates interpretability methods based on predictive accuracy, descriptive accuracy, and relevancy. Additionally, they advocated for the integration of

transparent models and post-hoc interpretation, suggesting that combining these approaches could enhance predictive accuracy and expand the utility of models in certain scenarios.

In a recent study by Arrieta et al. (Arrieta *et al.*, 2020) , a novel taxonomy was proposed to categorize deep learning interpretability methods, distinguishing between transparent and post-hoc approaches and further dividing them into sub-categories. This taxonomy, tailored specifically for deep learning methods, comprises four main categories: explaining deep network processing, elucidating deep network representation, interpreting producing systems, and hybrid methods combining transparent and black-box approaches. Additionally, the authors explored Responsible Artificial Intelligence, introducing criteria for AI implementation in organizations. Regarding interpretability techniques, Gradients, first introduced in (Simonyan *et al.*, 2013), rely on gradient-based attribution to quantify the impact of input dimension changes on predictions within a small input neighborhood. This method generates image-specific class saliency maps, highlighting discriminative areas for a given class within an image. An enhanced version proposed in (Kümmerer *et al.*, 2014) utilized the Krizhevsky network to significantly outperform existing saliency models, increasing the explained information by 67%. Moreover, in (Zhao *et al.*, 2015), a task-specific pre-training scheme was developed to enhance multi-context modeling for saliency detection.

Integrated Gradients (Sundararajan *et al.*, 2017) is a gradient-based attribution method designed to explain predictions generated by deep neural networks by attributing them to the network's input features. It builds upon the concept of calculating gradients of the prediction output with respect to input features, as seen in the simpler Gradients method. One key property it satisfies is completeness, also known as Efficiency or Summation to Delta (Roth, 1988;Shrikumar *et al.*, 2017), ensuring that the attributions sum up to the difference between the target output and the output evaluated at a baseline. Moreover, the method adheres to two fundamental axioms of attribution methods: sensitivity and implementation invariance. Recognizing that many existing attribution methods fail to meet these axioms, Integrated Gradients offers a straightforward approach to achieve robust interpretability results. Another related study (Mudrakarta *et al.*, 2018) leverages attribution methods, including Integrated Gradients, to identify weaknesses in question-answer models more effectively than conventional approaches, thereby enhancing workflow efficiency.

DeepLIFT (Shrikumar *et al.*, 2017) is a widely used algorithm designed to enhance the interpretability of deep neural network predictions. It builds upon the Gradient Input method, improving saliency maps by multiplying the gradient with the input signal. By analyzing neuron activations and assigning contribution scores based on differences between outputs and inputs, DeepLIFT uncovers dependencies often missed by other methods, such as Integrated Gradients (Sundararajan *et al.*, 2017). Guided BackPropagation (Springenberg *et al.*, 2014), also known as guided saliency, questions the use of max-pooling in convolutional neural networks (CNNs) for small images and suggests replacing max-pooling layers with convolutional layers of increased stride. This modification preserves accuracy on various image recognition benchmarks. Deconvolution (Zeiler &Fergus, 2014), initially proposed as a technique for visualizing CNNs using De-convolutional Networks (DeconvNets) (Zeiler *et al.*, 2011), provides insights into intermediate feature layers. By mapping feature activity back to input feature space, DeconvNets reveal the original input patterns causing specific activations in feature maps, offering valuable interpretability for trained CNNs.

Class Activation Maps (CAMs), introduced in (Zhou *et al.*, 2016), offer a method for interpreting convolutional neural networks (CNNs) by highlighting discriminative regions in images used for classification. By computing averages of activations from convolutional feature maps just before the final output layer, CAMs create a feature vector. This vector is then weighted and fed to the final softmax loss layer, enabling identification of important image regions for classification by projecting weights onto convolutional feature maps. However, CAMs have limitations: they require specific network structures in the final layers, necessitating architecture changes and retraining for other networks, and they only visualize the final convolutional layers, offering insights solely into the last stages of image classification. Grad-CAM (Gradient-weighted Class Activation Mapping), introduced in (Selvaraju *et al.*, 2017), represents a significant advancement over CAM, as it enables visual explanations for any CNN, regardless of its architecture. By leveraging class-specific gradient information from the final convolutional layer, Grad-CAM produces coarse localization maps highlighting important image regions for classification, enhancing the transparency of CNN-based models. Additionally, the technique can be combined with existing pixel-space visualizations to create high-resolution, class-discriminative visualizations, known as Guided Grad-CAM. Comparative evaluations against pixel-space gradient visualizations showed superior performance in terms of localization and faithfulness.

Despite these advancements, Grad-CAM has limitations, such as difficulty in localizing multiple object occurrences, challenges in accurately determining class-regions coverage, and potential signal loss due to upsampling and downsampling processes.

Grad-CAM++ (Gradient-weighted Class Activation Mapping Plus Plus), an extension of Grad-CAM, as introduced in (Chattopadhay *et al.*, 2018), enhances visual explanations of CNN predictions by extending object localization to multiple instances within a single image. It achieves this by employing a weighted combination of positive partial derivatives of the last convolutional layer's feature maps with respect to a specific class score, assigning different weights to each pixel to capture their individual importance in the gradient feature map. This extension is particularly advantageous in multi-label classification scenarios, allowing for more nuanced interpretation of model predictions. Layer-wise Relevance Propagation (LRP), described in (Bach *et al.*, 2015), is an interpretability technique for complex deep neural networks, offering insight into their predictions by backward propagation. LRP decomposes nonlinear classifiers, ensuring the conservation of output magnitude as information is propagated from output neurons to input-layer neurons. This method is versatile, applicable to diverse data types and neural network architectures, including images and text.

Smilkov et al. (Smilkov *et al.*, 2017) introduced SmoothGrad, leveraging the interpretation of the loss function gradient with respect to the input as a sensitivity map to enhance visualization clarity. By integrating with methods like Integrated Gradients and Guided BackPropagation, SmoothGrad reduces noise, sharpening sensitivity maps. Two smoothing approaches, averaging perturbed maps and adding random noise, demonstrated additive benefits, producing sharper and more coherent visualizations compared to unsmoothed gradients. The RISE algorithm (Petsiuk *et al.*, 2018) facilitates the interpretation of deep neural network predictions for images by generating saliency maps for any black-box model. It employs a straightforward yet effective strategy: each input image undergoes element-wise multiplication with random masks, and the resulting images are classified by the model. By assessing the model's scores for the masked inputs across available classes, a saliency map for the original image is constructed via a linear combination of the masks, with coefficients determined by the model's scores for the masked inputs relative to the target class.

Yosinski et al. (Yosinski *et al.*, 2015) proposed regularization techniques as an additional step in saliency map creation to enhance interpretability. By imposing stronger prior distributions, the

methods promote bias towards more recognizable and interpretable visualizations, with optimal results achieved through their combined application. Each regularization technique independently contributes to improved interpretability. In (Linardatos *et al.*, 2020), an interpretability technique for neural networks in natural language processing (NLP) was introduced, involving the extraction of smaller, tailored input text pieces called rationales. These rationales provide explanations for predictions by attempting to reproduce the same output as the original full-text input. The architecture comprises a generator and an encoder trained jointly to identify text subsets highly associated with predicted scores. Deep Taylor decomposition (Montavon *et al.*, 2017) is a method that breaks down a neural network's output for a given input into contributions from individual pixels by backpropagating explanations from the output layer to the input. Primarily applied in computer vision tasks, it offers insights into the importance of single pixels in image classification without requiring hyperparameter tuning. This method, closely linked to relevance propagation, produces heatmaps enabling deep understanding of each input pixel's impact on classification, applicable to various network architectures and datasets.

Kindermans et al. (Kindermans *et al.*, 2017) demonstrated that while methods like Deconvolution, Guided BackPropagation, and LRP aid in interpreting deep neural networks, they may not provide theoretically correct interpretations, even in simple linear model settings. They revealed that the network's gradient direction doesn't estimate the signal in the data but reflects the relationship between signal and noise. Addressing this, they proposed PatternNet and PatternAttribution as theoretically sound interpretation methods for linear models. PatternNet improves upon existing visualizations by estimating the correct direction, while PatternAttribution identifies the contribution of signal dimensions to the output across network layers. Contrarily, Zafar and Khan (Zafar &Khan, 2019) argued that LIME's random perturbation and feature selection methods yield unstable interpretations, generating different interpretations for the same prediction, posing deployment challenges. To mitigate this uncertainty, they introduced DLIME, a deterministic version of LIME replacing random perturbation with hierarchical clustering and k-nearest neighbors (KNN) for cluster selection. Using medical datasets, they showcased DLIME's superiority over LIME in terms of Jacart Similarity among multiple explanations.

The local interpretable model-agnostic explanations (LIME) method, initially introduced in (Ribeiro *et al.*, 2016), stands as one of the most widely-used interpretability approaches for black-

box models. Operating through a straightforward yet potent mechanism, LIME generates interpretations for individual prediction scores from any classifier. It accomplishes this by simulating random data samples around the vicinity of a given input instance, for which a prediction was made. These generated instances are then used to make new predictions with the model, weighted by their proximity to the original input. Finally, a simple interpretable model, such as a decision tree, is trained on this augmented dataset, allowing for the interpretation of the initial black box model. Despite its effectiveness, LIME has limitations. In 2020, the first theoretical analysis of LIME (Garreau &Luxburg, 2020) was published, affirming its significance while also highlighting the potential consequences of poor parameter choices, which could lead to overlooking crucial features.

On the other hand, Shapley Additive explanations (SHAP) (Lundberg &Lee, 2017), drawing inspiration from game theory, strives to enhance interpretability by calculating importance values for individual features in predictions. The authors first establish the class of additive feature attribution methods, which encompasses six existing methods, including LIME, DeepLIFT, and Layer-Wise Relevance Propagation, all employing the same explanation model. Subsequently, they introduce SHAP values as a unified measure of feature importance, maintaining desirable properties such as local accuracy, and consistency. Various methods for estimating SHAP values are presented, supported by experiments showcasing their superiority in distinguishing output classes and their alignment with human intuition compared to many other existing methods.

In (Ribeiro *et al.*, 2018), a novel model-agnostic interpretability method was proposed, known as anchors, offering high precision and a probabilistic guarantee for any black-box model. Anchors are high-precision, if-then rules created to represent local, sufficient conditions for predictions. An anchor is defined as a rule that sufficiently determines the prediction locally, meaning changes to other feature values of the instance do not fundamentally alter the prediction. Anchors are incrementally constructed using a bottom-up approach. Initially, each anchor is initialized with an empty rule applicable to every instance. Through iterative steps, new candidate rules are generated, and the one with the highest estimated precision replaces the previous rule for that anchor. If a candidate rule meets the anchor definition, the process terminates. While this approach aims to discover the shortest anchor, it doesn't directly optimize for the highest coverage, although shorter anchors are likely to achieve high coverage. A user study demonstrated that anchors lead to higher

human precision compared to linear explanations, requiring less user effort in both application and understanding.

The contrastive explanations method (CEM), originally introduced in (Dhurandhar *et al.*, 2018), offers contrastive explanations for any black box model. It identifies not only which features should be minimally present for a specific prediction but also which features should be minimally absent. While many interpretation methods focus solely on the presence of features, CEM considers the critical absence of features, which is essential for forming natural interpretations, as demonstrated in domains like healthcare and criminology. Luss et al. (Luss *et al.*, 2019)extended the CEM framework to images with richer structure by defining monotonic functions to introduce additional concepts into an image without removing existing ones. Wachter et al. (Wachter *et al.*, 2017) proposed a lightweight model-agnostic interpretability method called counterfactuals, providing counterfactual explanations. A counterfactual explanation describes the smallest change needed in feature values to alter the output prediction to a predefined desired output. Unlike other interpretability methods focusing on understanding a black-box system's inner workings, counterfactual explanations reveal external factors necessary for producing a desired output. While counterfactual explanations may not suffice for understanding a system's functionality or decision-making rationale, they serve as an initial step in achieving transparency, explainability, accountability, and regulatory compliance.

Van Looveren et al. (Van Looveren &Klaise, 2021) identified several shortcomings in the previous counterfactual approach (Wachter *et al.*, 2017), particularly its lack of consideration for local, class-specific interpretability and its computational inefficiency due to the growing dimensionality of the feature space during the counterfactual searching process. To overcome these limitations, they proposed an enhanced, faster, and model-agnostic technique for generating explainable counterfactual explanations for classifier predictions. Their novel method integrates class prototypes, constructed using either an encoder or class-specific k-d trees, into the cost function. This incorporation facilitates quicker convergence of perturbations to interpretable counterfactuals, thereby eliminating the computational bottleneck and enhancing the method's practical applicability. To evaluate the effectiveness of their approach and the quality of the generated counterfactuals, the authors introduced two new metrics focusing on local interpretability at the instance level. Through experiments conducted on both image data (MNIST

dataset) and tabular data (Wisconsin Breast Cancer dataset), they demonstrated that prototypes contribute to producing counterfactuals of superior quality. Additionally, they highlighted the challenge of generating meaningful perturbations and counterfactuals for categorical features due to the absence of a natural notion of distance or rank among different variable values. To address this issue, the authors proposed using embeddings based on pairwise distances between categorical variable values. Empirical results showcased the effectiveness of these embeddings when combined with their method on census data.

The work outlined in (Kim *et al.*, 2016) on prototypes was expanded upon in (Gurumoorthy *et al.*, 2019) by introducing non-negative weightings to prototypes based on their contribution. This extension created a cohesive framework encompassing both prototypes and criticisms/outliers. Additionally, the proposed framework allows for the utilization of any symmetric positive definite kernel, leading to objective functions with favorable properties. Subsequently, ProtoDash was introduced as a fast and mathematically sound approximation algorithm for prototype selection. Operating within the proposed framework, ProtoDash optimally selects prototypes and learns their non-negative weights. Permutation importance (PIMP) (Altmann *et al.*, 2010) is a heuristic approach designed to address bias in feature importance measures by normalizing them. The method operates under the assumption that the random importance of a feature.

L2X (Chen,Song, *et al.*, 2018) is an instance-wise feature selection method designed for real-time applications, doubling as a tool for model interpretation. Its objective is to identify the subset of input features that carry the most information regarding the prediction for a given training example. This subset is determined by a feature selector employing vibrational approximation, which aims to maximize the mutual information between input features and the corresponding label. Additionally, the study introduces a novel metric called post-hoc accuracy to quantitatively evaluate L2X's performance. Experiments conducted on both real and synthetic datasets demonstrate the effectiveness of L2X, not only in terms of post-hoc accuracy but also in human-judgment evaluations, particularly with nonlinear additive and feature-switching datasets. PDPs (Partial Dependence Plots) were proposed by Friedman (Friedman, 2001) as a visualization tool to interpret any black box predictive model. These plots illustrate the impact of specific features or feature subsets on the model's predictions. PDPs showcase how a particular set of features influences the average predicted value while marginalizing the remaining features. Although PDPs

offer a simplified view and may not capture all feature interactions accurately, they can still provide valuable insights for interpreting black box models, particularly when interactions are of low order. PDPs are versatile, capable of visualizing relationships for both single and multi-class problems, as well as feature interactions.

ICE plots, originally proposed in (Goldstein *et al.*, 2015), represent a model-agnostic interpretability method that builds upon the concept of Partial Dependence Plots (PDPs) while addressing their limitations. Recognizing that PDPs may struggle to capture complex relationships, especially in the presence of substantial interaction effects, ICE plots offer a refinement. In ICE plots, each plot depicts the functional relationship between the predicted value and a feature for individual instances, providing insight into the distribution of individual conditional expectation functions. This allows for the identification of heterogeneities and their extent across the dataset. ALE (Accumulated Local Effect) plots (Apley &Zhu, 2020) are closely related to PDPs and aim to overcome one of their significant shortcomings: the assumption of feature independence. Instead of averaging predictions, ALE plots calculate the average differences in predictions to block the effects of correlated features, thus providing a more accurate representation of feature effects. Many methods, such as Grad-CAM (Selvaraju *et al.*, 2017) and deconvolution neural networks (Zeiler &Fergus, 2014), produce saliency maps to explain deep learning models, leveraging gradient information for this purpose. Grad-CAM, in particular, has garnered significant attention and influence in terms of citations per year. For explaining any black-box model, LIME (Ribeiro *et al.*, 2016) and SHAP (Lundberg &Lee, 2017) are prominent and comprehensive methods for visualizing feature interactions and importance. Despite being older and less sophisticated, Friedman's PDPs (Binder *et al.*, 2016) remain popular. Both LIME and SHAP are model-agnostic and applicable to various data types.

In the study (Aldughayfiq *et al.*, 2023), the authors investigated the applicability of LIME and SHAP, two widely-used explainable AI techniques, in generating local and global explanations for a deep learning model based on the InceptionV3 architecture trained on retinoblastoma and non-retinoblastoma fundus images. They curated a dataset comprising 400 retinoblastoma and 400 non-retinoblastoma images, labeled them accordingly, and divided the dataset into training, validation, and test sets. The deep learning model was trained using transfer learning from a pre-trained InceptionV3 model. Subsequently, the authors applied LIME and SHAP to generate explanations

for the model's predictions on both the validation and test sets. The results obtained from this analysis demonstrated the effectiveness of LIME and SHAP in identifying the regions and features within the input images that significantly contribute to the model's predictions. By providing insights into the decision-making process of the deep learning model, these explanations offer valuable information for understanding its behavior and enhancing its interpretability.

In general, in the literature review there as a gap in the dataset, hence most of the related works are used their own private dataset available in specific countries, geographical disparity is also observed. There is a pressing needs of interpretability and explainability in this area. Hence, medicinal plants identification and classification requires interpretability for having a better insight to the final result and decisions. The other gaps found in the literature is the pretrained models are highly resource intensive, it is not affordable for devices having small computational resources.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1. Chapter Overview

This chapter describes the research methodology employed for the identification and classification of Ethiopian indigenous medicinal plants species. The primary emphasis is on determining a suitable research design, methods, tools and techniques aligned with the specific objectives of the research. The discussion encompasses various aspects, including research design, data collection methods, sites, and periods of data collection. Additionally, a detailed description of the dataset, its objectives, and values is provided. The chapter also explains dataset pre-processing techniques, optimization methods, and the performance evaluation metrics adopted in the research. The tools employed throughout the research process are clearly outlined in the chapter.

Classifying medicinal plants species from digital images is a difficult task. The deep learning approach using CNN architecture has proved to be capable of adequately dealing with the most challenges associated with medicinal plants species identification and classification. In this study, we presented a systematic review of the primary studies related to deep learning from the domain of medicinal plants classification and recognition. The study assessed the deep learning approach used to classify and identify medicinal plants species with the consideration of the geographical distribution of the paper, the availability of medicinal plants dataset sources, the techniques used for pre-processing the dataset, the feature extraction techniques used, and the deep learning classifier used.

The review can help the researchers understand how a deep learning approach can be used to classify and identify Medicinal Plants Species. A clear understanding in these domains equips researchers with the necessary tools for effective utilization of deep learning models in medicinal plants identification and classification.

Overall, this comprehensive exploration of the research methodology ensures the clarity of the research process flow and provides a strong foundation for understanding the strategies and techniques employed in the study.

3.2. Research Design

Choosing an appropriate research design is essential for effectively addressing research objectives. In this study, a design science has been opted to be employed. The Design Science Research

Process (DSRP) consists of several phases including problem identification and motivation, Literature Review, identification of the Objectives, Methods, Tools and Techniques, Model development, and Communication (Peffers *et al.*, 2020). Design science aims to create effective artifacts and fulfil stakeholders' expectations(Hevner *et al.*, 2008). The rationale behind selecting design science lies in its suitability for crafting artifacts in deep learning and machine learning. These artifacts encompass constructs, models, methods, and real-life prototypical products to address specific issues in image recognition, identification, detection, and classification domains. Another reason for adopting the design science research methodology is its application in addressing unresolved problems in innovative ways and finding more efficient solutions to previously solved ones(Hevner *et al.*, 2008).

Therefore, this research adopts experimental design science approach. It begins by exploring different deep learning interventions to identify and classify optimal solutions systematically, using a systematic literature review approach. The challenges associated with the identification and classification of medicinal plants species are investigated rigorously. Subsequently, the research conducts different experiments to select the most suitable deep learning approach and ultimately designs an interpretable model for the identification and classification of Ethiopian indigenous medicinal plants species using the Ethiopian custom dataset.

In this research study, a primary dataset of Ethiopian medicinal plants species was collected using appropriate data collection and preparation mechanisms as it can be indicated in Table-6 and section 3.5. The dataset was developed by following the necessary designing and development phases. Afterwards, the dataset was categorized into a training dataset, validation and testing datasets. Appropriate data processing steps are used for data formatting, clearing, scaling, transformation, and enhancements. Data augmentation techniques are also employed as they provided a solution for issues related to data imbalance.

The literature review highlights the promising performance of deep learning in image identification and classification (He *et al.*, 2016;Krizhevsky *et al.*, 2012;Simonyan &Zisserman, 2014). Consequently, this research adopts a deep learning approach to design an interpretable deep learning model, employing pre-trained models with Transfer Learning for the identification and classification of Ethiopian medicinal plants. The design incorporates knowledge distillation concepts, aiming to optimize performance while minimizing resource consumption. By employing the strengths of pre-trained models, the research seeks to enhance the efficiency and accuracy of

identifying Ethiopian indigenous medicinal plants species. The interpretable deep learning model aligns with the research objective, ensuring not only strong performance but also an insightful understanding of the fundamental procedures required during the model design.

3.3. Data Collection, and Site Selection

In this research, leaf images data from Ethiopian indigenous medicinal plants at the Gullele Botanical Garden, situated in the northern region of Addis Ababa City was collected. Research reveals that the Gullele Botanic Garden (GBG) covers cover 705 hectares of land located in the northern parts of Addis Abeba City Administration, in the sub-cities of Gullele and Kolfe Keraniyo and the area's elevation ranges from 2450 to 2995 meters above sea level (Seta &Belay, 2021). Gullele Botanic Garden (GBG) is the world's largest and youngest botanical garden (Seta &Belay, 2021). It is surrounded by a community of various social and cultural groups. The data collection was performed in two periods, from September to December 2022 and from January to March 2023. The dataset is collected in two periods to capture seasonal variations of leaf images. This approach ensures that the dataset reflects differences in leaf characteristics due to changing seasons, providing a comprehensive representation of leaf conditions. By including data from distinct seasonal periods, the dataset can account for variations in colors, texture, and other attributes influenced by seasonal changes.

Leaf images were selected for their year rounded availability, offering a continuous and accessible source of data throughout the year for the research study. Each image was accurately taken, selected, and cropped to focus on the leaf area, and then saved in JPG format. A total of 2200 leaf images were collected, representing 44 species of Ethiopian medicinal plants, with 50 samples per species then augmented to form a sample of 12,438 leaf images. Two botanist experts were involved in identifying indigenous medicinal plants, recording traditional uses, and collaborating with traditional practitioner communities during the data collection process. The species were chosen based on ecological significance and traditional uses, with insights from these experts. Prioritizing diversity and confirming image data availability, we conducted a preliminary analysis to identify promising candidates. Involving both botanists ensured the accuracy and reliability of species identification. Species were selected only when both botanists agreed, and their choices were verified against recorded documentation. Excluding species in cases of disagreement eliminated potential inaccuracies and discrepancies, thereby enhancing the overall quality and credibility of the dataset.

This rigorous approach ensures that the collected data is robust, trustworthy, and suitable for further ecological and botanical analysis. Constraints such as ecological diversity, seasonal variability, data availability, expert agreement, documentation verification, practicality, and cultural relevance were addressed, and the rationale for each selection was documented. Validation with experts ensured alignment with ecological, cultural, and practical considerations. This systematic approach guarantees a representative subset for meaningful and robust image analysis in plants species identification and classification.

Equal-sized sampled images were systematically captured from each species to prevent biases stemming from dataset imbalances. We maintained diversity by acquiring 25 leaf samples from the front side and 25 from the back side of every species. Knowledgeable botanists from the Gullele Botanical Garden labeled the images using a standardized naming convention. This sampling approach facilitated the creation of a comprehensive dataset encompassing Ethiopian medicinal plants species, thereby improving the accuracy of our model in species identification.

3.4. Dataset Description of Ethiopian Indigenous Medicinal Plants Species

A comprehensive overview of the dataset is presented in table-1, highlighting key characteristics such as the collection source, dataset availability, acquisition techniques, image format, and dataset types. Providing insights from fundamental characteristics to more intricate details, the dataset description aims to offer clarity and a deeper understanding of the information it contains. The detailed breakdown provides researchers and analysts with valuable insights into the preparation, structure, and availability of the Ethiopian indigenous medicinal plants species dataset. This information aids in their feature research for the conservation and preservation of these endangered medicinal plants species.

The dataset contains the identification of 44 species of Ethiopian indigenous medicinal plants, and its availability will support researchers in developing innovative techniques to enhance the challenges of identification, recognition, and classification in this field. Advancements in this field have significant benefits for traditional medicine healers, pharmacists, botanists, and other stakeholders involved in conserving knowledge and preserving the availability of these plants species. The proposed dataset can be utilized to create a practical, user-friendly, and interactive application that recognizes and classifies Ethiopian medicinal plants species, their parts, and their uses.

The Ethiopian indigenous medicinal plants dataset serves as a foundation for the research community, providing a starting point for further exploration and development. It is a valuable resource that researchers can build upon by incorporating additional leaf images and species. The dataset can be valuable in testing image recognition classifiers for the identification of various medicinal plants. By using the dataset's images of medicinal plants leaves, classification algorithms can be trained, tested, and validated. The data can be utilized for various Artificial Intelligence, machine-learning and deep learning applications, including image classification and image detection, medicinal plants identification, medicinal plants part and use identification, medicinal plants leaf property detection and even for medicinal plants phenotyping.

Implementing deep learning models for medicinal plants identification and detection using images of their leaves is an exciting area for future study and development. In order to create efficient deep learning and machine learning systems, researchers in this area require easy access to representative dataset. However, currently there is no standard image dataset of Ethiopian medicinal plants leaves for identifying and classifying Ethiopian indigenous medicinal plants species. Leaves of plants are generally available throughout the year compared to flowers and fruits that appear seasonally. As such leaves provide a more consistent resource for identification year around. The objective behind building this Ethiopian medicinal plants dataset was to reduce manual work and increase efficiency by the automatic identification of Ethiopian indigenous medicinal plants using machine learning and deep learning.

The other goal is to create a standardized Ethiopian indigenous medicinal plants dataset that will help spread the world about deep learning and its potential applications for the general public in some way, no matter how small. Therefore, we believe it is critical to create and disseminate such a dataset to promote study of Ethiopian indigenous medicinal plants. In the opinion of many leading experts in the field of deep learning, the benefits of this field have not yet been fully exploited for societal goods like healthcare and traditional medicine healers and others especially for Ethiopian context as the majority of the peoples are dependent on these indigenous medicinal plants.

Table 1 Ethiopian indigenous medicinal plants species dataset description.

Subject	*Computer Science and Engineering, Computational Biology, Medicine, Botany.*
Specific subject area	Computer Vision, Image Classification, Image Processing, Deep Learning, Machine Learning, Image Identification
Data format	*JPG*
Type of data	Medicinal Plants leaf image
Data collection	To gather images for this dataset, the collection process involved placing the plants leaf on an A4 paper to minimize background effects and maintain consistent leaf image quality. The images were captured using a Samsung Galaxy A23 with a 50-megapixel camera (resolution 1080 x 2408 pixels). The device used for capturing the images was equipped with a lighting control feature, allowing for precise and controlled image capture. This ensured that the dataset comprises high-quality visuals that accurately represent the leaf images.
Data source location	The dataset was collected from the Gullele Botanical Garden, located in the northern part of Addis Ababa City. Nevertheless, it's crucial to emphasize that the medicinal plants within this garden were sourced from multiple regions throughout the country, thus encompassing a broad spectrum of geographical features and

	environmental conditions. The botanical garden itself exhibits a rich diversity of geographical and environmental characteristics.
Data accessibility	Repository name: figshare Data identification number(DOI): https://DOI.org/10.6084/m9.figshare.24137802.v1 Direct URL to data: https://figshare.com/articles/dataset/Ethiopian_Indigenous_Medicinal_Plants_Dataset/24137802

3.5. Preprocessing Ethiopian Indigenous Medicinal Plants species Dataset

It is widely recognized that unprocessed images are not suitable for analysis and need to be converted into processed formats such as JPEG, JPG, or TIFF for further examination. In this study, the captured images were converted to the JPG format and saved accordingly. Before conducting image analysis, it is crucial to perform data processing to ensure the integrity of the experimental data. Proper image preparation plays a vital role in obtaining satisfactory results. In the data preparation phase, the dataset folders were organized and named based on the scientific names of the plants species. During the data pre-processing stage of this study, various techniques were employed, including image normalization, formatting, manual removal of low-quality images, image resizing, cropping of irrelevant sections, and other enhancement methods. These steps were undertaken to enhance the quality and suitability of the images for subsequent analysis.

3.5.1. Image Normalization

Image normalization is a crucial technique used in image processing to enhance the image quality by adjusting the pixel values to a standard scale(Sun *et al.*, 2023). Its main purpose is to reduce image variations due to various factors like lighting conditions and noise, which can adversely affect the image's quality. In technical terms, the process of image normalization involves rescaling

the pixel values of an image using a mathematical formula to conform to a specific range, typically between 0 and 1 or -1 and 1(Lee *et al.*, 2021). This technique can improve the quality of images significantly and enhance the accuracy of various image analysis tasks. In this study, the image normalization process was performed by multiplying each pixel value by 1/255.

3.5.2. Image Resizing

Image resizing is a widely used technique in image processing that involves changing the size of an image, either by scaling it up to make it larger or down to make it smaller(Talebi &Milanfar, 2021). This approach can be utilized for multiple purposes, including resizing an image to fit a specific space, enhancing image quality, or reducing file size. In the context of classification, the gathered image data may have different sizes, necessitating their transformation to match the desired dimensions of the model being used. In order to standardize the dataset of Ethiopian medicinal plants species that was collected, we performed image rescaling and set their dimensions to 224 × 224 with 3 colors channels. This resizing process ensures uniformity in the dataset and prepares the images for further analysis and classification. In other experiments, EfficientNet necessitated distinct sizes for various architectures. Consequently, resizing was performed to specific dimensions (224 × 224, 260 × 260, and 380 × 380 pixels) with three color channels tailored to each EfficientNetB0, EfficientNetB2, and EfficientNetB4 pre-trained model.

3.5.3. Image Cropping

In the field of image processing, the act of choosing a particular section of an original image and discarding the unnecessary parts is known as image cropping(Lu *et al.*, 2020). This technique is commonly used to achieve different goals, including enhancing the image's composition, bringing focus to a specific subject, or eliminating any distracting elements in the background(Takahashi *et al.*, 2019). Furthermore, image cropping is a useful tool for resizing images when they are too big to fit into a designated area or when a smaller version is necessary for quicker processing or smaller file sizes. In this study, we conducted image cropping on leaf images of Ethiopian medicinal plants species to remove irrelevant sections and enhance the quality of the images. The purpose of this process was to improve the performance of the classification algorithm by focusing on the essential features of the leaves. By eliminating unnecessary sections, we aimed to optimize the dataset and

ensure that the classification algorithm receives clear and relevant visual information for accurate analysis.

3.5.4. Image Augmentation

A lack of an adequate dataset or an inconsistent class balance within datasets is the most frequently mentioned issue in the field of deep learning. Because the performance of deep learning or machine learning models is dependent on the quantity and diversity of training data, collecting such data can be costly and time-consuming in many cases. Currently, companies and AI researchers leverage data augmentation to reduce dependence on training data collection and preparation to build an efficient deep learning model. In fact, data augmentation is a set of mechanisms to increase the amount of data artificially by generating new data points from existing data. This includes making small changes to the data using manual data augmentation techniques like image cropping, change in size, shape and position of the image. Deep learning is engaged in this process to generate new data points. In this regard, Image augmentation refers to the practice of expanding a dataset by applying diverse modifications to the original images, such as alterations in brightness and contrast, rotation, flipping, or zooming(Kostrikov *et al.*, 2020;Takahashi *et al.*, 2019). This technique aims to generate additional data for training deep learning models, especially for computer vision tasks like object recognition and classification(Mikołajczyk &Grochowski, 2018). By implementing image augmentation, the model can handle variations in the input data more effectively, leading to more accurate and robust predictions(Kostrikov *et al.*, 2020;Perez &Wang, 2017). Additionally, it is beneficial in situations where the initial dataset is small as it can help prevent overfitting and improve the generalizability of the model as well as address issues with small dataset(Perez &Wang, 2017). After applying image augmentation techniques, the dataset's total size has expanded to 12,438 instances. The parameter used in the augmentation techniques of this study are presented in the following table (table-2).

Table 2 Parameters of augmentation techniques.

No	*Operation*	*Values*	*Properties*
1	rotation _range	90	Randomly rotate images with random angles between 0 and 90 degrees
2	width_shift_range	0.2	Shift the image along X-axis by 20%
3	height_shift_range	0.2	Shift the image along Y-axis by 20%
4	Shear_range	0.2	Shear the image by 20%
5	zoom_range	0.2	Zoom In and Zoom Out by 20%
6	horizontal_flip	True	Enable horizontal flipping
7	vertical_flip	False	Disable vertical flipping
8	fill_mode	Nearest	Fill the area with the nearest pixel and stretch it

3.5.5. Data Splitting

Data splitting is a process utilized in machine learning and deep learning to divide a dataset into several subsets for distinct purposes such as training, validation, and testing(Azimi *et al.*, 2021;Ramcharan *et al.*, 2017). Its primary objective is to assess a model's performance on unseen data and prevent overfitting, where the model becomes too specific to the training data and performs poorly on new data(de Luna *et al.*, 2019). Typically, the dataset is divided into two or three subsets. The training subset is used to train the model, the validation subset is used to fine-tune model hyper-parameters and select the best model, and the testing subset is employed to evaluate the final model's performance (Sun *et al.*, 2017).

In this study, data-splitting techniques were employed for classifying Ethiopian indigenous medicinal plants species. In this research, various experiments were conducted, with the first experiment involving the use of 35 Ethiopian medicinal plants species collected in the initial round as indicated in the data collection period. The dataset, consisting of 9,265 images of Ethiopian medicinal plants species after augmentation, was divided into training (80%), testing (10%), and validation (10%) sets. Following the standard practice of allocating a larger portion to training data, more than two-thirds of the total, the training dataset comprised 7,742 images representing 35 species. Additionally, the validation dataset included 830 images, and the testing dataset consisted of 693 images.

In other experiments, the dataset, consisting of 12,438 images after augmentation, was divided into 80% for training, 10% for testing, and 10% for validation. Following the common practice of allocating a larger portion, more than two-thirds, to the training data, the training dataset comprised 9,766 images from 44 species. Additionally, the validation subset included 1,189 images, and the testing subset consisted of 1,438 images.

3.6. Optimization Techniques

Training deep learning is challenging and complex due to the intrinsic complexity of the algorithm since optimization techniques or algorithms are required to be used for increasing or optimizing the performance of the model because the model performance is directly affected by the optimization algorithms. However, the basic principles or functions of the optimization algorithm are basically targeted at the training dataset and loss function because the main aim of this optimization algorithm is targeted at reducing training errors. Recently, different deep learning optimization algorithms are deployed, for example, ADAM optimizer, AdaGrad, RMSProp, Gradient Descent and etc. In the proposed research work, state-of-the-art optimization algorithms will be used.

In the training setup phase this work, Stochastic Gradient Descent (SGD) (Tian *et al.*, 2023) and a categorical cross-entropy loss (Rajaraman *et al.*, 2021) were used as the optimizer and the loss, respectively. To prevent overfitting, all the pre-trained models developed in this study were carefully trained with the training data. To ensure this, a validation dataset was used during each epoch of the training phase to assess the model's performance. Specifically, overfitting was deemed absent when the validation accuracy exhibited stability or improvement while the training loss demonstrated stability or decrease. Thus, we can confidently assert that the identified models were adequately trained without overfitting. For convenience, table-3 presents the hyperparameters used, including the selected values from the search spaces for the network. We conducted comprehensive experimentation and experimental evaluation to carefully select these hyperparameters. The objective is to improve the model’s performance by preventing overfitting and ensuring effective training.

Table 3 Hyperparameters specifications.

Hyperparameters	***Properties***
Epochs	20
Activation	Relu,Softmax
Regularization	Batch Normalization
Optimizer	SGD,ADAM
Learning Rate	0.0001
Alpha	(0.3,0.5)
Temperature	(4,5)
Momentum	0.9
Batch Size	32,64,128
Image Size	224,224
Output Classes	35,44

3.7. Performance Evaluation Metrics

To determine the best pre-trained model for building our ensemble model and to evaluate the effectiveness of our novel ensemble model for identifying Ethiopian medicinal plants parts and their usage, we conducted an extensive evaluation. The same evaluation metrics has been used for evaluating the performance of our novel interpretable deep learning model. This evaluation included a thorough comparison of experimental results, taking into account metrics such as accuracy, precision, F1 score, and recall. Throughout this evaluation process, we carefully examined the performance of the models during the training and validation phases to ensure a comprehensive assessment. The mathematical formulations for each performance metric are given as follows:

True Positive (TP): Instances that are actually positive and are correctly predicted as positive.

True Negative (TN): Instances that are actually negative and are correctly predicted as negative.

False Positive (FP): Instances that are actually negative but are incorrectly predicted as positive.

False Negative (FN): Instances that are actually positive but are incorrectly predicted as negative.

Accuracy: is a measure of how the model predicts the class of a given data.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \text{---------- (3.1)}$$

Precision: measures how often a positive value prediction is correct.

$$Precision = \frac{TP}{TP+FP} \text{ ---------- (3.2)}$$

Recall: describes how sensitive the classifier is while detecting positive instances. It is commonly known as sensitivity.

$$Recall = \frac{TP}{TP+FN} \text{ ---------- (3.3)}$$

F1-score: describes the mean/harmonic mean of the above two matrices, i.e. Precision and Recall. Its lowest value is 0, meaning one of the two values is 0. This condition indicates perfect precision or recall.

$$F1 - Score = \frac{2}{\left(\frac{1}{recall}+\frac{1}{Precision}\right)} \text{ ---------- (3.4)}$$

3.8. Research Tools

In the course of this research, the execution of fundamental tasks such as image processing, identification, and classification played a crucial role. To accomplish these activities, a comprehensive array of software and open-source libraries was employed, including Jupyter Notebook, TensorFlow, Keras, Pandas, sci-kit-learn, and various other valuable tools for plotting and measuring. The selection of these tools was intricately linked to the specific nature of the identification and classification tasks, necessitating a careful examination to identify the most suitable tools capable of addressing the challenges posed by these tasks. The use of Jupyter Notebook facilitated a dynamic and interactive environment for data analysis, code development, and visualization. TensorFlow and Keras, being prominent deep learning frameworks, provided robust support for building, training, and evaluating the pretrained and the designed novel models. Pandas, a powerful data manipulation library, played a pivotal role in handling and analyzing the dataset, ensuring a seamless integration of data into the research pipeline. Sci-kit-learn, a versatile machine learning library, contributed to the implementation of various deep learning algorithms/models and evaluation metrics.

In the pursuit of training our deep learning models, the research uses the capabilities of Google Colab Pro+. This platform granted access to an A100 GPU and an impressive 83.5 GB memory capacity, marking the highlight in deep learning GPUs. The deployment of such advanced computational resources played a pivotal role in accelerating the development and training

processes of the designed models. The high-performance GPU technology offered by Google Colab Pro+ significantly accelerated the experimentation and training phases, enhancing the efficiency and effectiveness of the overall research endeavour. The availability of these GPU technology through Google Colab Pro+ played a transformative role in enhancing the speed and performance of our deep learning models. This technological advancement proved instrumental in successfully executing the proposed tasks related to image identification and classification. The seamless integration of powerful computational resources with sophisticated software tools contributed to the model designed to identify and classify Ethiopian indigenous medicinal plants species in a good performance.

CHAPTER FOUR

EXPERIMENTAL RESULT AND DISCUSSION

4.1.Chapter Overview

Accurate and effective identification and classification of Ethiopian indigenous medicinal plants are vital for their conservation and preservation. However, the exiting identification and classification process is time-consuming, tedious, and demands the expertise of specialists. Botanists traditionally rely on traditional and experience-based methods for identifying various medicinal plants species. This research aims to develop an efficient deep learning model through transfer learning for the identification and classification of Ethiopian indigenous medicinal plants species. In this work, numerous experiments have been done with the use of pre-trained deep learning models, specifically VGG16, VGG19, Inception-V3, and Xception.

The comparative model analysis presented in this study makes a substantial contribution to scholarly discussions on medicinal plants research, particularly in the context of conserving and preserving endangered medicinal plants species. The noteworthy contributions of this chapter can be summarized as follows: Firstly, the paper conducts a performance evaluation of pretrained deep learning models, VGG16, VGG19, InceptionV3, and Xception, upon utilizing the Ethiopian indigenous medicinal plants species dataset. The benchmark results provide valuable insights into the strengths and limitations of these pre-trained models, offering a precise understanding of their effectiveness in identifying and classifying Ethiopian indigenous medicinal plants species. Secondly, the practical implications of the research findings are significant for the accurate identification and classification of Ethiopian indigenous medicinal plants species. The precise identification and classification of these plants species hold the potential to harness their healing properties and contribute to the preservation of the rich biodiversity in Ethiopia. Thirdly, the study addresses common challenges in identification and classification tasks, such as class imbalance and network overfitting. Through the implementation of transfer learning techniques on Ethiopian indigenous medicinal plants custom dataset, the research successfully tackles these challenges. The observed significant performance improvements in the identification and classification tasks of Ethiopian indigenous medicinal plants species underscore the effectiveness of the applied transfer learning approach.

Fourthly, identify the traditional uses and parts predominantly applied by traditional practitioners in Ethiopian indigenous medicinal plants species using ensemble learning. Finally, design an interpretable deep learning model using knowledge distillation that addresses computational consumption issues for small handheld devices, facilitating the identification and classification of Ethiopian indigenous medicinal plants species.

4.2. Pretrained Models in Identifying and Classifying Ethiopian Indigenous Medicinal Plants

This section introduces pre-trained deep learning model with transfer learning to address the identification and classification challenges of Ethiopian indigenous medicinal plants species by employing custom dataset. The study specifically examines 35 species of Ethiopian medicinal plants collected during the initial phase of data collection. The dataset contains 9265 images of Ethiopian medicinal plants species after augmentation. Following the split of the total number of datasets, the training data consists of 7742 images of 35 species, while the validation dataset subset has 830 and the testing subsets have 693 images. Finally, transfer learning is applied for the identification and classifications of Ethiopian medicinal plants species using pre-trained deep learning model.

Transfer learning is a method used in deep learning to transfer the acquired knowledge and learned features of a pre-trained model to a new problem. This technique can lead to improved performance while using fewer computational resources and data (Pathak *et al.*, 2022). In this section, we have developed a deep learning model using transfer learning for the identification and classifications of Ethiopian indigenous medicinal plants species.

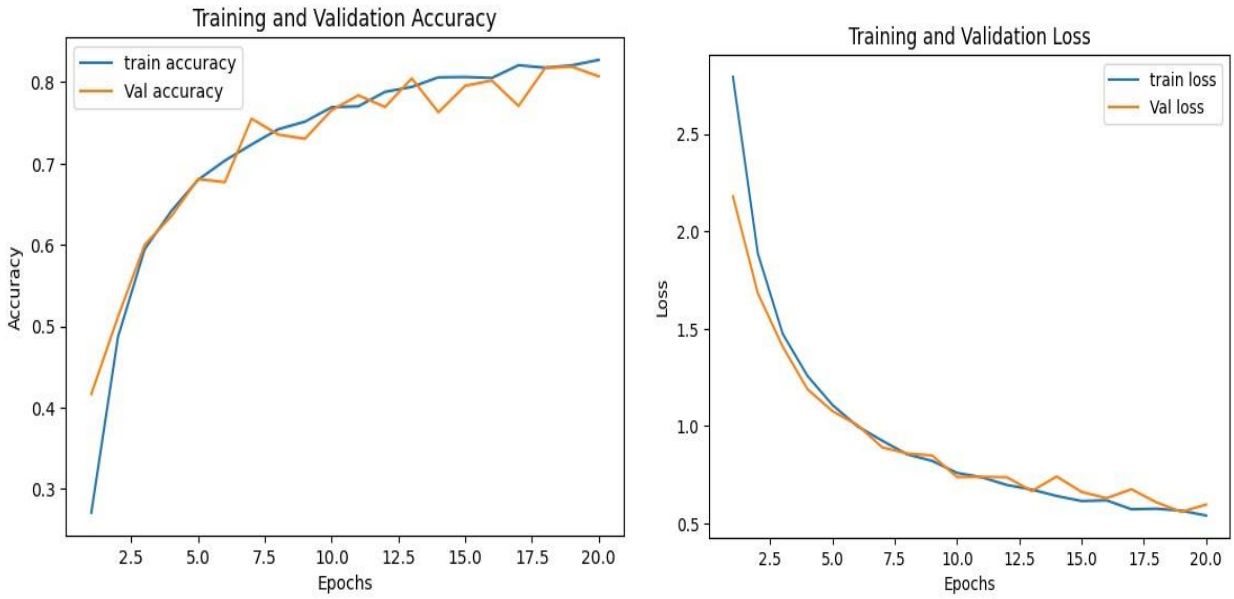
In this work, various experiments were conducted to explore the effectiveness of different pre-trained models for the identification and classification tasks of Ethiopian medicinal plants species. As outlined in table-4 of chapter three, to ensure consistency and reliability in our findings, we standardized the hyperparameters for each of these pre-trained models. The table (table-4) presents the results of various experiments conducted to identify the most effective pre-trained model for the identification and classification of Ethiopian indigenous medicinal plants species.

Table 4 Experimental results of various pre-trained models without and with fine-tuning.

Experimental results without model fine-tuning					
Models	Execution Time	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
VGG16	4:07:42.18	83%	0.54	75%	0.71
VGG19	3:21:34.08	79%	0.66	77%	0.7
Inception-V3	3:20:19.45	87%	0.38	83%	0.59
Xception	4:18:36.45	90%	0.29	88%	0.38
Experimental results with model-fine tuning					
VGG16	2:30:58.56	95%	0.13	92%	0.20
VGG19	2:46:02.9	95%	0.15	94%	0.19
Inception-V3	3:04:42.37	93%	0.22	91%	0.30
Xception	2:24:03.68	92%	0.30	87%	0.42

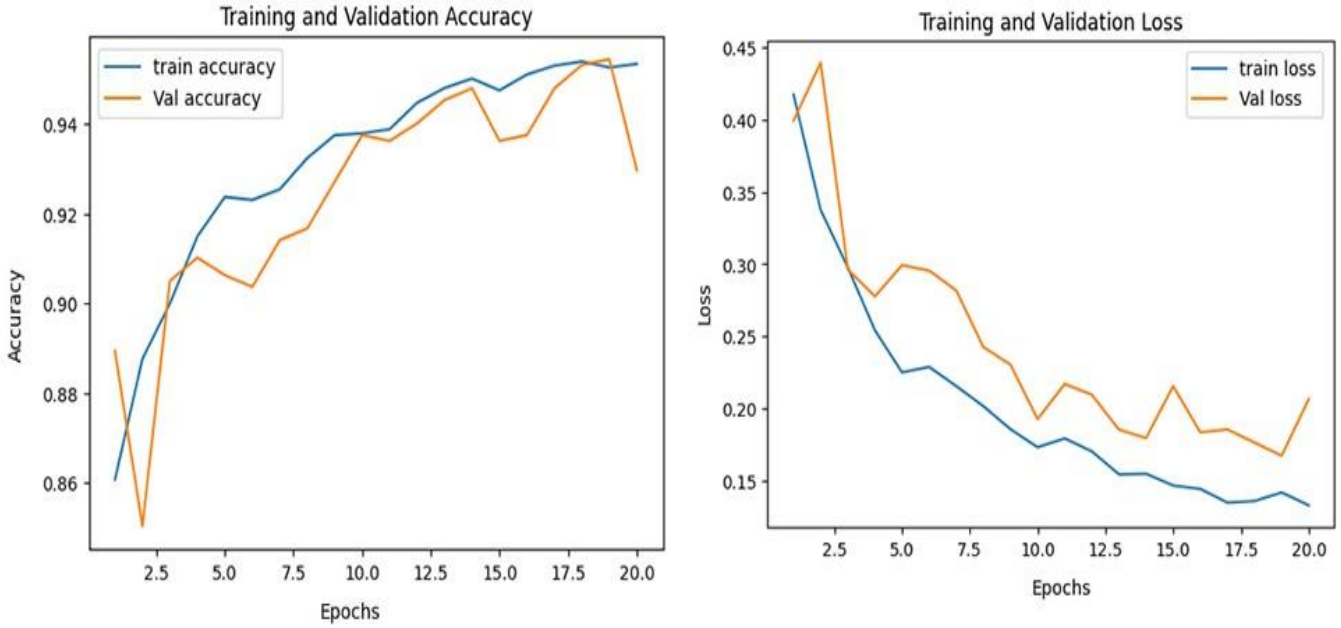
4.2.1. VGG16 pre-trained model

Initially, the VGG16 pre-trained model was evaluated with and without fine-tuning on a GPU infrastructure. The training duration was 4 hours, 7 minutes, and 42 seconds before applying fine-tuning hyperparameters, and 2 hours, 30 minutes, and 58 after applying fine-tuning.



(A) **(B)**

Figure 5 (A) Training and Validation Accuracy of VGG16 without fine-tuning; (B) Training and Validation Loss of VGG16 without fine-tuning.



(C) (D)

Figure 6 (C) Training and Validation Accuracy of VGG16 with fine-tuning; (D) Training and Validation Loss of VGG16 with fine-tuning

The analysis of the results revealed that the pre-trained model achieved a training accuracy score of 83% and a validation accuracy of 75%. The training and validation losses were recorded as 0.51 and 0.71, respectively, before fine-tuning (figure 5 A and B). However, after fine-tuning the pre-trained models, the training accuracy improved to 95%, and the validation accuracy increased to 92%. This indicates a significant improvement in accuracy, with a 12% increase in training accuracy and a 17% increase in validation accuracy.

By adjusting the hyperparameters, the performance of the pre-trained models are improved while minimizing the execution or training time. The experimental results demonstrated that the VGG16 model exhibited slight overfitting problems without fine-tuning. However, after applying fine-tuning, the validation and training losses were reduced to 0.13 and 0.20, respectively, indicating a better balance. The experimental results, shown in figure 6 C and D, confirmed that the VGG16 model performed well in classifying both training and test image data after fine-tuning. The experimental result indicated that the fine-tuning process helps to reduce overfitting and underfitting issues of the pretrained deep learning models.

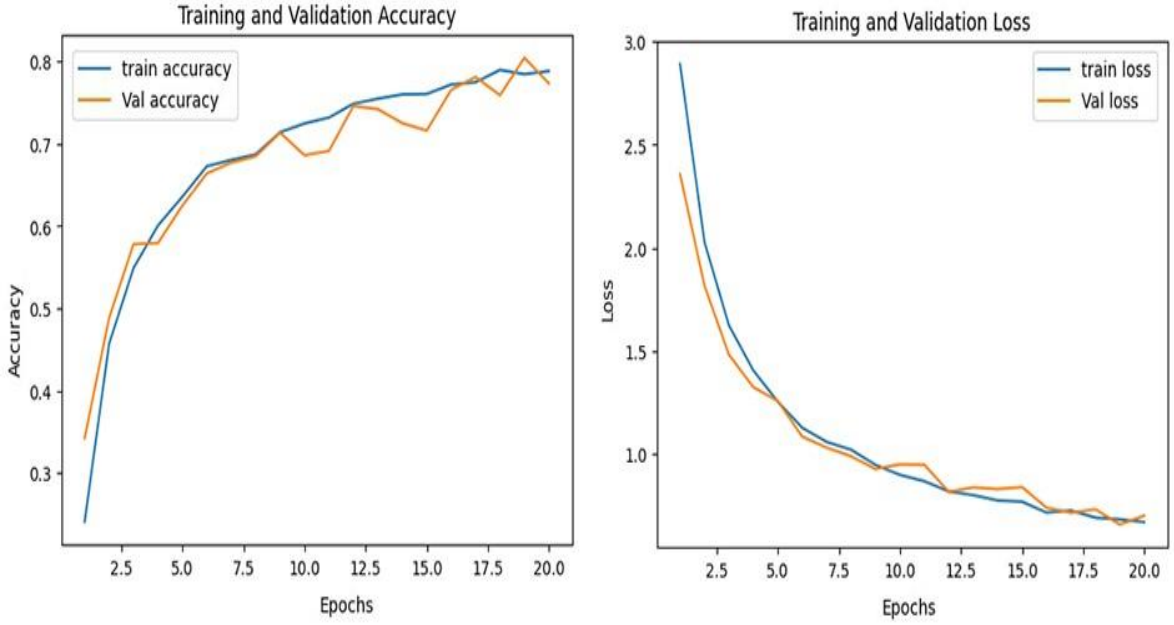
4.2.2. VGG19 pre-trained model

The second experiment involved evaluating the VGG19 pre-trained model using the same computing infrastructure and parameters as the previous experiment. The training duration was 3 hours, 21 minutes, and 34 seconds before applying fine-tuning parameters, and 2 hours, 46 minutes, and 2 seconds after applying fine-tuning.

The experimental results showed that the VGG19 model required less execution time compared to the VGG16 pre-trained model. Before fine-tuning, the model achieved a training accuracy of 79% and a validation accuracy of 77%. After applying fine-tuning parameters, the training accuracy improved to 95%, and the validation accuracy reached 94%. These results indicate that the VGG19 pre-trained model performs better in terms of accuracy and is not affected by overfitting or underfitting issues like those of VGG16.

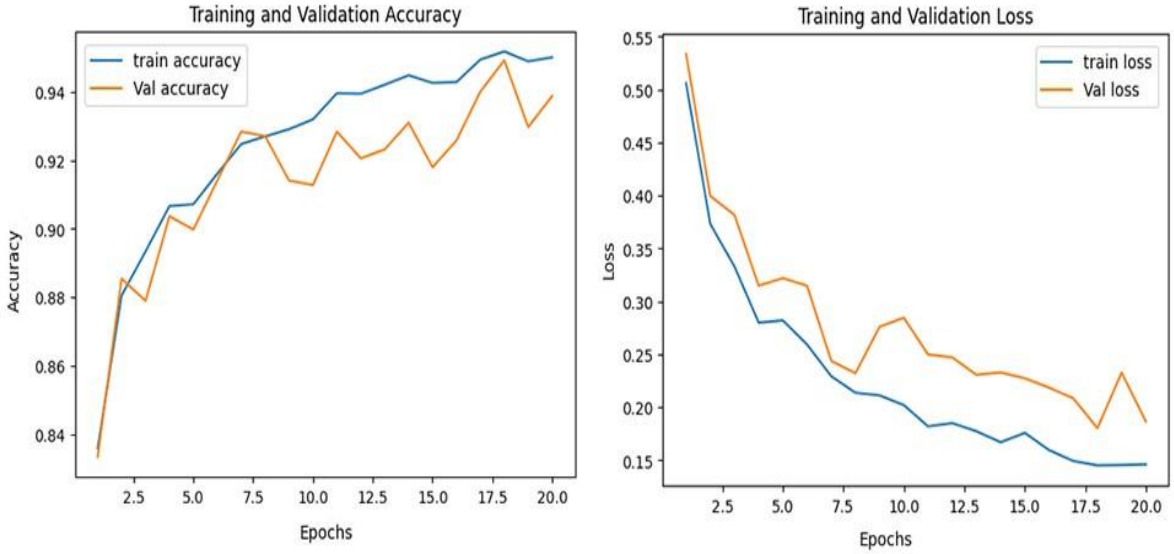
The accuracy and loss scores of the VGG19 pre-trained model for classifying Ethiopian indigenous medicinal plants are depicted in figure-7 A and B. The training and validation losses were found to be 0.66 and 0.70, respectively. However, after applying fine-tuning hyperparameters, the training loss decreased to 0.15, and the validation loss decreased to 0.19. Figure-8 C and D demonstrate that the models fit well after fine-tuning. The results suggest that further parameter

adjustments could potentially enhance the performance of the VGG19 pre-trained model. Additionally, compared to the VGG16 pre-trained model, the VGG19 model showed slightly better performance in terms of validation accuracy, indicating its ability to accurately classify images as the dataset size increases.



(A) **(B)**

Figure 7 (A) Training and Validation Accuracy of VGG19 without fine-tuning; (B) Training and Validation Loss of VGG19 without fine-tuning.



(C) **(D)**

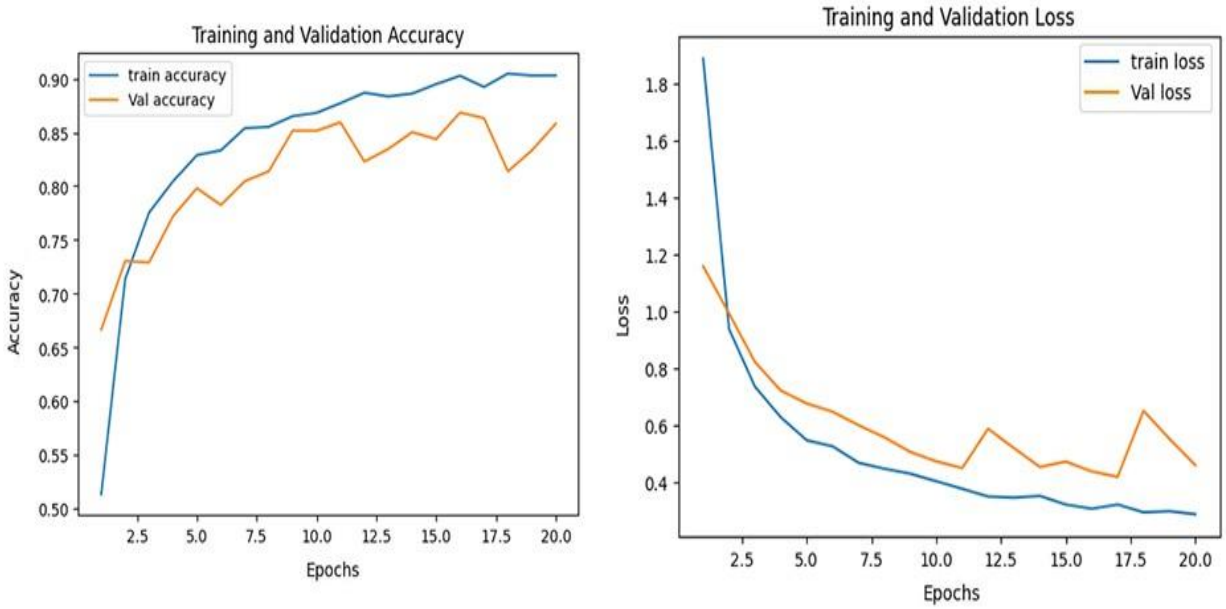
Figure 8 (C) Training and Validation Accuracy of VGG19 with fine-tuning; (D) Training and Validation Loss of VGG19 with fine-tuning.

4.2.3. Inception-V3

The third experiment involved evaluating the Inception-V3 pre-trained model using the same GPU infrastructure with and without fine-tuning. The training duration was 3 hours, 20 minutes, and 19 seconds, and after applying fine-tuning parameters, the execution time was 3 hours, 4 minutes, and 42 seconds. The execution time remained relatively consistent before and after fine-tuning.

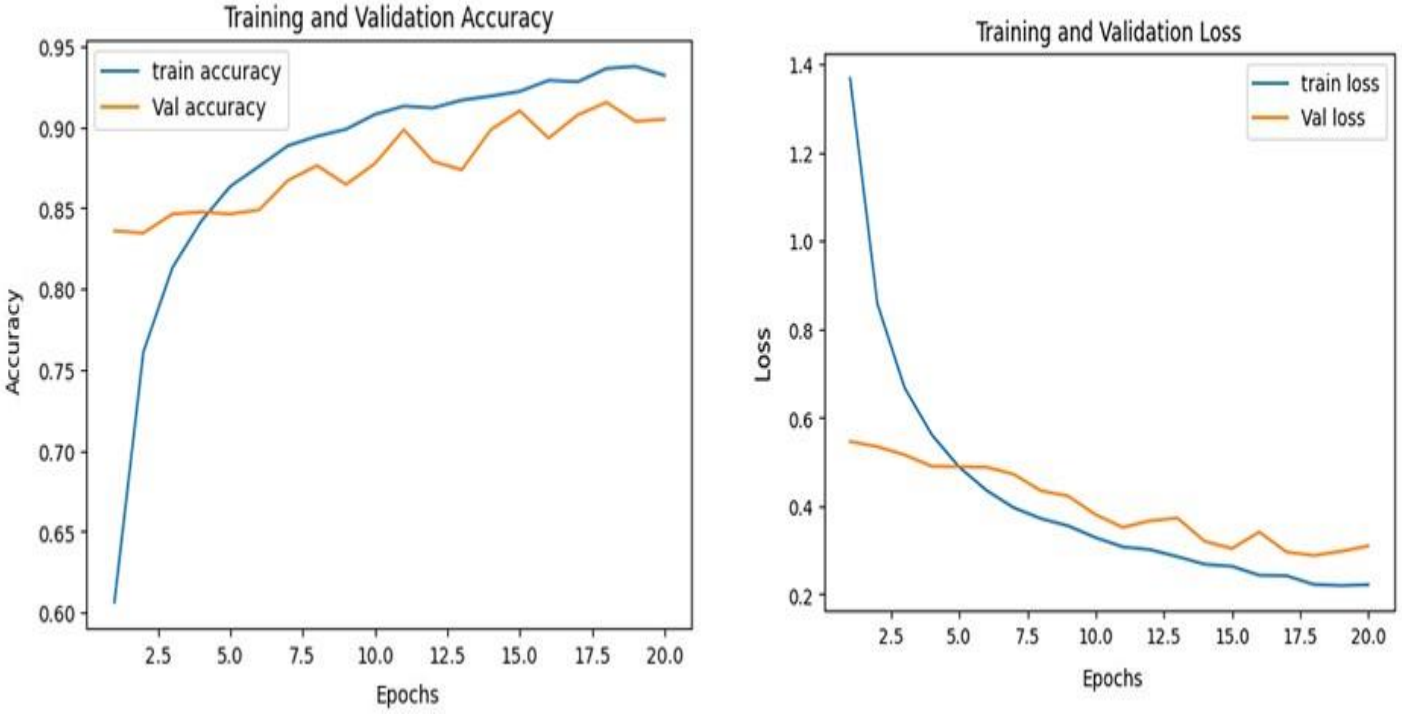
Before fine-tuning, the model achieved a training accuracy of 87% and a validation accuracy of 83%. After applying fine-tuning parameters, the training accuracy increased to 93%, and the validation accuracy reached 91%. The accuracy and loss scores of the Inception-V3 pre-trained model for classifying Ethiopian indigenous medicinal plants are depicted in figure (figure 9 A and B). The training loss was 0.38, and the validation loss was 0.59. The experimental result indicates overfitting issues during training because the validation loss is much higher than that of the training loss.

Figure (figure 10 C and D) illustrate the performance scores of the Inception-V3 model with and without fine-tuning parameters. The training and validation loss scores after fine-tuning were 0.22 and 0.30, respectively, indicating a reduction in overfitting problems.



(A) **(B)**

Figure 9(A) Training and Validation Accuracy of Inception-V3 without fine-tuning; (B) Training and Validation Loss of Inception-V3 without fine-tuning.



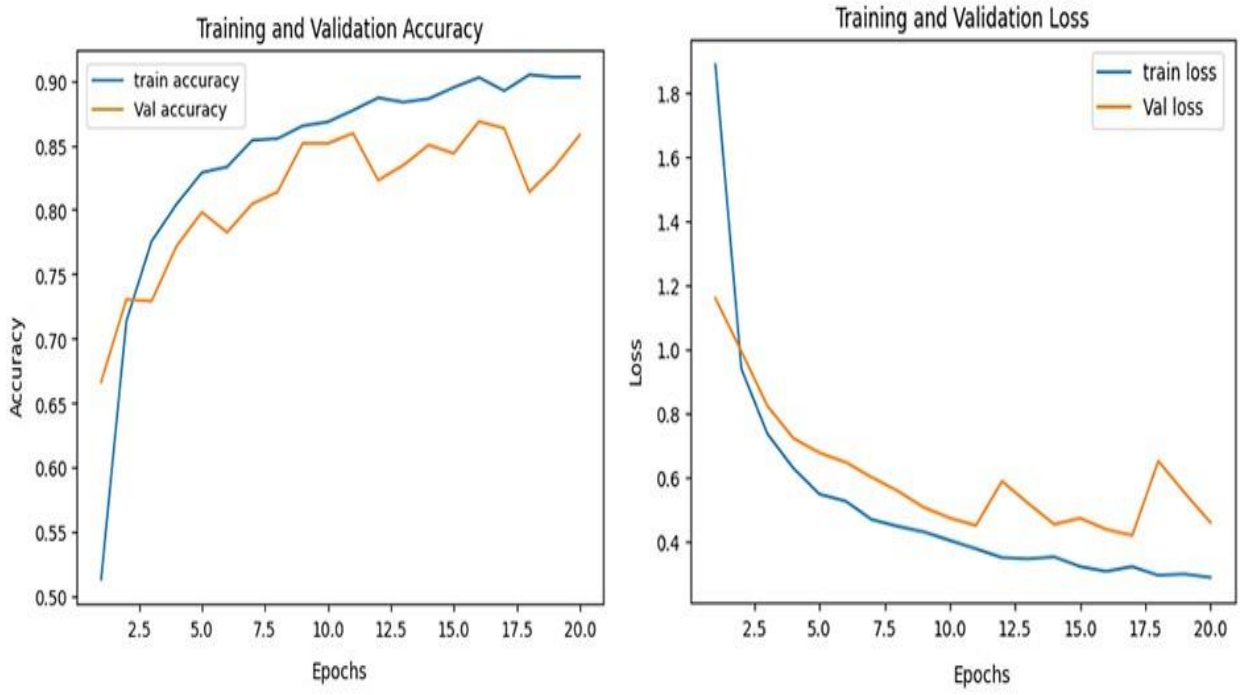
(C) **(D)**

Figure 10 (C) Training and Validation Accuracy of Inception-V3 with fine-tuning; (D) Training and Validation Loss of Inception-V3 with fine-tuning

4.2.4. Xception

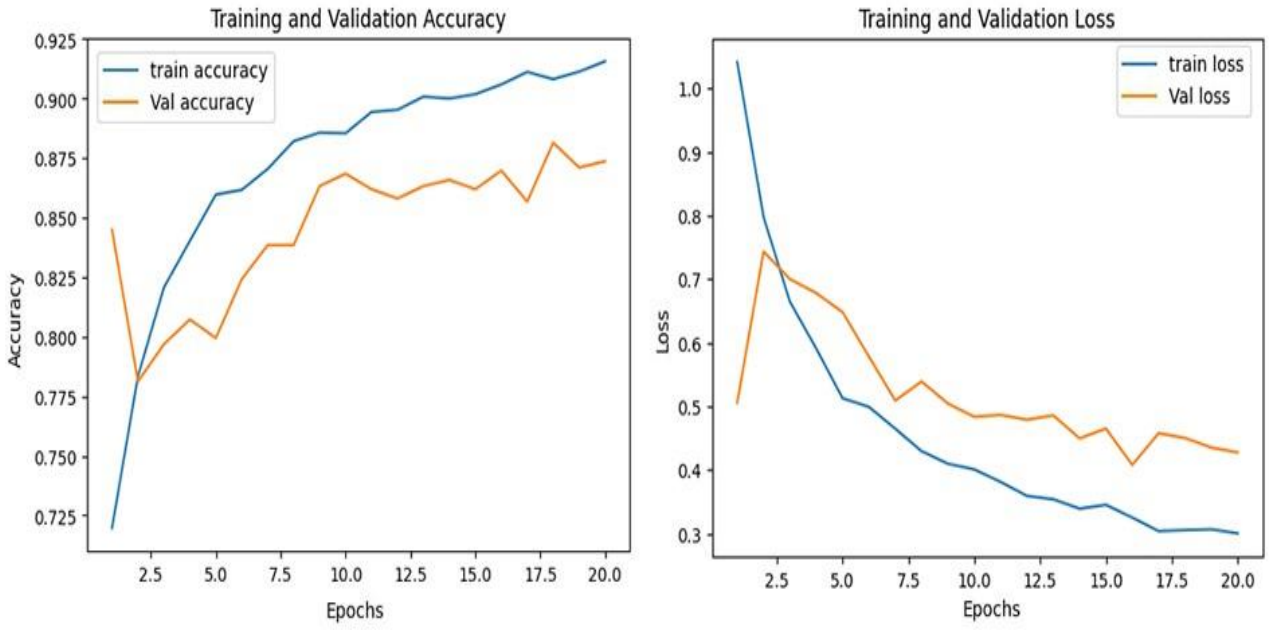
The fourth experiment focused on evaluating the Xception pre-trained model for classifying Ethiopian indigenous medicinal plants species. The training process took 4 hours, 18 minutes, and 36 seconds before fine-tuning parameters were applied. However, after applying fine-tuning, the training duration decreased to 2 hours, 24 minutes, and 3 seconds, effectively cutting the execution time in half. Before fine-tuning, the Xception model achieved a training accuracy of 90% and a validation accuracy of 88%. After applying fine-tuning parameters, the training accuracy increased to 92%, while the validation accuracy slightly decreased to 87%.

The accuracy and loss scores of the Xception model without fine-tuning were visualized in Figure (figure 11 A and B). The training loss was 0.29, and the validation loss was 0.38, indicating the presence of some overfitting issues due to some variation in training and validation loss. Figure (figure 12 C and D) illustrate the performance of the model after fine-tuning, showing that further parameter adjustments could enhance its performance. However, the model still exhibited some overfitting challenges even with fine-tuning.



(A) **(B)**

Figure 11(A) Training and Validation Accuracy of Xception without fine-tuning; (B) Training and Validation Loss of Xception without fine-tuning.



(C) **(D)**

Figure 12 (C) Training and Validation Accuracy of Xception with fine-tuning; (D) Training and Validation Loss of Xception with fine-tuning

4.3. Identifying Ethiopian Medicinal Plants Parts and Uses Using Ensemble Learning

This section seeks to identify the traditional uses and the specific parts predominantly utilized by traditional practitioners in Ethiopian indigenous medicinal plants species. The primary research objective is to accurately identify the parts and their associated uses for Ethiopian medicinal plants species. However, the precise identification of specific plants parts and their uses has been challenging due to the complexity of traditional healing practices. To tackle this issue, a majority-based ensemble deep learning approach is employed to identify the medicinal plants parts and their uses in Ethiopian indigenous medicinal plants species.

Ethiopia, with its diverse flora and numerous ethnic groups, each possessing unique methods for utilizing medicinal plants, is a rich repository of traditional knowledge. The country's array of languages, cultures, and belief systems adds to the vast diversity in traditional practices concerning the use, management, and conservation of plants resources. (Abera, 2014;Tefera *et al.*, 2019). This research aims to highlight Ethiopia’s traditional healing treasures by implementing an efficient ensemble deep learning framework for Ethiopian medicinal plants species. Utilizing the state-of-

the-art EfficientNet architecture, we can improve the accuracy and efficiency of identifying the parts of Ethiopian indigenous medicinal plants species and their associated uses.

Advancements in deep learning frameworks like EfficientNet have been effectively utilized in various fields, including image recognition and classification. This study aims to reveal Ethiopia's traditional healing treasures by using an efficient ensemble deep learning framework to identify the parts and uses of Ethiopian indigenous medicinal plants. EfficientNet, introduced by Tan and Le (Tan &Le, 2019) has gained attention for its outstanding performance in image processing tasks. The EfficientNet architecture indeed includes 8 different models, denoted as EfficientNetB0 through EfficientNetB7. Each model is characterized by varying depths, widths, and resolutions, with B0 being the baseline and B7 representing the largest and most complex variant. This range of models allows for flexibility in choosing the most suitable architecture based on the specific requirements of a given task, considering factors such as computational resources, model size, and the desired level of accuracy.

The architecture has been successfully applied in various fields, including medical imaging and plants classification tasks, demonstrating its potential for accurate and efficient deep learning-based identification systems. Our study builds on advancements in deep neural networks, specifically leveraging the state-of-the-art EfficientNet architecture, known for its exceptional performance in image identification and classification tasks. The proposed framework has the potential to significantly impact medicinal plants research in Ethiopia by providing a more efficient and precise method for identifying parts of indigenous medicinal plants and understanding their specific uses.

Pre-trained models expedite convergence during fine-tuning, reducing computational time. This approach is particularly beneficial in scenarios with limited task-specific data or resource constraints, offering a time-efficient and effective solution for various machine learning applications. In our study, we employed ensemble learning by combining the results from EfficientNetB0, EfficientNetB2, and EfficientNetB4 to enhance performance, taking advantage of transfer learning. However, ensembles, which combine multiple models, inherently require more memory and computational resources than single models. Replicating these models within an

ensemble increases memory demands, necessitating independent storage of each model's parameters (Mhawi *et al.*, 2022;Yang,Lv, *et al.*, 2023).

Combining diverse model predictions introduces computational complexity through methods like averaging or voting. Training multiple individual models, each with unique parameters, adds to the resource-intensive nature of ensembles. Including diverse models, which vary in architecture or training data subsets, expands feature space exploration but increases computational demands. However, by incorporating transfer learning techniques, we improved the performance of our ensemble learning model while simultaneously reducing the need for extensive computational resources and data. This enhancement was achieved by using a pre-trained deep learning model and employing a majority-based voting mechanism within the ensemble learning framework to further boost overall performance. Figure-13 below provides a comprehensive explanation of the details of our ensemble learning-based model.

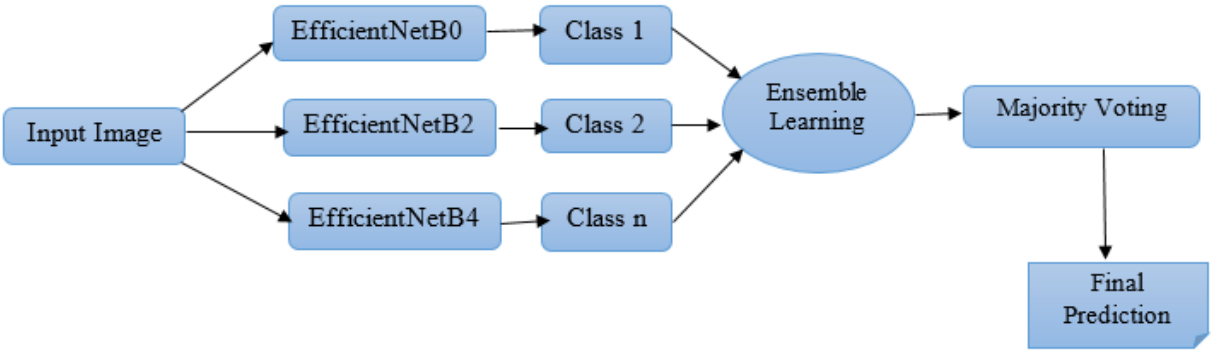


Figure 13 the proposed ensemble deep learning framework.

Mathematical formulation of the ensemble model involved a series of steps to articulate the proposed model. The subsequent process outlines the expression in mathematical terms:

Step 1: Individual Prediction: For a given input sample X_i, generate individual predictions from each of the base models:

$$Individual\ Prediction = \{hB_0\ (X_i), hB_2(X_i), hB_4(X_i)\} - - - -(4.1)$$

For a given input sample X_i, each EfficientNet model (M) provides an individual prediction. Each EfficientNet model independently produces a prediction for the input sample X_i. $hB_0\ (X_i)$, $hB_2\ (X_i)$, and $hB_4\ (X_i)$ represent the prediction of EfficinetNetB0, EfficientNetB2, and EfficientNetB4, respectively.

Step 2: Majority Voting Mechanisms: Use a majority voting mechanism (mode) to determine the final ensemble prediction:

$$Ensemble\ Prediction\ H(X_i) = \text{mode}\{hB_0(X_i), hB_2(X_i), hB_4(X_i)\} - - - - - - (\mathbf{4.2})$$

The ensemble prediction is determined by the mode (most frequent predictions) of the individual prediction. The model operation selects the prediction with the majority of votes among the M models.

Step 3: Handling Ties: In the case of a tie in the voting (multiple models have the same number of votes), a tie-breaking strategy is employed:

$$Handling\ Tie\ H(X_i) = argmin_j\{hB_0(X_i), hB_2(X_i), hB_4(X_i)\} - - - (\mathbf{4.3})$$

The choice of the tie-breaking ensures the final prediction is unambiguous and may involve selecting the predictions from the first model in the list. The strategies of tie-breaking involve selecting the prediction from the model with the minimum index (j) in the list. The $argmin_j$ function returns the index (j) corresponding to the minimum values among the tied predictions. Then, the final prediction ($H(X_i)$) is determined by selecting the prediction from the model with the minimum index, where M is the number of base models (EfficientNetB0, EfficientNetB2, and EfficientNetB4); $H_j(X_i)$ is the prediction of the jth base model for the ith input sample X_i; and $H(X_i)$ is the ensemble prediction for the ith input sample X_i.

4.3.1 Benchmark Models

In this subsection, we present the results of our experiments using pre-trained convolutional neural network (CNN) models, specifically the EfficientNetB0, EfficientNetB2, and EfficientNetB4 architectures, applied to classify leaf images from a custom dataset of Ethiopian indigenous medicinal plants species. These models were chosen for their innovative architectural approach, which features a unique scaling method that uniformly adjusts the network's critical dimensions, including depth, width, and resolution. This novel approach contrasts with traditional methods, which typically involve ad-hoc modifications to these architectural parameters.

The rationale for using EfficientNet lies in its proven superiority in achieving higher accuracy and computational efficiency compared to earlier ConvNets. For instance, the EfficientNet model attained a state-of-the-art top-1 accuracy of 84.3% on the ImageNet dataset, with a model size that is 8.4 times smaller and an inference speed that is 6.1 times faster than the best-performing existing

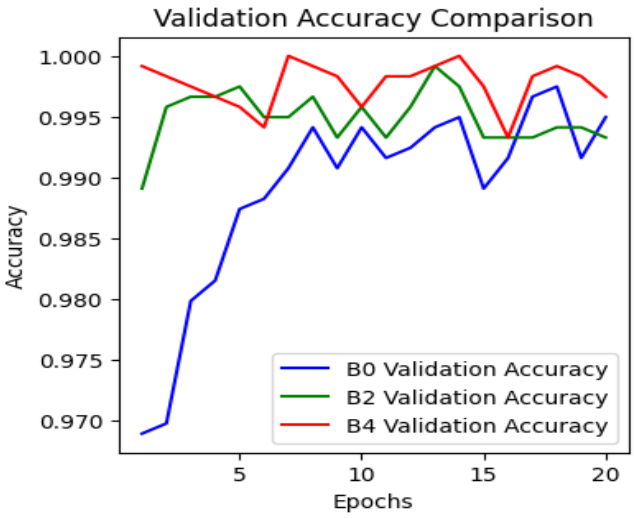
ConvNet (Tan &Le, 2019). Additionally, EfficientNet's versatility is evident in its success with transfer learning, achieving state-of-the-art accuracy not only on ImageNet but also on various datasets such as CIFAR-100 (91.7%), Flowers (98.8%), and three other transfer learning datasets (Tan &Le, 2019). These achievements have been made while using significantly fewer parameters than traditional approaches. To assess the practical applicability of EfficientNet models, we applied them to the task of identifying various parts and uses of Ethiopian indigenous medicinal plants. The results of this application are summarized in Table 5, providing insights into the models' accuracies in this specific domain.

Table 5 Summary of the performance of the benchmark model.

No	***Pre-trained Models***	***Accuracy***	***F1 Score***	***Precision***	***Recall***
1	EfficeintNetB0	0.9974	0.9974	0.9915	0.9974
2	EfficeintNetB2	0.9991	0.9991	0.9991	0.9991
3	EfficeintNetB4	0.9993	0.9991	0.9991	0.9991

Table-5 presents an overview of the test accuracy achieved by the individual pre-trained EfficientNet models that were trained on our custom dataset of Ethiopian indigenous medicinal plants species. Upon analyzing the results, we discovered that the pre-trained EfficientNetB0 model achieved outstanding performance, boasting a test accuracy score of 99.74%, an F1 Score of 99.74%, a precision rate of 99.15%, and a recall of 99.74%. These metrics collectively indicate highly accurate predictions, signifying a successful outcome. EfficientNetB2 demonstrated a similar level of excellence, achieving accuracy, F1-score, precision, and recall rates of 99.91%, highlighting its robust performance. Remarkably, the EfficientNetB4 pre-trained model exhibited exceptional performance, attaining an accuracy of 99.93% and an F1 score, precision, and recall of 99.91%. These results indicate that the model correctly predicted the majority of instances across various classes, reflecting its remarkable accuracy and reliability. The recall values specifically underline the model’s proficiency in identifying instances from each class accurately. The F1 score, which serves as a balanced metric encapsulating both precision and recall, further confirms the overall effectiveness of these models in our analysis. Figure-14 illustrates that the

loss consistently diminishes as training unfolds, suggesting that the models become more proficient at minimizing prediction errors.



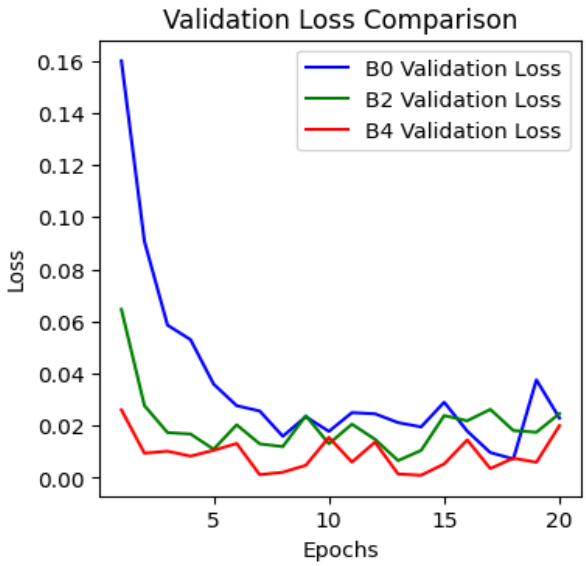


Figure 14 Validation Accuracy and Loss of benchmark models.

4.3.2 Performance Analysis of the Proposed Ensemble Learning Model

Ensemble learning leverages the strengths of diverse machine learning algorithms, training them independently on the same data with varied perspectives. Through the amalgamation of predictions using voting mechanisms like averaging, ensemble learning enhances overall performance, surpassing the capabilities of individual algorithms (Dong *et al.*, 2020). Ensemble learning encompasses both hard and soft techniques. In hard voting, the final class label for a sample is determined by the majority vote. On the other hand, soft ensemble learning employs a weighted probability approach to make predictions, considering the confidence levels of individual models in the ensemble (Dagnew *et al.*, 2021). The hard ensemble model, utilizing majority voting to amalgamate predictions from multiple models, yielded promising outcomes when implemented on the customized dataset of Ethiopian indigenous medicinal plants. Figure-19 visually illustrates the operational process of the proposed ensemble learning model. Within this figure, the ensemble node, as indicated, combines insights from three model components to yield predictions that surpass the accuracy achievable by any individual model, namely EfficientNetB0, EfficientNetB2, and EfficientNetB4. This fusion of predictions from diverse models significantly enhances overall accuracy. This ensemble approach integrates the pre-trained EfficientNetB0, EfficientNetB2, and

EfficientNetB4 models to address the identification challenges related to Ethiopian indigenous medicinal plants parts and uses. To provide a comprehensive assessment of our model's performance, please refer to table 6, which offers a detailed breakdown of results in comparison to the benchmark pre-trained models.

Table 6 Summary of the proposed models' prediction performance in comparison with benchmark models.

No	Pre-trained Models	Accuracy	F1 Score	Precision	Recall
1	EfficeintNetB0	0.9974	0.9974	0.9915	0.9974
2	EfficeintNetB2	0.9991	0.9991	0.9991	0.9991
3	EfficeintNetB4	0.9993	0.9991	0.9991	0.9991
4	Proposed model	0.9996	0.9992	0.9991	0.9992

As indicated in table-6, our proposed model attains an outstanding accuracy rate of 99.96%, surpassing the performance of individual pre-trained EfficientNet-based frameworks. The hard ensemble methodology consistently yields elevated F1 scores, precision, and recall for each class. Notably, achieving an accuracy of 99.96% underscores the ensemble's model's accurate identification of every instance in the custom Ethiopian dataset. Overall, the hard ensemble demonstrates robust performance across all evaluation metrics, including an accuracy of 99.96%, precision of 99.92%, recall of 99.92%, and F1 scores of 99.92%. This underscores its efficacy in accurately identifying the parts and corresponding uses of Ethiopian indigenous medicinal plants species. For a more in-depth understanding of the model's performance, table-7 provides a comprehensive comparison of validation and test accuracy between our proposed model and the benchmark models.

Table 7 Test and Validation accuracy scores of the proposed and benchmark models.

No	***Pre-Trained Models***	***Test Accuracy***	***Validation Accuracy***
1	EfficeintNetB0	0.9974	0.9958
2	EfficeintNetB2	0.9991	0.9992
3	EfficeintNetB4	0.9993	1.000
4	Proposed model	0.9996	0.9998

Table-7 displays the results of test and validation accuracy for both the benchmark models and our proposed model, focusing on the identification of parts and associated uses within Ethiopian indigenous medicinal plants species. An in-depth analysis of these findings reveals noteworthy insights. To begin with, the EfficientNetB0 benchmark model attains a commendable test accuracy score of 99.66% and a validation accuracy of 99.58%. Similarly, the EfficientNetB2 benchmark model stands out with a perfect test accuracy score of 99.91% and a robust validation accuracy of 99.92%. Additionally, the EfficientNetB4 benchmark model demonstrates high performance, achieving a test accuracy score of 99.93% and a flawless validation accuracy of 100%. Most notably, our proposed model, leveraging a majority vote-based ensemble method, indicates good and consistency results, achieving 99.96% test accuracy and 99.98% validation accuracy. This outcome underscores the reliability and precision of our ensemble deep learning model in accurately discerning Ethiopian medicinal plants species parts and their respective uses. These significant findings are further visually represented in figure-15, highlighting the performance of our proposed ensemble model. As depicted in figure-15, the accuracy results of the proposed ensemble model outshine those of EfficientNetB0, EfficientNetB2, and EfficientNetB4.

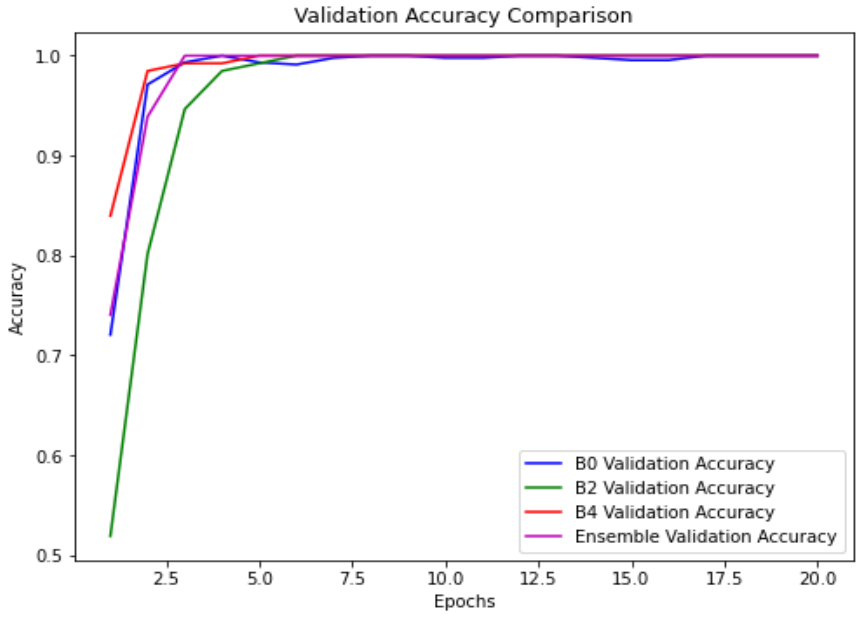
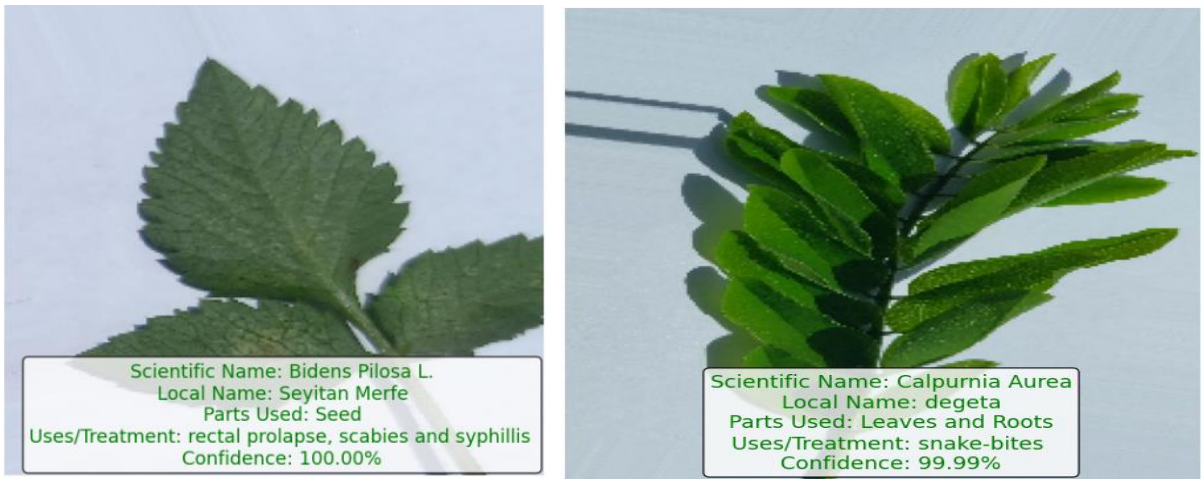


Figure 15 Validation accuracy performance of ensemble learning and benchmark models.

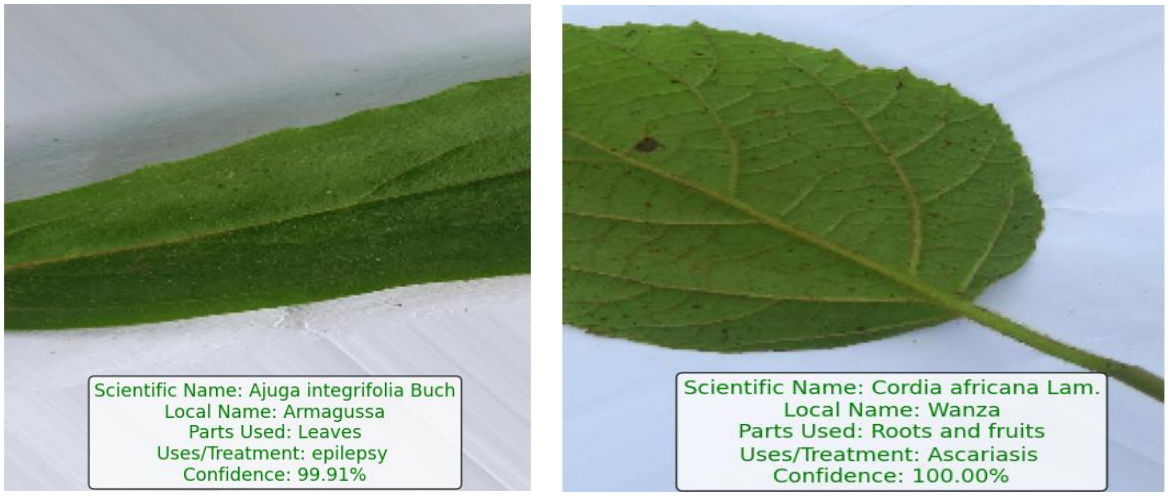
Our proposed ensemble learning approach, employing a majority vote system, has demonstrated significant test accuracy when applied to our dataset. As illustrated in figure-16, we conducted tests on a variety of indigenous Ethiopian medicinal plants, and the results are detailed below:

In figure-16A, we evaluated an Ethiopian medicinal plants known scientifically as *Bidens pilosa L.*. Traditionally, its seeds have been utilized to address rectal prolapse, scabies, and syphilis. Our proposed model achieved an impressive identification accuracy of 99.96% for this specific Ethiopian indigenous medicinal plants, with confidence intervals of 100%. Figure-16 B features *Calpurnia aurea*, along with its scientific nomenclature, and *Degeta* as its local name. This plants species is conventionally employed to treat snake bites, with its leaves and roots being the primary components used for this purpose. The ensemble learning model exhibited confident scores of 99.99% and accurate identification of this species. Figure-16 C indicates a medicinal plants scientifically known as *Ajuga integrifolia Buch,* locally referred to as *Armagussa*. Traditionally, this plants species is used to address epilepsy, with its leaves being predominantly used as the medicinal components. The proposed model demonstrated a notable confidence score of 99.91% for this plants species.

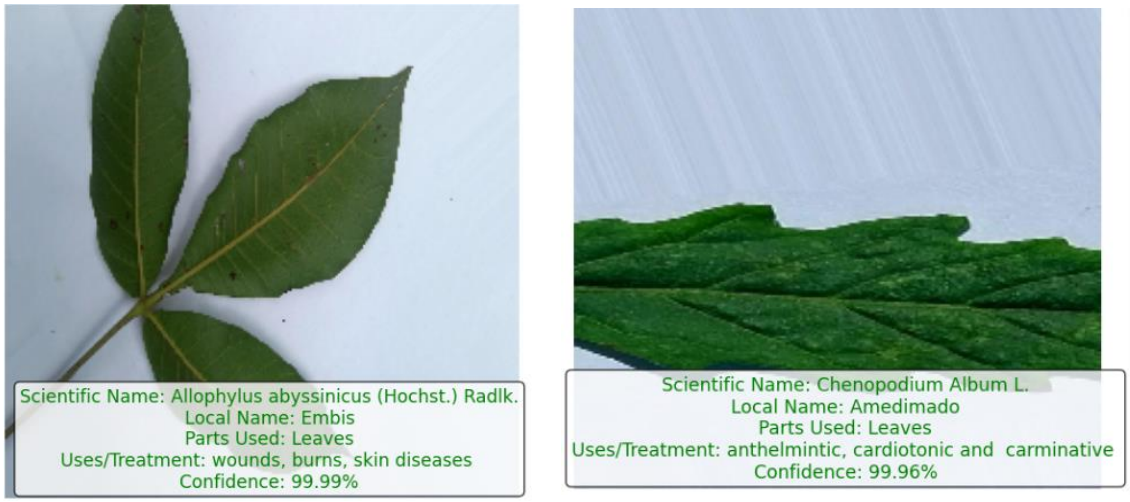
In Figure 16 D, we depict a medicinal plants scientifically recognized as *Cordia African Lam* and locally referred to as *Wanza*. This Ethiopian medicinal plants is traditionally employed to treat ascariasis, primarily using its roots and fruits. The ensemble learning model demonstrated flawless identification, achieving a 100% confidence score for this medicinal plants species. Figure-16 E highlights an indigenous medicinal plants known scientifically as *Allophylus abyssinicus (Hochst.) Radlk*. and locally as *Embis*. Traditionally, this plants is used for treating wounds, burns, and skin diseases, with its leaves used as the preferred remedy. The model yielded an impressive confidence score of 99.99% for this species. Lastly, in figure-16 F, we introduce a medicinal plants scientifically labeled as *Chenopodium album L.* and locally referred to as *Amedimado*. These medicinal plants are employed to treat anthelmintic, cardiotonic, and carminative conditions, with their leaves being the traditional remedy. The ensemble learning model confidently identified this species, with a confidence score of 99.96%. These findings highlight the efficiency of our suggested ensemble learning approach in precisely identifying diverse Ethiopian indigenous medicinal plants species, as demonstrated by the consistently high confidence scores obtained for each.



(A) (B)



(C) (D)



(E) (F)

Figure 16 Sample test data for Ethiopian indigenous medicinal plants parts and uses.

(A) Bidens pilosa L.: Traditionally used for addressing rectal prolapse, scabies, and syphilis. (B) Calpurnia aurea: Traditionally used for treatment of snake bites. (C) Ajuga integrifolia Buch: Traditional remedy for epilepsy. (D) Cordia africana Lam.: traditionally used for treating Ascariasis. (E) Allopyhylus abyssinicus: Traditionally applied for the management of skin diseases, wounds, and burns. (F) Chenopodium album L.: Traditionally used for anthelmintic, cardiotonic, and carminative purposes.

4.4. Interpretable Deep Learning for Ethiopian Indigenous Medicinal Plants Identification and Classification

In this Section, an interpretable deep learning model is developed for the identification and classification of Ethiopian indigenous medicinal plants species. To address computational resource constraints, a knowledge distillation approach is employed, aiming to create an efficient lightweight model through deep learning. The teacher-student architecture concept is applied to mitigate challenges associated with resource-intensive models. The implementation includes an ensemble knowledge distillation technique, enhancing the learning process of the student model. This approach not only improves the model's interpretability but also facilitates efficient knowledge transfer from more complex teacher models to lightweight student models. The ensemble knowledge distillation strategy contributes to the optimization of the student model's performance, making it well-suited for the identification and classification tasks related to Ethiopian indigenous medicinal plants, while concurrently managing computational costs.

In recent years, deep learning models have exhibited remarkable success in both industrial and academic domains, spanning a wide spectrum of applications, including computer vision(Alzubaidi *et al.*, 2021;Esteva *et al.*, 2021;Høye *et al.*, 2021) and natural language processing(Landolt *et al.*, 2021;Lauriola *et al.*, 2022;Torfi *et al.*, 2020). Insufficient training data is a common challenge in effectively training deep learning models for many applications(Alzubaidi *et al.*, 2021;Kang &Gwak, 2020). Deep learning also grapples with the inherent black-box nature and interpretability issues (Ekanayake *et al.*, 2022;Petch *et al.*, 2022;Singh *et al.*, 2020). Hence, there is a pressing need to create compact networks that exhibit strong generalization capabilities without an overwhelming reliance on extensive datasets. Moreover, addressing the black-box nature requires the integration of interpretable deep learning techniques. Additionally, it's crucial to acknowledge that most deep learning models come with

high computational demands, rendering them unsuitable for resource-constrained devices like mobile phones and embedded systems(Polson &Sokolov, 2020;Thompson *et al.*, 2020).

To address the constraints of computational resources, various model compression methodologies have been introduced. These techniques encompass methods such as low-rank factorization(Bejani &Ghatee, 2020;Scetbon *et al.*, 2021;Swaminathan *et al.*, 2020), parameter sharing(Dai *et al.*, 2020;Wang,Bai, *et al.*, 2020), and pruning(Hoefler *et al.*, 2021;Yeom *et al.*, 2021), as well as the adoption of transferred or compact convolutional filters(Tabernik *et al.*, 2020;Thomson *et al.*, 2020). These approaches aim to diminish the model's size while retaining comparable performance levels. The application of Low Rank Factorization (LRF) to convolution filters (Denton *et al.*, 2014) serves to accelerate the computationally intensive convolution operations and concurrently reduces the volume of convolutional kernel parameters. Moreover, the Low Rank Decomposition technique, such as Singular Value Decomposition (SVD), proves to be a stable and highly efficient method for compressing the weights within fully connected layers(Xue *et al.*, 2013). This is achieved by assessing the in formativeness of the model parameters.

Pruning is a widely adopted optimization strategy for neural networks today, even though the concept was recognized long before the ascent of deep learning into mainstream research. While earlier methods involved permanent removal of connections (Zhu &Gupta, 2017) or nodes (Alvarez &Salzmann, 2016;Cheng *et al.*, 2018) in a single operation, it was subsequently discovered that an iterative approach, involving removal and retraining, yields higher compression rates with less impact on accuracy (Gale *et al.*, 2019). The practice of pruning convolutional filters has unveiled intriguing insights into determining the optimal level of network pruning. Approaches rooted in the design of transferred/compact convolutional filters employ specialized structural configurations of convolutional filters to curtail the parameter space, thereby conserving storage and computational resources(Cheng *et al.*, 2017).

Knowledge distillation, within a teacher-student framework, is a potent deep learning technique. A pretrained teacher network generates probabilities to train a more compact student network, offering architectural flexibility. It employs relative class probabilities and data relationships, providing insights into sample similarities, departing from one-hot labels for mutually exclusive classes(Gou *et al.*, 2021). Despite its apparent simplicity, knowledge distillation has shown great promise in various fields, with image classification and identification standing out as prominent

applications. Knowledge distillation methods can be classified into three primary categories. First, in response-based distillation, the neural responses from the final output layer of the teacher network serve as the source of knowledge. For example, Gao, Teng, and Hong (Gao &An, 2021) explored the use of logits by reevaluating hyperparameter settings, while Xu, Chuanyun, et al. (Xu *et al.*, 2023) employed soft logit from the teacher model for knowledge distillation.

Second, feature-based distillation involves knowledge sources that encompass the output of intermediate layers, including feature maps and the last layer. This approach can be observed in the works of (Li,Guo, *et al.*, 2023;Zhang &Ma, 2020), which enhance the student model's performance by utilizing intermediate feature maps from the teacher network. Furthermore, the introduction of Attention Transfer (Komodakis &Zagoruyko, 2017) enables the transfer of attention maps, which represent the relative importance of layer activations. In relation-based distillation, knowledge transfer explores the relationships between data samples and various layers. Cheng et al. (Cheng *et al.*, 2021) proposed the distillation of the FSP matrix from adjacent layers of a teacher network pre-trained on ImageNet into a student network with the same structure to address the anomaly detection problem. These diverse knowledge distillation methods provide numerous options for effective and efficient knowledge transfer in the field of deep learning.

Interpretable deep learning techniques are used to overcome the challenges presented by the black-box nature of deep learning models(Li,Xiong,Li,Wu,Zhang,Liu,Bian,&Dou, 2022). These techniques aim to provide insight into the inner workings of these complex models, enabling users to understand how and why certain decisions are made. By employing interpretable deep learning methods, such as attention mechanisms(Gao *et al.*, 2021), saliency maps(Ismail *et al.*, 2021), gradient-based techniques(Zeng *et al.*, 2023) , LIME and SHAP(Aldughayfiq *et al.*, 2023), researchers and practitioners can gain valuable insights into model predictions. Interpretable deep learning refers to the practice of designing and training deep neural networks in a way that allows humans to understand, explain, and interpret the model's decisions and predictions(Li,Xiong,Li,Wu,Zhang,Liu,Bian,Dou, *et al.*, 2022;Zhang,Wang, *et al.*, 2020). Interpretable deep learning is essential in applications where model transparency and insight into decision-making are crucial, such as agriculture, healthcare, finance, and legal systems(Preuer *et al.*, 2019).

The increasing number of convolutional layers in neural networks has made it more challenging to understand how the deep learning models work, especially in complex fields like medicinal plants. The intricate scenes, various types of medicinal plants, and numerous environmental factors add to the complexity, emphasizing the critical need for improved deep learning model interpretability. Despite these challenges, there is an obvious lack of research's addressing this specific issue in the context of medicinal plants. In our investigation, we aimed to address this gap by enhancing the local interpretability of deep learning models by using Ethiopian indigenous medicinal plants species dataset.

This research presents an Interpretable Deep Learning approach, integrating a teacher-student framework and knowledge distillation techniques. Three highly performing teacher models and a compact pretrained model are employed to design the teacher-student model in this research, using knowledge distillation. These models have been meticulously designed and rigorously evaluated using our own datasets of Ethiopian Indigenous Medicinal plants species. To facilitate the knowledge distillation process, we employ a simplified student network represented by MobileNetV2, while opting for more advanced models such as VGG16, VGG19, and InceptionV3 as the teacher network for knowledge transfer due to their remarkable performance. MobileNetV2, a lightweight neural network by Google in 2018 (Sandler *et al.*, 2018), is designed for mobile and embeded devices. Its depth wise separable convolutions and efficient architecture make it ideal for resource-constrained environments, prioritizing speed and performance.

This study employs knowledge distillation mechanisms, which involve transferring knowledge from large, complex models to smaller, more practical ones suitable for real-world deployment. Knowledge distillation is performed more commonly on neural network models associated with complex architectures including several layers and model parameters. As deep learning has become prominent in various identification and classification tasks over the past decade, knowledge distillation techniques have gained traction for practical real-world applications(Tang *et al.*, 2020).

In our proposed approach, we used a multi-teacher-student framework, where a single student model learns from multiple teacher models. This multi-teacher distillation process involves the student model acquiring knowledge from several distinct teacher models (Zuchniak, 2023). By leveraging a collaborative teacher models, the student gains access to a diverse range of

knowledge, offering advantages over learning from a single teacher model alone. Combining knowledge from multiple teachers involves aggregating responses across all models, commonly through distillation techniques. The transferred knowledge typically includes logits and feature representations, providing the student with comprehensive insights from various perspectives.

A pretrained models (VGG16, VGG19 and InceptionV3) has been used as a teacher model and a light weight pretrained model MobileNetV2 are used as a baseline student model. In our experiments of previous experments, we found that VGG16, VGG19, and InceptionV3 demonstrate significant accuracy. These pretrained models have been extensively trained on vast datasets like ImageNet, minimizing the necessity for training from scratch and delivering strong performance with reduced data and computational resources (Chollet, 2017;Simonyan &Zisserman, 2014;Szegedy *et al.*, 2016).

MobileNetV2(Sandler *et al.*, 2018;Xiang *et al.*, 2019) is a convolutional neural network architecture optimized for deployment on mobile and embedded devices. It builds upon the original MobileNetV2 design by introducing inverted residuals, linear bottlenecks, and other improvements to enhance performance and efficiency. Key features include inverted residuals with linear bottlenecks, depth wise separable convolutions, and shortcut connections for gradient flow. The network utilizes expansion layers to increase representational power without significantly increasing computational cost. Additionally, MobileNetV2 offers flexibility through width and resolution multipliers, enabling scaling to different resource constraints(Peng *et al.*, 2023;Reza, 2023). These characteristics make MobileNetV2 well-suited for various computer vision tasks, including image classification, object detection, and semantic segmentation in resource-constrained environments such as mobile devices and edge computing platforms. Its efficient design and adaptability have contributed to its widespread adoption in practical applications requiring real-time processing and limited computational resources.

In our research, we introduce interpretable deep learning as a solution to the inherent black-box nature of deep learning models in identifying and classifying Ethiopian indigenous medicinal plants species. Our approach aims to enhance the transparency and comprehensibility of deep learning systems by providing users with insights into the reasoning behind their decisions. Leveraging LIME (Local Interpretable Model-agnostic Explanations), our interpretable deep learning model offers insight into how these complex models make predictions. Ultimately, our

research strives to mitigate the black-box nature of deep learning models, enabling users to trust and validate their outputs with greater confidence and understanding.

4.4.1. Proposed Architecture

In our proposed interpretable deep learning architecture, we leverage knowledge distillation techniques to transfer knowledge from multiple teacher models to a student model, fostering collaborative learning. This approach combines diverse knowledge sources, enriching the student model's understanding of Ethiopian medicinal plants species classification. Cosine similarity measures play a pivotal role in assessing how closely the predicted leaf images align with the trained dataset, ensuring accurate identification and classification between the teacher and student model. Lime interpretation provides valuable insights into the model's decision-making process, elucidating its reasoning behind specific predictions. This interpretability aspect not only enhances trust in the model's outcomes but also offers valuable feedback for refining its performance.

Overall, our architecture integrates knowledge distillation, similarity measures, and Lime interpretation to establish a robust framework for accurately identifying and classifying Ethiopian medicinal plants species while ensuring interpretability of the model's decisions. The proposed architecture is depicted in figure-17 below.

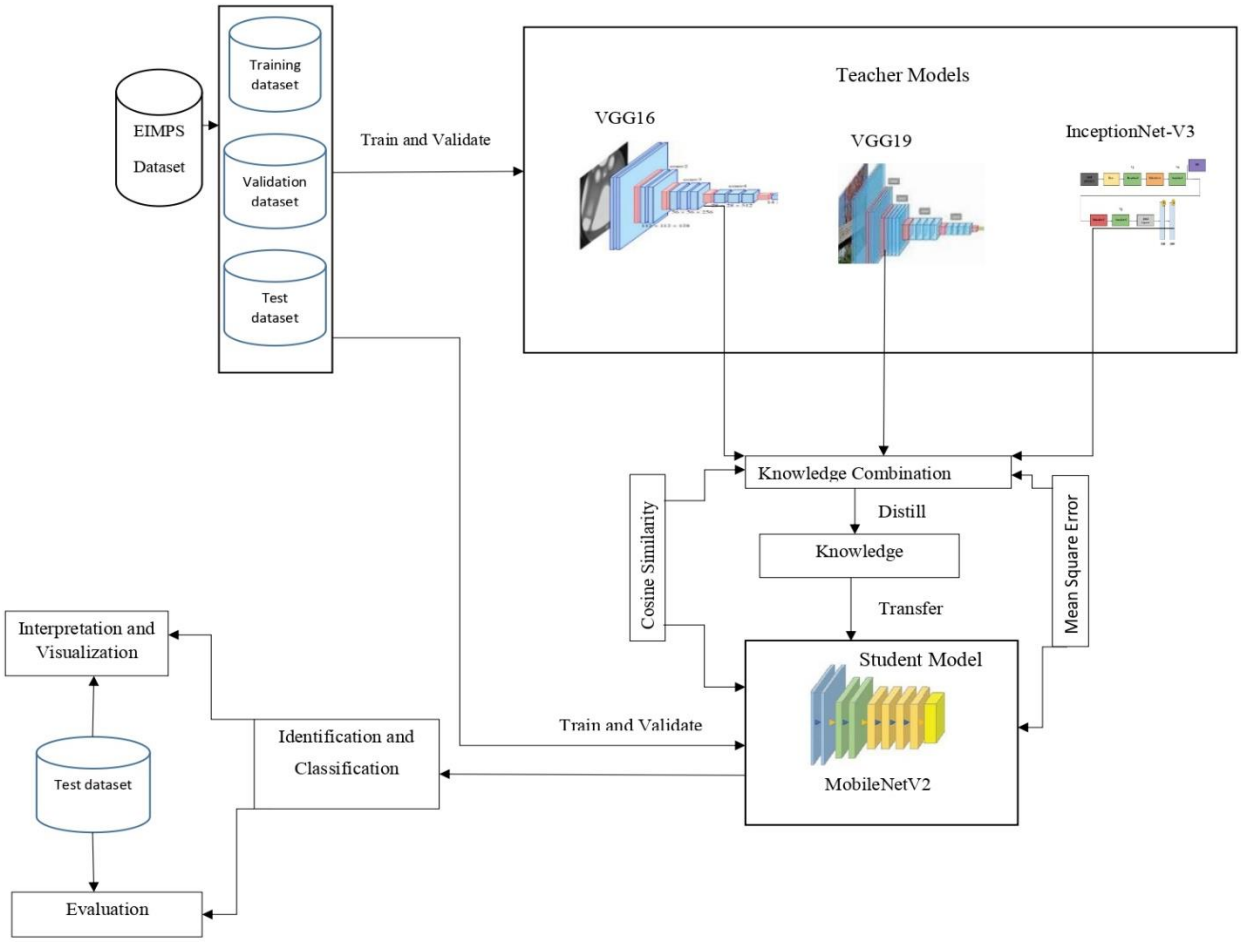


Figure 17 Proposed Architecture of Interpretable Distilled Student Model

4.4.1.1. Knowledge Distillation

In the conventional knowledge distillation (KD) approach, information transfer from a complex teacher model to a simpler student model relies on utilizing the softened class probabilities generated by the teacher(Bang *et al.*, 2021;Wang *et al.*, 2021). These softened scores, obtained through a softmax operation over the teacher's logits, serve as more informative targets for the student's training compared to traditional one-hot encoded labels. During training, the student learns not only to replicate the teacher's hard decisions but also to mimic its nuanced probabilistic reasoning. This is achieved by minimizing a loss function that quantifies the discrepancy between the student's predictions and the teacher's softened probabilities. By distilling knowledge in this manner, the student model gains insight into the underlying class distribution learned by the teacher, often resulting in improved performance and generalization capabilities.

The total loss of the student model's training is given by the sum of the student loss and the distillation loss, weighted by the alpha parameter, which controls the balance between the two losses.

$$\boldsymbol{total\_loss} = \boldsymbol{alpha} * \boldsymbol{student\_loss} + (\mathbf{1} - \boldsymbol{alpha}) * \boldsymbol{distillation\_loss} \text{ --------- } \mathbf{(4.4)}$$

Here, **student_loss** is the loss calculated between the student's predictions and the ground truth. The student loss, typically calculated using a categorical cross-entropy loss function, measures the discrepancy between the predictions made by the student model and the ground truth labels. The formula for the student loss Ls can be expressed as:

$$\boldsymbol{Ls} = -\frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} \sum_{\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{C}} \boldsymbol{Y_{ij}} . \log\left(\boldsymbol{P_{ij}}\right) \text{ ---------- } \mathbf{(4.5)}$$

Where,

- N is the total number of samples in the dataset.
- C is the number of classes.
- Y_{ij} is an indicator function that equals 1 if sample i belongs to class j, and 0 otherwise.
- P_{ij} is the predicted probability that sample i belongs to class j according to the student model.

Distillation_loss is the loss calculated between the soft student predictions and the soft teacher predictions(Yang,Zeng, *et al.*, 2023). The distillation loss, often computed using the Kullback-Leibler (KL) divergence between the soft student predictions and the soft teacher predictions, measures the disparity between the distributions of the predicted probabilities produced by the student and teacher models(Ji *et al.*, 2020;Kim *et al.*, 2021). The formula for the distillation loss $\boldsymbol{Ld}$ can be expressed as:

$$\boldsymbol{L_d} = \boldsymbol{D_{KL}}\left(\boldsymbol{P_{Teacher}} || \boldsymbol{P_{Student}}\right) \text{ ---------- } \mathbf{(4.6)}$$

Where:

- $\mathbf{D_{KL}}$ represents the Kullback-Leibler divergence.
- $\mathbf{P_{teacher}}$ is the soft distribution of predicted probabilities produced by the teacher model.
- $\mathbf{P_{student}}$ is the soft distribution of predicted probabilities produced by the student model.

The Kullback-Leibler (KL) divergence, also known as relative entropy, is a measure of how one probability distribution diverges from a second, expected probability distribution(Ji *et al.*, 2020). It is often used in statistics and information theory to quantify the difference between two probability distributions. The KL divergence between two probability distributions **P** and $\boldsymbol{Q}$ over the same probability space is defined as:

$$D_{KL}\,(P||Q) = \sum_i P(i).\,log\left(\frac{P(i)}{Q(i)}\right) \quad \text{---------- (4.7)}$$

Where i ranges over all possible outcomes. Note that the KL divergence is not symmetric, meaning $D_{KL}\,(P||Q)\,!= D_{KL}\,(Q||P)$

In the context of knowledge distillation, $P_{teacher}$ refers to the soft distribution of predicted probabilities produced by the teacher model, while $P_{student}$ refers to the soft distribution of predicted probabilities produced by the student model. Mathematically, the soft distribution $P_{teacher}$ produced by the teacher model can be represented as:

$$\boldsymbol{P_{teacher} = softmax(\frac{z_{teacher}}{T})} \quad \text{---------- (4.8)}$$

Similarly, the soft distribution $P_{student}$ produced by the student model can be represented as:

$$\boldsymbol{P_{student} = softmax(\frac{z_{student}}{T})} \text{---------- (4.9)}$$

Where, $z_{teacher}$ and $z_{student}$ represent the raw logits or scores produced by the teacher and student models, respectively. **T** denotes the temperature parameter, which controls the softness of the probability distributions. A higher temperature results in softer distributions, while a lower temperature leads to sharper distributions. The softmax function transforms the raw logits into a probability distribution over the classes, ensuring that the predicted probabilities sum up to one.

Generally, in the context of knowledge distillation, $P_{teacher}$ and $P_{student}$ represent the soft distributions of predicted probabilities generated by the teacher and student models, respectively. In knowledge distillation, soft distributions represent the predicted probabilities produced by both the teacher and student models for each class. Softening is achieved by applying a temperature parameter to the logits of the models before applying the softmax function, resulting in more informative probabilities that capture the model's uncertainty or ambiguity in its predictions. The

goal is for the student model to learn not only from the hard labels (ground truth) but also from the soft predictions of the teacher model. By aligning the soft distributions between the teacher and student models, the student can gain additional insights into the data distribution and potentially improve its performance, particularly when dealing with noisy or scarce ground truth labels. The Kullback-Leibler divergence loss function is commonly used to measure the difference between the soft distributions, guiding the student to mimic the behavior of the teacher model's predictions. Hence, knowledge distillation enables effective transfer of knowledge from a complex teacher model to a simpler student model, enhancing the student's performance through the integration of soft predictions.

The hyperparameter α plays a pivotal role in knowledge distillation, serving to balance the influence of two critical components: the student loss and the distillation loss (Chen *et al.*, 2022;Zhou,Song*, et al.,* 2021;Zong *et al.,* 2022). Spanning a range from 0 to 1, α dictates the relative emphasis placed on each loss during the training process. When α, the distillation loss exclusively shapes the total loss, leading the student model to learn solely from the soft predictions of the teacher model, disregarding the ground truth labels. This approach proves advantageous in scenarios with noisy ground truth labels or when the teacher model's predictions are deemed more reliable. Conversely, when α total loss is determined solely by the student loss, neglecting the distillation loss. Here, the student model learns exclusively from the ground truth labels, resembling conventional supervised learning. Intermediate values of α, ranging between 0 and 1, enable a blending of the student loss and the distillation loss, allowing practitioners to strike a balance between learning from ground truth labels and incorporating the distilled knowledge from the teacher model. By fine-tuning α, practitioners can tailor the training process to suit the characteristics of the dataset and achieve the desired behavior in the student model, thereby enhancing its performance and generalization capabilities.

The temperature parameter in knowledge distillation plays a crucial role in controlling the softness of probability distributions generated by the softmax function(Sun *et al.,* 2024). Instead of using hard labels directly as targets for training the student model, softened probabilities from the teacher model offer more nuanced information about class likelihoods. The softmax function is applied to the logits produced by the model, and the temperature parameter scales these logits before softmax

application. Mathematically, for a single logit z_i, the softmax function with temperature **T** is defined as:

$$\boldsymbol{Softmax_T(z_i) = \frac{e^{z_i/T}}{\sum_j e^{z_j/T}}} \text{ ---------- } \textbf{(4.10)}$$

A higher temperature value increases the range of logits, resulting in softer probability distributions where probabilities for all classes become more similar. Conversely, a lower temperature value sharpens the distribution, assigning higher probabilities to confident predictions. The temperature parameter balances exploration and exploitation during training. Adjusting it varies the degree of smoothing in probability distributions. A higher temperature encourages the student model to explore a wider range of possibilities, potentially capturing more nuanced data patterns. Conversely, a lower temperature emphasizes confident predictions, leading to more focused learning. Thus, the temperature parameter offers a flexible mechanism to adjust the balance between exploration and exploitation, crucial for effective knowledge distillation and student model training.

4.4.1.2.Multi-Teacher Distillation

In the conventional setup relying on a single teacher network (Zhou,Xu, *et al.*, 2021), a student network is trained to align with both the ground truth and the knowledge passed from the teacher network, which includes soft logits and intermediate feature embedding. For instance, MetaDistill (Zhou,Xu, *et al.*, 2021) enhances the knowledge transfer capability of a teacher network through meta-learning, introducing a pilot update mechanism to optimize teacher and student training as a bi-level optimization problem, facilitating improved knowledge transfer based on feedback from student learning. Structured Feature Transfer Network (Park *et al.*, 2021) adopts a modular approach, dividing teacher and student networks into multiple blocks, where both are simultaneously trained to minimize differences in feature representations and logits. Xu et al(Xu *et al.*, 2020). Explore self-supervised learning as an auxiliary task to extract richer knowledge from the teacher network, employing contrastive prediction to maximize agreement between data points and their transformed versions in latent space. To address potential performance degradation due to a significant gap between teacher and student (Mirzadeh *et al.*, 2020), introduce a teacher assistant, an intermediate-sized network facilitating multi-step teacher-student learning. Bergmann et al(Bergmann *et al.*, 2020), employ multiple student networks simultaneously supervised by a

powerful pre-trained teacher network, enabling accurate pixel-precise anomaly segmentation in high-resolution images. Consequently, anomalies can be detected when students fail to replicate the teacher's output. Inspired by recent advances, multiple teacher networks have been introduced in Knowledge Distillation, allowing a student to learn from various sources of knowledge simultaneously. Consequently, a student network can effectively learn diverse knowledge under the guidance of multiple teacher networks. In multi-teacher distillation, averaging the knowledge from multiple teachers (Nguyen,Lee, *et al.*, 2021;Papernot,Abadi, *et al.*, 2016;Tarvainen &Valpola, 2017;Yang *et al.*, 2020) is a common approach to incorporate diverse knowledge, although potentially sub-optimal as it assigns identical importance weight to each teacher. Hence, using a multi-teacher knowledge distillation approach presents an exciting opportunity to enrich student learning by harnessing insights from multiple teachers. Hence, we selected for a multi-teacher-student architecture to accurately identify Ethiopian indigenous medicinal plants species, leveraging interpretable deep learning techniques.

4.4.1.3. Cosine Similarity

Cosine similarity plays a pivotal role in knowledge distillation, particularly when comparing the learned representations of a teacher model to those of a student model (Ham *et al.*, 2023;Li,Lin, *et al.*, 2022;Sheng *et al.*, 2024). In this process, the teacher model, usually more complex and accurate, imparts its knowledge to a smaller, more lightweight student model. The core of this comparison lies in representing the models' parameters as vectors, with each parameter encapsulating the learned knowledge and decision-making processes of the model. By flattening the parameters into 1D arrays, the models' representations become amenable to cosine similarity calculations. This similarity measure quantifies the alignment between the parameter vectors of the teacher and student models, providing a numerical assessment of their agreement. A high cosine similarity indicates a strong correspondence between the learned representations, suggesting successful knowledge transfer from the teacher to the student. Conversely, a low cosine similarity implies misalignment or insufficient knowledge transfer, signaling areas for improvement in the distillation process. By monitoring changes in cosine similarity throughout the distillation process, practitioners gain insights into the dynamics of knowledge transfer and model alignment, guiding optimization strategies to enhance distillation outcomes. Thus, cosine

similarity serves as a critical evaluation metric in knowledge distillation, aiding in the assessment and improvement of model alignment and knowledge transfer effectiveness.

The cosine similarity between two vectors **a** and **b** is a measure of their similarity, computed as the cosine of the angle between them. Mathematically, the cosine similarity sim (**a**, **b**) is calculated as the dot product of the vectors **a** and **b** divided by the product of their Euclidean norms:

$$sim(a,b) = \frac{a.b}{\|a\|\|b\|} \text{---------- (4.11)}$$

Where:

a.b represents the dot product of the vectors a and b.

$\|a\| \; and \; \|b\|$ denote the Euclidean norms (lengths) of vectors a and b, respectively.

In the vector notation, the dot product of two vectors a and b is calculated as:

$$\boldsymbol{a.b = a_1b_1 + a_2b_2 + a_3b_3 + ---a_nb_n} \text{ ---------- (4.12)}$$

And the Euclidean norm (or magnitude) of a vector v is given by:

$$\boldsymbol{\|v\| = \sqrt{v1^2} + v2^2 + \cdots + vn^2} \text{ ---------- (4.13)}$$

4.4.1.4. Mean Square Error (MSE)

In knowledge distillation, Mean Square Error (MSE) is commonly used as a metric to measure the disparity between the soft probabilities predicted by the teacher model and those predicted by the student model (Dao *et al.*, 2021;Kim *et al.*, 2021;Takamoto *et al.*, 2020). This discrepancy helps gauge how well the student model is mimicking the behavior of the teacher model. By minimizing the MSE loss, the student model is trained to closely match the soft predictions of the teacher, thereby effectively distilling the knowledge from the teacher model. Minimizing MSE encourages the student model to capture the nuanced relationships and uncertainties present in the teacher's predictions, leading to improved performance and generalization capabilities in the student model. It quantifies the average squared difference between predicted values ($\overline{y_i}$) and their corresponding actual values (y_i) within a dataset. Mathematically, MSE involves squaring the difference between each predicted and actual value, summing these squared differences, and then averaging them

across all predictions (n). The resulting MSE value provides insights into prediction accuracy: lower MSE values indicate closer alignment between predictions and actual values, signifying superior model performance, while higher MSE values denote larger discrepancies, indicating poorer performance. Despite its simplicity and interpretability, MSE may be sensitive to outliers due to the squaring operation, warranting caution, especially in datasets with extreme values.

The Mean Squared Error (MSE) is calculated using the following formula:

$$\boldsymbol{MSE = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i - y_i)^2} \text{---------- (4.14)}$$

$(\overline{y_i})$ Represents the predicted value for the ith observation.

y_i Represents the actual (ground truth) value for the ith observation

n Total number of observations

4.4.1.5. LIME Interpretation

In the context of our designed knowledge distillation based interpretable deep learning, LIME (Local Interpretable Model-agnostic Explanations) interpretation provides a valuable tool for understanding the decisions made by the distilled student model. LIME (Dieber &Kirrane, 2020;Garreau &Luxburg, 2020;Garreau &Mardaoui, 2021) offers insight into how the complex student model arrives at its predictions by providing interpretable explanations for individual predictions. By examining local changes in the input data and their effects on the model's output, LIME helps elucidate the reasoning behind specific predictions. This interpretability aspect is crucial in KD, as it allows users to understand and trust the decisions made by the student model, despite its simplified architecture compared to the teacher models. Ultimately, LIME interpretation enhances the transparency and trustworthiness of the distilled student model's predictions, making it more suitable for practical deployment in real-world applications. Lime is chosen due to its computational efficiency(Islam *et al.*, 2021), instance-level interpretability(Islam *et al.*, 2021), model-agnostic nature(Natesan Ramamurthy *et al.*, 2020), and robustness in capturing feature interactions(Islam *et al.*, 2021). Grad-CAM, like LIME, provides model-agnostic explanations. However, unlike LIME, Grad-CAM is resource-intensive, requiring substantial computational power to calculate gradients and generate visual explanations, especially for large and deep

convolutional neural networks(Selvaraju *et al.*, 2017). Grad-CAM generates heat map that highlights important regions in an image, but it doesn't provide explicit features leave importance. More important for visual explanation rather than detail feature analysis. Hence, it is challenged in accurately determine class region coverage. This can limit its efficiency and scalability.

LIME also holds significant importance in the identification and classification tasks of Ethiopian medicinal plants species for various reasons. Firstly, it provides valuable insights into the decision-making process of complex medicinal plants species leaf image identification and classification models. By explaining individual predictions at the local level, LIME helps end users comprehend the underlying rationale behind the model's outputs. This transparency is essential for building trust and confidence in the model's predictions. Additionally, LIME's explanations can aid in identifying biases or anomalies within the model, offering valuable insights for improving its performance. By analyzing the features that contribute most too each prediction, it helps us to refine the model's architecture or training data to mitigate biases and enhance accuracy. Overall, LIME plays a crucial role in enhancing the interpretability of Ethiopian indigenous medicinal plants identification and classification models, thereby facilitating their deployment in real-world applications.

The study's results emphasize the successful use of interpretable deep learning models and transfer learning techniques to overcome the difficulties associated with identification and classification of Ethiopian indigenous medicinal plants species. The challenges associated with identifying and classifying Ethiopian medicinal plants species, including interpretability issues, are successfully addressed through the use of a custom dataset, knowledge distillation concepts and pre-trained deep learning models trained on the ImageNet dataset.

The teacher-student approach, known as knowledge distillation, transfers knowledge from a complex model (teacher) to a simpler one (student), improving computational efficiency without sacrificing performance. Multiple teachers can contribute to a single student, enhancing its learning. This compression maintains accuracy while preventing overfitting through regularization and enhancing generalization via knowledge transfer. The student model inherits acquired knowledge, facilitating interpretation. This method fosters progress, particularly in the identification and classification of Ethiopian indigenous medicinal plants, by consolidating valuable insights into more efficient models.

To improve model interpretability and enhance classification accuracy, a multi-teacher-student strategy, coupled with LIME interpretation and cosine similarity, are used. Furthermore, knowledge distillation techniques were applied to minimize computational resources and increase efficiency in the training phase.

Transfer learning demonstrated crucial in the methodology, offering significant advantages such as reduced training time and enhanced performance. Training deep learning models from scratch can be resource-intensive, especially on GPU machines. However, by using transfer learning, this research uses the knowledge already encoded in pre-trained models, speeding up learning significantly. Moreover, the ability to freeze specific model layers while fine-tuning others through transfer learning led to improved accuracy. Overall, the experimental results highlight the effectiveness of employing interpretable deep learning models and transfer learning techniques in tackling the challenges of identifying and classifying Ethiopian indigenous medicinal plants species.

4.4.2. Experimental Results of the Proposed Lightweight Interpretable Deep Learning

In this subsection, we present the outcomes of our experimentation involving an interpretable deep learning applied to the task of classifying leaf images from a custom dataset of Ethiopian indigenous medicinal plants species. In this work, VGG16, InceptionNetV3 and VGG19 pretrained models were selected as a teacher model. These models were selected due to their performance on the ImageNet dataset. We used MobileNetV2 for student model, a compact model which is suitable for mobile devices due to its less resource consumption and resource utilization.

After training the model for 20 epochs, remarkable performance metrics were attained by our distilled student model: a training accuracy of 99.83% and a validation accuracy of 99.16%. Moreover, the model demonstrated robustness on unseen data, achieving a commendable accuracy of 99.45% on the test dataset. Throughout the training process, the loss consistently decreased, indicating effective learning from the data.

4.4.3. Experimental Result Analysis

The performance of the distilled student model was evaluated using metrics such as accuracy, F1-score, recall, and precision. These metrics offer insights into the model's accuracy, its ability to

correctly classify positive instances, and its balance between precision and recall. Apart from accuracy, the model's performance was also assessed using test and validation accuracy. This broader evaluation provides an understanding of how well the model generalizes to new, unseen data beyond the training set. Additionally, the model's loss was measured using training, validation, and test losses. Analyzing these losses across different datasets helps determine whether the model effectively learns from the data and improves its predictive capabilities without overfitting or underfitting.

The experimental outcomes of the proposed interpretable deep learning model are briefly summarized in table 8 and 9, offering insights into the accuracies of the models within this specific domain. The results demonstrate the effectiveness of the interpretable deep learning approach in accurately identifying and classifying Ethiopian indigenous medicinal plants species, thereby highlighting its potential for practical applications in the domain.

Table 8 Overall performance of the student model after distillation

No	***Models***	***Accuracy***	***F1 Score***	***Precision***	***Recall***
1	Distilled Student model	0.9946	0.9945	0.9949	0.9946

Table-8 presents the overall performance of our designed distilled student model. We measured the accuracy, f1-score, precision and recall achieved by the designed models that were trained on our custom dataset of Ethiopian indigenous medicinal plants species. Upon careful analyzing the results, we discovered that the designed model achieved outstanding performance, boasting a test accuracy score of 99.46%, this indicates that the model reliably predicts the correct class for almost all instances in the dataset, highlighting its robust predictive capabilities and trustworthiness in tasks related to identification and classification of Ethiopian indigenous medicinal plants species. Our distilled student model also achieved an F1 Score of 99.45%, which serves as a balanced metric encapsulating both precision and recall. This further validates the overall effectiveness of these models in our analysis. Additionally, the consistent decrease in loss throughout training suggests that the distilled student models progressively improve at minimizing prediction errors. Such a high F1 score implies that the model effectively balances the trade-off between minimizing false positives and false negatives, which is very crucial for tasks where both precision and recall are important.

The designed model also achieves an impressive precision rate of 99.49%. Precision signifies the proportion of true positive predictions out of all positive predictions made by the model. Such a high precision score indicates that when the model predicts a positive class, it is highly likely to be correct. This translates to the model making very few false positive errors, which is advantageous in numerous identification and classification tasks. Moreover, the distilled student model records a recall of 99.46%. Recall measures the proportion of actual positive instances that the model correctly identifies. A recall score this high indicates that the model effectively captures the majority of positive instances in the dataset, thus minimizing false negatives. In real-world applications, a high recall score ensures that crucial positive instances are not overlooked or missed by the model. The balanced nature of these scores suggests that the model achieves high precision in identifying true positives while also capturing a significant proportion of actual positives (recall), crucial for effective identification and classification tasks. Overall, the exceptional performance of the Distilled Student model across these metrics underscores its efficacy and reliability in accurately classifying instances, highlighting its potential for practical use in real-world applications where accurate prediction is paramount.

Table 9 Overall training, validation and test accuracy and loss of student model

No	***Evaluation Parameter***	***Experimental Result***
1	Training accuracy	0.9983
2	Training Loss	0.0080
3	Validation accuracy	0.9916
4	Validation loss	0.0257
5	Test accuracy	0.9945
5	Test loss	0.011

Table-9 serves as a comprehensive summary of the performance metrics obtained from training, validation, and testing phases of our proposed distilled student model. This model is specifically tailored for the critical task of identifying and classifying Ethiopian indigenous medicinal plants species. Upon a thorough examination of these performance metrics, several key insights emerge. Firstly, during the training phase, the model exhibits an outstanding accuracy rate of 99.83%. This means that during the training process, the model accurately predicts the correct labels for approximately 99.83% of the training data. Accompanying this high accuracy is an impressively

low training loss of 0.0080. The training loss quantifies the difference between the predicted outputs and the actual labels in the training data, with a lower value indicating better performance. In this case, the low training loss suggests that the model effectively learns the underlying patterns and features of the training data, resulting in minimal errors during the training process.

As the model's capabilities extend beyond the training data to unseen instances, its performance remains robust. As it can be indicated in table-9, the validation accuracy of 99.16% indicates the model's ability to generalize well to new, unseen data. This is further supported by the corresponding validation loss of 0.0257, which, although slightly higher than the training loss, still reflects strong performance on the validation dataset. These findings affirm the reliability and precision of our distilled student model in accurately identifying and classifying Ethiopian medicinal plants species. Moreover, when subjected to evaluation on a separate test dataset, the model continues to demonstrate exceptional performance. With a test accuracy of 99.45% and a test loss of 0.011, the model showcases its effectiveness in accurately predicting labels for both familiar and unfamiliar data instances. These results not only validate the robustness of the model but also highlight its potential for practical application across various domains, where accurate identification and classification of medicinal plants species are crucial for research, conservation, and medicinal purposes.

4.4.4. Visualization Results using LIME

We used the widely adopted explainable AI technique, LIME, to generate local explanations for the predictions made by our distilled student model on the test sets. LIME generated segmentations of the images, highlighting the crucial regions for identification and classification of Ethiopian medicinal plants leaf images. While LIME successfully identified most of these features, some segmentations, particularly those of the outer regions of the leaf image, were inaccurate. Nevertheless, the model's important regions were identified in the majority of the images.

As LIME employs a perturbation approach for model interpretation, it involves altering different parts of an image and observing how these alterations affect the model's output. The core concept behind this approach is that perturbing important parts of an image significantly impacts the model's output, whereas perturbing unimportant parts has little effect. This underlying idea suggests that when important parts of an image are perturbed, the output of the model is strongly affected, while perturbations in unimportant areas have minimal impact on the model's output.

LIME creates simulated data around the original prediction using perturbation to generate explanations for individual predictions. This process involves random creation of simulated data, leading to variability in the explanations provided by LIME for a single prediction. We observed this instability while testing the XAI model multiple times, attempting to obtain explanations for a single prediction, and it is also evident in LIME's explanations for our model on Ethiopian indigenous medicinal plants species leaf images.

Our study results indicate that LIME is successful in producing explanations for the identification and classification tasks of Ethiopian indigenous medicinal plants species. LIME's model-agnostic nature allows it to explain any model without requiring access to its internal workings. To understand which parts of the interpretable input contribute to the prediction, we perturb the input around its neighborhood and observe the behavior of the model's predictions. We have explained correct predictions by employing a segmentation methods within the 'explain_instance ()' function to assess their impact.

The simulated data created through the process of perturbation is done randomly, which means that every time LIME is executed the explanations returned for a single prediction will be different. This instability in LIME was found while testing the XAI model numerous times trying to return explanations for a single prediction and can also be seen in LIME's explanations for deep learning predictions in the Ethiopian indigenous medicinal plants species leaf images. The results from our study signify that LIME is successful in producing stable explanations, its stability depends on the number of samples present in the dataset.

In figure-18, we tested various indigenous Ethiopian medicinal plants using our distilled student model, with interpretation results provided below. In figure-18 A, we evaluated an Ethiopian medicinal plants known scientifically as *Calpurnia Aurea*, locally referred to as *degeta*. Our proposed distilled model clearly reason out why the species is classified as Calpurnia Aurea. As it can be clearly shown in the figure, the feature similarity scores of the predicted class of the species is scored 99.04%, which indicates that the features of the predicted image and test images are very similar with test image. As evident from the figure, the feature similarity scores of the predicted class stand at 99.04%, underscoring the striking resemblance between the features of the predicted image and those of the test images. This high score suggests a close match between the predicted

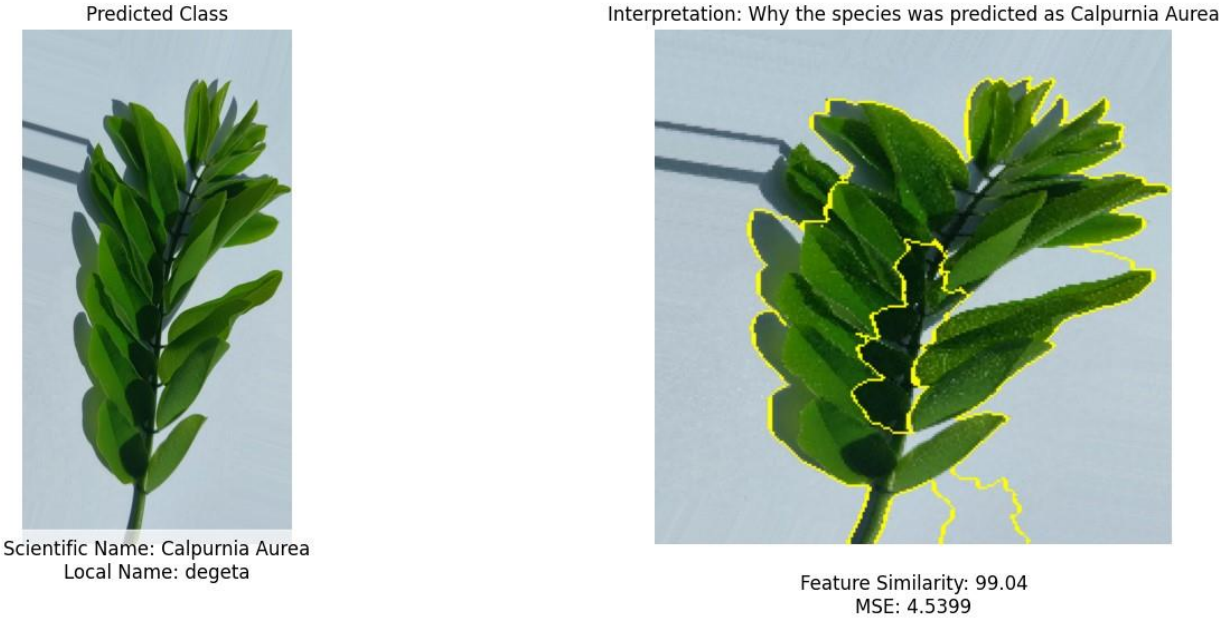
image and the test images, indicating robust consistency and accuracy in the model's classification. The notable similarity strengthens confidence in the model's ability to discern and classify instances accurately, affirming its reliability in identifying the species with a high degree of precision.

The mean score error (MSE) of the predicted image is 4.53%. An MSE of 4.539% indicates the model's predictions deviate, on average, by a fraction of total data variability. This level of error is very good, as lower MSE signifies better performance. The minimal deviation observed underscores the model's proficiency in capturing the intricate nuances of botanical features and making precise predictions. From a scientific perspective, the convergence of feature similarity scores and the low MSE values corroborate the efficacy of our proposed distilled model in Ethiopian medicinal plants species identification and classification tasks. These metrics serve as quantitative measures of the model's performance and provide valuable insights into its predictive capabilities. By accurately identifying and classifying indigenous Ethiopian medicinal plants species.

Figure-18 B the predicted class of Ethiopian indigenous medicinal plants species is classified as *Bidens Pilosa L.*. The distilled student model exhibited a feature similarity scores of 98.77% indicative of a substantial correspondence between the predicted class and the test images which is still good similarity scores is achieved the MSE scores of this species is also less than 5%, which is 4.538%. Moving on to figure-18 C, our investigation centered on the classification of *Ajuga integrifolia Buch*, locally referred to as *Armagussa*. Through meticulous analysis and interpretation, our proposed distilled student model showcased its ability to identify and classify this indigenous medicinal plants species with remarkable precision. Notably, the model achieved a notable feature similarity score of 99.36%, indicative of a strong resemblance between the predicted class and the test images. Additionally, the assessment of MSE scores further elucidated the model's predictive performance. With an MSE score of 4.15%, the model demonstrated a high degree of accuracy in its predictions, with minimal deviation observed from the actual values. This low MSE value reaffirms the model's proficiency in the identification and classification tasks, highlighting its reliability and efficacy in accurately identifying and classifying indigenous Ethiopian medicinal plants species.

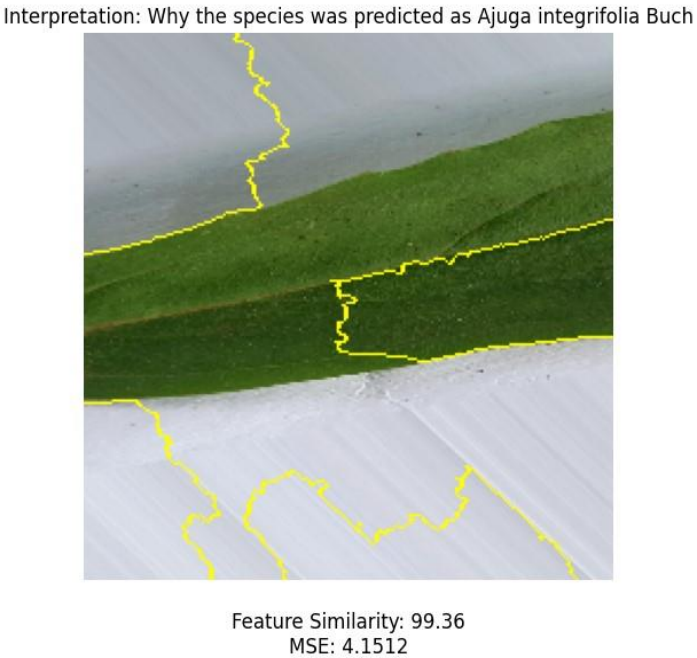
In figure-19 D, we depict a medicinal plants scientifically recognized as *Cordia African Lam* and locally referred to as *Wanza*. The distilled student model demonstrated flawless interpretation of achieving a 97.84% feature similarity score for this medicinal plants species with MSE scores of 4.534%. This score signifies a substantial correspondence between the predicted class and the test images, indicating the model's ability to accurately identify and classify Cordia African Lam based on its distinct botanical features. Furthermore, the low mean squared error (MSE) score of 4.534% emphasizes the precision of the model's predictions, with minimal deviation observed from the actual values. Figure-19 E highlights an indigenous medicinal plants known scientifically as *Allophylus abyssinicus (Hochst.) Radlk.* , and locally as *Embis*. The model yielded an impressive interpretable similarity score of 97.91% for this species with MSE scores of 4.54%.

These findings underscore the efficiency and effectiveness of our proposed interpretable deep learning approach in precisely identifying diverse Ethiopian indigenous medicinal plants species. The consistently high feature similarity scores and low MSE scores obtained across different plants species demonstrate the reliability and robustness of our model. Moreover, our approach addresses common challenges associated with black box models, enhancing interpretability and transparency in agricultural particularly in the identification and classification of Ethiopian indigenous medicinal plants species.



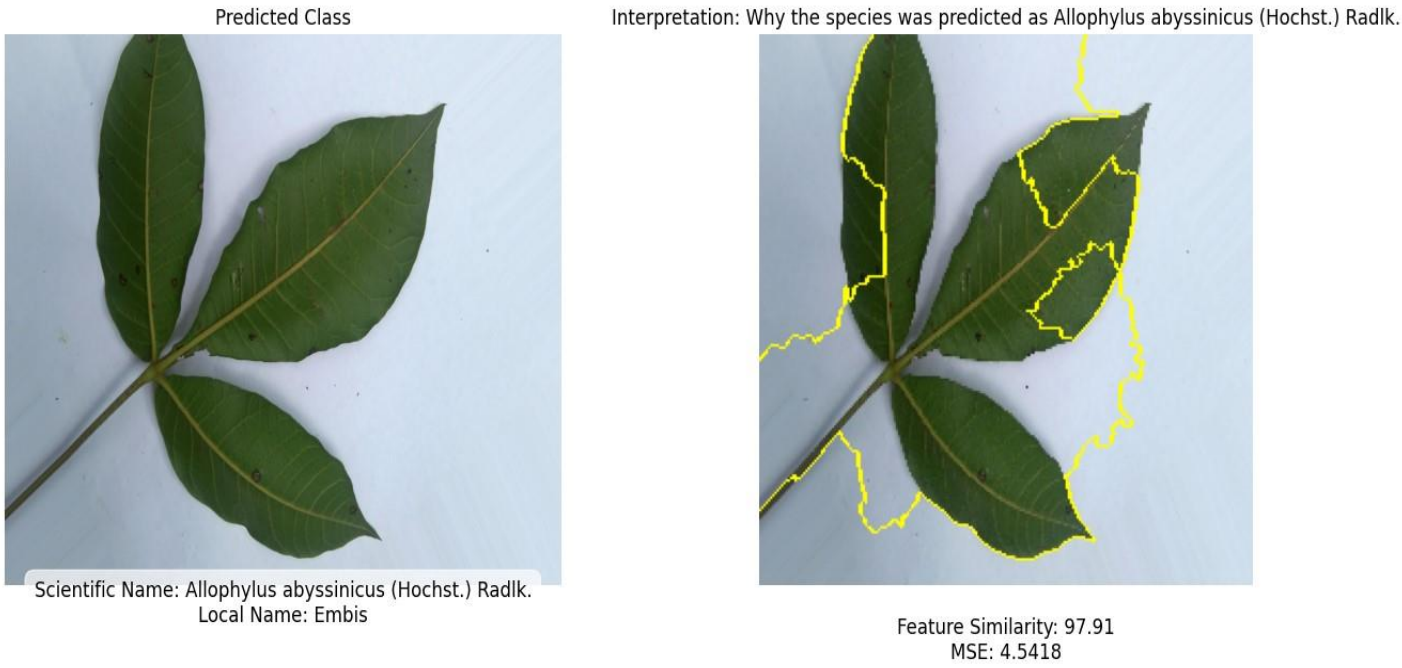
A) Calpurnia Aurea

B) *Bidens Pilosa L.*



C) *Ajuga Integrifolia Buch*

D) Cordia Africana Lam.



E) Allophylus abyssinicus(Hochst.Radik.

Figure 18 Sample test data for Ethiopian indigenous medicinal plants species

4.5. Discussion of Experimental Results

4.5.1. Deep Learning for Ethiopian Medicinal Plants Identification and Classification

Conserving indigenous medicinal plants is of utmost importance in traditional medicine(Ssenku *et al.*, 2022). To ensure their protection, we must embrace the latest technologies for identification and classification, involving local communities in the conservation process. Deep learning models have proven effectiveness in image classification, and transfer learning-based models which can simplify training complexity and data volume requirements.

For the classification of Ethiopian indigenous medicinal plants species using our custom dataset, we utilized four pre-trained models: VGG16, VGG19, Inception-V3, and Xception. These pretrained models are extensively trained on large scale datasets such as ImageNet, reducing the need for training from scratch and achieving good performance with less data and computation (Chollet, 2017;Simonyan &Zisserman, 2014;Szegedy *et al.*, 2016). Using pretrained deep learning models with CNNs provides a remarkable benefits, automatically extracting features from raw images (Abisha &Bharathi, 2023;Brahimi *et al.*, 2017;Haile *et al.*, 2022). The pretrained VGGNet model achieved the runner-up position in ImageNet(Sakib *et al.*, 2019;Simonyan &Zisserman, 2015). The Inception pretrained model introduced by Szegedy et al (2015), setting a new standard in ILSVRC14 competitions. The author (Chollet, 2017) developed Xception, a pretrained model which is known for its recognized for its innovative architecture. Therefore, we used these pretrained models in our work because they demonstrated perform very well in the ImageNet Challenges. Furthermore, we implemented a fine-tuning approach to enhance the performance of our transfer learning model, resulting in improved accuracy.

The experimental result demonstrates performance of pre-trained models and highlights their effectiveness in the identification and classification tasks of Ethiopian indigenous medicinal plants species. The experimental results of VGG19 outperforms VGG16, while Inception-V3 initially exhibits overfitting, alleviated through fine-tuning. Xception, though the performance, requires continuous parameter refinement. Fine-tuning substantially enhances accuracy, mitigating overfitting. VGG16 achieves a notable 92% validation accuracy, and VGG19 achieves an impressive 94% validation accuracy. Whereas Inception-V3 achieves 91%, and Xception reaches 87% in validation accuracy. Though remarkable performance were found in the experimental results, further adjustments and fine-tuning are advised for optimal model performance, for the identification and classification of Ethiopian indigenous medicinal plants species. The

experimental results also clearly demonstrate that pretrained models faced challenges including overfitting, under fitting, and training time complexity, which negatively impacted their performance. To tackle these issues, we implemented fine-tuning through adjustments of various hyperparameters. The outcomes revealed significant performance improvements in the fine-tuned models.

During our experiments, we measured the training complexity of the pre-trained models. Fine-tuning notably reduced time complexity, except for Inception-V3, which already had a relatively short execution time. By further adjusting hyperparameters, we could potentially minimize time complexity even further. To address the same challenge, the researchers employed two primary strategies(Azadnia *et al.*, 2022;Roopashree &Anitha, 2021). They expanded the dataset by increasing the number of images and applying image augmentation techniques to diversify the data and improve the model's generalization. Additionally, they made a crucial architectural change by replacing the traditional fully connected layer with the Global Average Pooling (GAP) layer, reducing the model's complexity and parameters. According to their result experimental results and analysis the GAP layer's computation of average feature maps per channel not only enhanced computational speed but also reduced overfitting risks, resulting in a more efficient and effective model during training and inference.

In other research studies, similar approaches have been applied to identify and classify medicinal plants species that are specific to certain countries. For instance, in the study by Pacifico et al.(2019),the researchers aimed to develop a non-invasive method for automatically classifying medicinal plants species using colour and texture features. They collected leaf images and used image processing techniques to extract features. By combining colour and texture features, their SVM model achieved impressive classification accuracy ranging from 76% to 93% for different plants species. In the research paper by Van Hieu et al.(2020), the focus was on identifying plants species native to Vietnam. They utilized a dataset of 28,046 images from 109 different indigenous plants species found in Vietnamese forests. To extract deep convolutional features, MobileNetV2, Inception ResnetV2, ResnetV2, and VGG16 models were employed. Experimental findings showed that VGG16 encountered overfitting issues, while Inception ResnetV2 took longer for evaluation. Although ResnetV2 achieved average accuracy, MobileNetV2 emerged as the superior model in plants recognition. With an impressive accuracy of 83.2%, MobileNetV2 is particularly well-suited for mobile applications due to its compactness.

The study by Malik et al. (2022), suggested the design of an automated real-time system for identifying medicinal plants species in the Borneo region. To achieve the plants species identification task, the study uses an EfficientNet-B1 pretrained deep learning model which was tested by a dataset comprising both public and private plants species information. The results indicated that their proposed pretrained model achieved Top-1 accuracies of 87% and 84% on the test sets for private and public datasets, respectively. Compared to other works (Borman *et al.*, 2022;Malik *et al.*, 2022;Pacifico *et al.*, 2019;Van Hieu &Hien, 2020), our approach significantly improves the classification and identification performance of Ethiopian indigenous medicinal plants species compared to other researchers. Achieving a remarkable accuracy of 95%, we conclude that fine-tuning proved to be a highly effective strategy in enhancing the performance of deep learning models. The fine-tuned models consistently outperformed their non-fine-tuned counterparts in terms of accuracy and loss. Additionally, fine-tuning reduced training time complexity, making it advantageous for situations with limited computational resources or when quicker model iteration is desired. Our study highlights the significance of preserving indigenous medicinal plants and demonstrates the value of advanced technologies and community involvement in this conservation effort.

In this experiment, the dataset comprises 1853 leaf images from 35 indigenous Ethiopian indigenous medicinal plants, potentially limiting the system's applicability to a broader range of Ethiopian indigenous medicinal plants species. To address this, future work could prioritize increasing the dataset of Ethiopian medicinal plants species and increasing the number of leaf images per species. Moreover, our approach depends completely on leaf images which may not be adequate for Ethiopian medicinal plants species identification and classification. Hence, other parts such as flowers, fruits, or roots must be considered in the future to develop a multi-modal identification and classification systems of Ethiopian indigenous medicinal plants species. The interpretability issues of pretrained models also represent a significant challenge. These limitations emphasize the need for continuous research and development in medicinal plants species identification and classification.

4.5.2. Identifying Ethiopian Indigenous Medicinal Plants Parts and Traditional Uses using Ensemble Learning

In this research, we carried out an extensive investigation into the identification of 44 Ethiopian indigenous medicinal plants parts and uses. Our dataset underwent meticulous preparation, with images partitioned into training, testing, and validation sets. Before initiating model training, preprocessing steps were implemented to enhance the features of the images. This involved eliminating background effects by capturing leaf images of the plants against a white background during the image-capturing process. Then, image processing techniques such as image normalization, resizing, manual removal of low-quality images, and cropping are applied to eliminate irrelevant sections. The purpose of these steps was to elevate the quality and augment the features of the images, laying a robust groundwork for subsequent model training. Image-augmentation techniques can be used to increase the number of images to minimize overfitting and underfitting challenges during training. Image augmentation is also a valuable tool for mitigating the challenges posed by dataset imbalances. Transfer learning emerged as a pivotal component of this study, offering several advantages. By leveraging pre-trained EfficientNet models originally trained on the ImageNet dataset, the study circumvented the time-consuming process of training models from scratch (Wu &Lin, 2022). This expedited the learning process while maintaining a high level of accuracy. Furthermore, leveraging transfer learning enabled fine-tuning of specific layers in the models, thereby enhancing their performance in our target domain. The evaluation of benchmark models, specifically EfficientNetB0, EfficientNetB2, and EfficientNetB4, uncovered their remarkable proficiency in precisely categorizing the different parts and uses of Ethiopian indigenous medicinal plants. For instance, the outcomes presented in Table 3 and Table 4 indicate that all these frameworks delivered nearly the same levels of accuracy, with EfficientNetB4 achieving the highest accuracy at 99.93%. Notably, EfficientNet adopts a technique known as the compound coefficient, which simplifies and enhances model scaling. Instead of making random adjustments to width, depth, or resolution, compound scaling uniformly modifies each dimension using a predefined set of scaling coefficients (Tan &Le, 2019).

EfficientNet-based U-Net models were used for segmenting CT images of kidney tumors, achieving impressive IoU scores (0.976 to 0.980) (Abdelrahman &Viriri, 2023). Notably, B7 excelled in kidney segmentation, while B4 performed best in tumor segmentation. Thus, the EfficientNet framework offers high accuracy in kidney disease segmentation and classification (Abdelrahman &Viriri, 2023). The EfficientNet-based models have gained recognition for their innovative scaling approach, which results in exceptional accuracy. This highlights the suitability

of EfficientNet models for transfer learning tasks related to computer vision activities. Moreover, this approach serves as a foundation for implementing the novel ensemble deep learning models. In this research, an ensemble deep learning model was developed, incorporating a majority voting mechanism to address the challenges associated with identifying the various parts and uses of Ethiopian indigenous medicinal plants species.

Through extensive experimentation and training, the ensemble deep learning model yielded promising results, attaining the maximum accuracy of 99.96% among the developed models and showcasing exceptional performance. Notably, an accuracy score of 99.98% was obtained in the validation phase compared to 99.96% in the test phase, surpassing the individual benchmark models. This accomplishment underscores the synergy derived from combining insights from multiple models, resulting in improved accuracy and reliability in identifying various Ethiopian indigenous medicinal plants species and their uses; this emphasizes the efficacy of ensemble learning in enhancing the overall system performance. Table 12 provides a comparison of the state-of-the-art benchmark models' performance with the proposed ensemble deep learning model. Overall, the result demonstrates a comparable accuracy result in identifying parts of Ethiopian indigenous medicinal plants species and their associated traditional uses, achieving an outstanding accuracy of 99.96% through the hard ensemble. These outcomes illustrate the efficacy of our proposed ensemble deep learning models, the success of the applied preprocessing techniques, and the advantages of ensemble learning. While our ensemble learning approach demonstrates strong performance, it occasionally misclassified a few images due to data-augmentation techniques that may have hidden crucial features, thereby affecting model accuracy. Therefore, to enhance the model's performance, it is essential to incorporate a variety of augmentation techniques.

To provide a comparative analysis, a study on Bangladeshi medicinal plants classification (Uddin *et al.*, 2023) employed an ensemble strategy. The researchers integrated VGG16, ResNet50, DenseNet201, InceptionV3, and Xception models. Their approach yielded a 98% accuracy using the hard ensemble method and an elevated accuracy of 99% with the soft ensemble configuration. Similarly, in the domain of Hepatitis C disease prediction, an ensemble learning model based on artificial intelligence was introduced in (Edeh *et al.*, 2022). This model leverages three components, namely MLP, Bayesian Network, and QUEST, achieving an impressive accuracy score of 95.59%. In (Joshi *et al.*, 2021), progressive transfer learning was utilized to classify leaf types, referencing datasets like Flavia, LeafSnap, and MalayaKew (MK-D1 and MK-D2). The

results revealed the following accuracy rates: Flavia 100%, MK-D1 99.05%, MK-D2 99.89%, and LeafSnap 97.95%. In another study (Abdollahi, 2022), CNNs, specifically the MobileNetV2 model, achieved 98.05% accuracy in identifying 30 Indian medicinal plants species from leaf images. Furthermore, (Roopashree &Anitha, 2021) utilized EfficientNetB4 models, both regular and pre-trained, to classify 38 plants species, recording a 99% accuracy. In (Pukhrambam *et al.*, 2022), a DenseNet-based CNN was utilized for medicinal plants classification in Manipur, yielding a 99.56% accuracy on the IMPPAT dataset. This research also introduced the Ensemble Deep Learning–Automatic Medicinal Leaf Identification (EDL–AMLI) classifier, which, based on weighted model outputs, surpassed established pre-trained models like MobileNetV2, InceptionV3, and ResNet50, achieving a remarkable 99.9% accuracy (Sachar &Kumar, 2022). Hence, our ensemble learning approach demonstrates commendable performance in comparison to other methodologies.

The experimental results of this work shows that individual EfficientNet architectures yield comparable results to their ensemble counterparts. It was also observed that experiments using VGG16, VGG19, and InceptionV3 produced results similar to those of the EfficientNet pretrained models in the context of Ethiopian medicinal plants species identification and classification. To design the lightweight interpretable deep learning model, a multi-teacher-student approach was implemented using ensemble techniques. However, the EfficientNet pretrained model consumes more memory due to its width and depth, making it computationally inefficient for resource-constrained environments. Therefore, VGG16, VGG19, and InceptionV3 were chosen for building the proposed lightweight interpretable deep learning model.

4.5.3. Proposed Lightweight Interpretable Deep Learning Model

In this research, we carried out an extensive investigation into the identification and classification of 44 Ethiopian indigenous medicinal plants species using our distilled student model. We have used a combined teacher model using knowledge distillation for designing our interpretable distilled student model. VGG16, VGG19 and InceptionNetV3 pretrained models has been used as a teacher model with the help of transfer learning for enhancing their performance in our target domain. Hence, these models provide a comparable accuracy for the identification and classification tasks using ImageNet Dataset. These models are combined together for transferring their knowledge to the student model using knowledge distillation concepts. Knowledge distillation is a deep learning technique that involves transferring the knowledge acquired by a

complex teacher model to a simpler student model. Rather than directly learning from the original training data, the student model is trained to mimic the outputs of the teacher model. This process typically uses the outputs of the teacher model, either as soft labels (probabilities) or as features, to guide the training of the student model. For our distilled student model, we have use soft labels from the combined teacher models. The primary motivation behind the use of knowledge distillation in our work is that the knowledge distillation concepts are the capabilities to compress the knowledge stored in the teacher model into a more compact form that can be efficiently used by the student model. This compression leads to several benefits, including model compression, improved generalization, transfer learning capabilities, and regularization effects. By distilling the knowledge from a large, pre-trained combined teacher models, in our case, VGG16, VGG19 and InceptionNetV3, knowledge distillation enables the creation of smaller, task-specific student models that retain much of the performance of the original teacher model while being more suitable for deployment in resource-constrained environments. We have used MobileNetV2 which is a pretrained deep learning model tailored for mobile and embedded vision tasks. MobileNetV2 strikes a balance between model complexity and computational cost. Its inverted residual blocks optimize feature extraction while minimizing the number of parameters, making it suitable for resource-constrained devices like smartphones and IoT gadgets. The addition of linear bottlenecks further enhances computational efficiency by combining lightweight depth wise and pointwise convolutions. Moreover, MobileNetV2 incorporates shortcut connections to aid gradient propagation, improving training stability and convergence. These design principles enable MobileNetV2 to achieve high accuracy in various visual recognition tasks while maintaining low computational overhead. Overall, MobileNetV2 represents a significant advancement in lightweight convolutional neural network architectures, empowering a wide range of applications in mobile and embedded systems. In order to design our distilled student model, we have used MobileNetV2 as a student model.

Through extensive experimentation and training, the distilled student model provides promising results, attaining the maximum accuracy of 99.83% among the developed models and indicating exceptional performance. Notably, an accuracy score of 99.16% was obtained in the validation phase compared to 99.45% in the test phase. However, according to the study in (Mengisti Berihu &Obeng, 2024), their student model achieves accuracy of 96.91% on the test dataset. The study uses a deep neural network and knowledge distillation, leveraging a dataset of 4,026 images from

8 species of Ethiopian medicinal plants leaves. Knowledge from a ResNet50 teacher model was transferred to a lightweight, 2-layer student model. The training process incorporated optimization strategies such as oversampling, data augmentation, and learning rate adjustment. To understand the model's decisions, LIME and Grad-CAM post-hoc explanation techniques were employed to highlight the influential image regions that contributed to the classification.

Similarly, the study (Nikam *et al.*, 2022) uses pretrained InceptionV3, Xception, and ResNet50 models on the VNPlants-200 dataset, each trained for 20 epochs. Results indicated that ResNet50 achieved 80% training accuracy and 62% test accuracy, InceptionV3 achieved 82% training accuracy and 71% test accuracy, while Xception achieved 86% training accuracy and 66% test accuracy. The study employed LIME to interpret the deep learning decisions, highlighting critical superpixels for each model's classification.

In comparison to other studies, our proposed Distilled Student Model provides outstanding performance metrics. Achieving a maximum accuracy of 99.83%, with validation and test accuracies of 99.16% and 99.45% respectively, our model demonstrates superior capability. In contrast, the study(Mengisti Berihu &Obeng, 2024) achieved 96.91% accuracy through a deep neural network and knowledge distillation, showcasing competitive but lower accuracy compared to ours. Meanwhile, (Nikam et al., 2022) uses of the VNPlants-200 dataset resulted in up to 86% training accuracy and 66% test accuracy, highlighting the benefits of knowledge distillation in enhancing model effectiveness. Both our approach and these studies employ LIME for interpretability, providing crucial insights into model decision-making processes.

Our study introduces a novel methodological approach featuring a multi-teacher framework and MobileNetV2, a pretrained model optimized for mobile and embedded systems. This strategic choice enhances our model's accuracy significantly, leveraging streamlined architecture and advanced optimization techniques. We further quantify the alignment between student and teacher models using similarity scores and Mean Squared Error (MSE), underscoring the efficacy and robustness of our approach in achieving high-performance outcomes. This experimental results of the proposed architecture underscores the synergy derived from combining insights from multiple teacher models, resulting in improved accuracy and reliability in identifying and classifying various Ethiopian indigenous medicinal plants species. Table-16 provides a promising results in training, validation, and test experiments of our proposed distilled student model. Overall, our

proposed distilled student model provides exceptional accuracy in identifying and classifying Ethiopian indigenous medicinal plants species. These outcomes illustrate the efficacy of our proposed distilled student model. The experimental results reveal insightful performance metrics for the trained model. Notably, the training loss, computed at 0.80%, signifies the degree of error between predicted and actual values within the training dataset, indicating a proficient grasp of underlying patterns. Meanwhile, the validation loss, standing at 2.57%, reflects the model's performance on unseen data, crucial for assessing its generalization capability and guarding against overfitting. A relatively low validation loss suggests a favorable ability to generalize beyond the training data. Finally, the test loss, registering at 1.1%, serves as the ultimate measure of the model's efficacy on entirely new data, affirming its reliability and applicability in real-world scenarios. Collectively, these results indicate a well-performing model with promising potential for practical deployment.

After classification, we have used LIME (Local Interpretable Model-agnostic Explanations) to try and explain our distilled student models to identify and classify the Ethiopian indigenous medicinal plants species correctly and explain which part of the image triggers the algorithm to make one prediction out of various other classes. We have incorporated this for the explainable AI (XAI) techniques enable a better understanding of the model's working process and highlight the relevant parts of the image on which the model has been working in order to mitigates issues of black box in deep learning. Using LIME we can identify the pixels of the image which play a more active part in classification.

LIME works by generating a simple, interpretable explanation for a prediction by training a linear model on perturbed versions of the input data(Ghosh *et al.*, 2023). The idea is to generate a large number of perturbed samples that are similar to the original input and then to train a simple model on these perturbed samples. The coefficients of this simple model can then be used to determine which features are most impactful for the prediction. The images shown in Figure 4 have been generated by LIME. The display only includes the super-pixels in the correct region section, with the contour of the super pixel highlighted and the background included as well. In the correlation section, the image displayed areas of super-pixels colored in green, indicating an increase in the probability that the image belongs to a specified class. The top fifteen positive features have been highlighted in green during experimentation. Limiting the display to the top five positive coefficients in LIME's right panel is a common default setting aimed at providing a concise

summary of the most influential features(Garreau &Mardaoui, 2021). This approach helps streamline the interpretability process by focusing on the most significant factors driving the model's predictions. However, adjusting this parameter to display the top 15 positive coefficients can offer a more comprehensive understanding of the model's behavior and the features contributing to its decisions. Expanding the display to 15 coefficients allows for a more detailed examination of the model's interpretability. It provides insights into a broader range of features that positively influence the model's predictions, offering a deeper understanding of the underlying patterns and relationships in the data. This increased granularity can be particularly useful in complex datasets or scenarios where multiple features interact to influence the model's output.

The figure also showed that the color warmth was not the same for both samples in Figure 4. Therefore, the model may have misinterpreted the medicinal plants leaf image scars because its significant region was not similar to that of the correctly classified Ethiopian indigenous medicinal plants species leaf images.

In this work, we used cosine similarity score between the student and teacher models measures the alignment of their parameter vectors. It quantifies how similar the directions of their parameter vectors are in a high-dimensional space, regardless of their magnitudes. In knowledge distillation, cosine similarity is favored for aligning soft targets from the teacher model with student model predictions. It's robust to temperature scaling, emphasizes relative similarity, and enables efficient optimization, leading to improved generalization by focusing on directional relationships between class probabilities (Ham *et al.*, 2023;Li,Lin*, et al.*, 2022;Sheng *et al.*, 2024). Mean square error (MSE) between the student and teacher models is used in this work, hence it is a key measure of how closely the student model's predictions match those of the teacher model. It quantifies the discrepancy between the probability distributions produced by the two models. A lower MSE indicates that the student model is effectively capturing the knowledge transferred from the teacher model. This alignment is crucial in knowledge distillation for effectively transferring knowledge from a complex teacher model to a simpler student model. As it can be depicted in figure-21, in our designed model both cosine similarity score and MSE were good enough in the experimental results. Both scores are clearly indicated in the experimental results and test validation phases.

CHAPTER FIVE
CONCLUSION AND FUTURE WORK

5.1.Chapter Overview

In this section, we synthesize the comprehensive concepts investigated throughout the research endeavor, summarizing the culmination of our findings. This synthesis encapsulates the culmination of the thesis, highlighting significant outcomes and implications. Additionally, meticulous recommendations for future avenues are provided, revealing crucial directions for advancing research in the field of identifying and classifying Ethiopian indigenous medicinal plants species specifically, as well as medicinal plants species more broadly. These future works aim to guide further exploration and development within the field.

5.2.Conclusion

This research addresses the critical issues of limited research efforts and dataset availability for identifying and classifying Ethiopian indigenous medicinal plants. A systematic review of global and Ethiopian studies revealed significant gaps, notably the scarcity of datasets and geographical disparities, with a pronounced lack of research focus in developing countries like Ethiopia. While deep learning has shown promise in this domain, concerns about model interpretability persist. To address these issues, an interpretable deep learning model has been developed and trained using an Ethiopian medicinal plants dataset.

Through extensive experimental analysis, various pretrained models, including VGG16, VGG19, Xception, and InceptionV3, were evaluated. Among these, VGG19 demonstrated superior performance. Fine-tuning these models reduced execution time and improved performance, highlighting their efficiency for feature extraction and addressing identification and classification challenges in the domain. Consequently, VGG16, VGG19, and InceptionV3 provided remarkable performance in the experimental analysis and were used as teacher models in our proposed distilled student model. For resource-constrained devices, MobileNetV2 is recommended and has been used as the student model.

In this study, an ensemble model was developed for identifying the parts and uses of Ethiopian medicinal plants using a majority-based ensemble of deep learning models. EfficientNet frameworks, including EfficientNetB0, EfficientNetB2, and EfficientNetB4, were used as

benchmark models and applied in a majority vote-based ensemble technique. The experimental results demonstrate that this model accurately identifies the plants parts and their traditional uses, indicating the effectiveness of ensemble learning.

The interpretability challenges of deep learning in this domain stem from the black-box nature of existing pretrained models. To tackle these challenges, an interpretable deep learning approach was developed, leveraging knowledge distillation and combining teacher's approaches. Local Interpretable Model-Agnostic Explanations (LIME) was also employed to elucidate the decisions made by the distilled student model. By visualizing the critical regions of input images contributing to predictions, LIME enhances model interpretability, mitigating the black-box nature of deep learning models. Additionally, metrics such as cosine similarity score and mean square error (MSE) were also used to gauge the alignment between the student and teacher models, affirming the effectiveness of knowledge transfer.

The findings of the study highlight the effectiveness of using interpretable deep learning models and transfer learning techniques to tackle the challenges associated with identifying and classifying Ethiopian indigenous medicinal plants species. Through the integration of a custom dataset and the incorporation of knowledge distillation concepts using pre-trained deep learning models, interpretability issues were successfully addressed. The teacher-student approach facilitated knowledge transfer from complex models to simpler ones, thereby improving computational efficiency without compromising performance.

5.3. Future Works

This research provides a solid basis for researchers interested in advancing knowledge in this area. However, it also highlights several limitations that need to be addressed. Researchers should direct their efforts towards generating country specific datasets, considering the diverse and indigenous nature of many indigenous medicinal plants. Moreover, optimizing computational resources is essential to effectively tackle real-time challenges in identifying and classifying indigenous medicinal plants species. A thorough comprehension of these domains will equip researchers with the essential tools needed to develop deep learning models effectively for indigenous medicinal plants identification and classification. Accordingly, further research in this domain is of utmost importance for the advancement of traditional medicine and pharmaceuticals, enhancing recognition, and supporting conservation goals.

Based on the experimental findings, it is recommended to further explore and refine the application of LIME in interpreting deep learning models, particularly in agricultural contexts such as the identification and classification of indigenous medicinal plants species. Future research should prioritize addressing inaccuracies in LIME's segmentations, especially in outer leaf regions, to enhance model interpretability. Additionally, investigating the impact of dataset size on LIME's stability and exploring methods to mitigate instability could improve the reliability of explanations provided. Furthermore, in the context of agricultural applications such as the identification and classification of Ethiopian indigenous medicinal plants species, future research into explainable AI techniques and their integration with deep learning models is crucial for promoting transparency and trust in model predictions. This is especially important in domains where interpretability is paramount for ensuring the reliability and acceptance of AI-driven solutions. Ultimately, in the realm of medicinal plants, particularly concerning the identification and classification of indigenous medicinal plants species, advancing the understanding and application of interpretability in AI techniques like LIME holds significant promise. By improving model interpretability, these techniques can facilitate informed decision-making processes in this complex, where transparency and trust in AI-driven solutions are vital for their successful implementation and acceptance.

In the context of medicinal plants research, future endeavors in knowledge distillation and teacher-student models offer immense potential for advancing model compression and efficiency. To drive progress in this field, researchers can explore refining distillation techniques tailored specifically for the task of identifying and classifying medicinal plants species. This may involve investigating novel loss functions, regularization methods, and optimization strategies optimized for distillation tasks in this domain. Additionally, exploring multi-teacher distillation approaches could prove beneficial. Transferring knowledge from multiple teacher models trained on diverse datasets or representing different botanical expertise could enrich the learning process, leading to more comprehensive and robust knowledge transfer for medicinal plants identification.

Dynamic distillation methods provide another avenue for exploration. By adaptively adjusting the distillation process during training, such as dynamically selecting teacher models or adjusting distillation parameters based on the complexity of medicinal plants features, researchers can enhance the effectiveness of the learning process. Tailoring distillation methods specifically to the

domain of medicinal plants, considering the unique characteristics and complexities of plants morphology and biochemistry, is crucial. This focused approach could result in significant performance gains in accurately identifying and classifying medicinal plants species. Ensuring the robustness and generalization capabilities of distilled models across various datasets and real-world scenarios is paramount in medicinal plants research. Addressing challenges related to dataset variability, limited labeled data, and diverse environmental conditions is essential to ensure the practical applicability of distilled models in this domain.

Moreover, research into privacy-preserving distillation techniques and hardware-aware distillation methods is essential to address emerging challenges in data privacy and resource optimization. By prioritizing these areas, future work in knowledge distillation and teacher-student models aims to push the boundaries of model compression, efficiency, and applicability in medicinal plants research, ultimately advancing the field of botanical science and supporting efforts in traditional medicine and pharmaceutical development.

Future studies should also broaden the scope beyond datasets gathered exclusively from Gullele Botanical Garden to include a more comprehensive representation of indigenous medicinal plants throughout Ethiopia. This entails addressing the natural, ecological, and cultural variations inherent in different regions of the country to ensure a broader coverage of diverse plants species. Moreover, future research should focus on understanding and documenting cultural and ethnicity-specific uses of indigenous medicinal plants parts. This includes exploring the variability in traditional healing practices among various societal groups to enhance the accuracy of identification and classification within different cultural contexts. A national traditional medicinal plants knowledge database and associated datasets should be designed and developed to comprehensively accommodate all Ethiopian indigenous medicinal plants species and their corresponding traditional knowledge. This initiative aims to preserve and conserve these valuable botanical resources, ensuring their sustainable use and protection cultural heritage.

Additionally, the development of user-friendly interfaces and mobile applications would streamline the practical application of our system in real-world situations. This would enable field researchers, herbalists, and the general public to effortlessly access and employ the system for the identification of medicinal plants parts and their uses and classifications of Ethiopian medicinal plants species.

REFERENCES

Abdelrahman et al. (2023). Efficientnet family u-net models for deep learning semantic segmentation of kidney tumors on ct images. Frontiers in Computer Science, 5, DOI: 10.3389/fcomp.2023.1235622

Abdollahi. (2022). Identification of medicinal plants in ardabil using deep learning: identification of medicinal plants using deep learning. Paper presented at the 2022 27th international computer conference, computer society of iran (csicc). (pp. 1-6). IEEE. DOI: 10.1109/CSICC55295.2022.9780493

Abera. (2014). Medicinal plants used in traditional medicine by oromo people, ghimbi district, southwest ethiopia. Journal of ethnobiology and ethnomedicine, 10, 1-15. DOI: https://DOI.org/10.1186/1746-4269-10-40.

Abisha et al. (2023). An hybrid feature extraction and classification using xception-rf for multiclass disease classification in plants leaves. Applied artificial intelligence, 37(1), 2176614. DOI:https://DOI.org/10.1080/08839514.2023.2176614

Adadi et al. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (xai). IEEE access, 6, 52138-52160. DOI: 10.1109/ACCESS.2018.2870052

Admasu et al. (2019). Ethiopian common medicinal plants: their parts and uses in traditional medicine-ecology and quality control. Plants science-structure, anatomy and physiology in plants cultured in vivo and in vitro, 21, 78-101.

Ahn et al. (2019). Variational information distillation for knowledge transfer. Paper presented at the proceedings of the ieee/cvf conference on computer vision and pattern recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9163-9171).

Aldughayfiq et al. (2023). Explainable ai for retinoblastoma diagnosis: interpreting deep learning models with lime and shap. Diagnostics, 13(11), 1932. DOI: https://DOI.org/10.3390/diagnostics13111932

Alemneh. (2021). Ethnobotanical study of plants used for human ailments in yilmana densa and quarit districts of west gojjam zone, amhara region, ethiopia. Biomed research international, 2021, 6615666. DOI: 10.1155/2021/6615666

Alkhulaifi et al. (2021). Knowledge distillation in deep learning and its applications. 7, e474.

Allen-zhu et al. (2020). Towards understanding ensemble, knowledge distillation and self-distillation in deep learning. PeerJ Computer Science, 7, e474. DOI: https://DOI.org/10.7717/peerj-cs.474

Altmann et al. (2010). Permutation importance: a corrected feature importance measure. Bioinformatics, 26(10), 1340-1347 DOI: https://DOI.org/10.1093/bioinformatics/btq134

Alvarez et al. (2016). Learning the number of neurons in deep networks. Advances in neural information processing systems, 29.

Alzubaidi et al. (2021). Review of deep learning: concepts, cnn architectures, challenges, applications, Journal of big Data, 8, 1-74. DOI: https://DOI.org/10.1186/s40537-021-00444-8

Amenu et al. (2016). Review on woody plants species of ethiopian high forests. Journal of Resources Development and Management, 27.

Amsalu et al. (2018). Use and conservation of medicinal plants by indigenous people of gozamin wereda, east gojjam zone of amhara region, ethiopia: an ethnobotanical approach. 2018. DOI: 10.1155/2018/2973513

Apley et al. (2020). Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society Series B: Statistical Methodology, 82(4), 1059-1086.82(4). DOI: https://DOI.org/10.1111/rssb.12377

Arrieta et al. (2020). Explainable artificial intelligence (xai): concepts, taxonomies, opportunities and challenges toward responsible ai. Information fusion, 58, 82-115. DOI: https://DOI.org/10.1016/j.inffus.2019.12.012

Asres et al. (2007). Identification and quantification of hepatotoxic pyrrolizidine alkaloids in the ethiopian medicinal plants solanecio gigas (asteraceae). International Journal of Pharmaceutical Sciences, 62(9), 709-713. DOI: https://DOI.org/10.1691/ph.2007.9.6766

Awas et al. (2010). Ethnobotany of berta and gumuz people in western ethiopia. Biodiversity, 11(3-4), 45-53. DOI: https://DOI.org/10.1080/14888386.2010.9712663

Awulachew et al. (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), 48-60. DOI: 10.11648/j.jdmp.20210703.11

Azadnia et al. (2022). An ai based approach for medicinal plants identification using deep cnn based on global average pooling. Agronomy, 12(11), 2723. DOI: https://DOI.org/10.3390/agronomy12112723

Azimi et al. (2021). A deep learning approach to measure stress level in plants due to nitrogen deficiency.Measurement, 173,108650.DOI:https://DOI.org/10.1016/j.measurement.2020.108650

Aziz et al. (2018). Traditional uses of medicinal plants practiced by the indigenous communities at mohmand agency, fata, pakistan. Journal of ethnobiology and ethnomedicine, 14(1), 1-16. DOI:10.1186/s13002-017-0204-5

Ba et al. (2014). Do deep nets really need to be deep? , Advances in neural information processing systems, 27.

Bach et al. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PloS one, 10(7), e0130140. DOI: https://DOI.org/10.1371/journal.pone.0130140

Balamurugan et al. (2023). Stage-wise categorization and prediction of diabetic retinopathy using ensemble learning and 2d-cnn. Intelligent Automation & Soft Computing, 36(1). DOI: 10.32604/iasc.2023.031661

Bang et al. (2021). Distilling from professors: enhancing the knowledge distillation of teachers. Information sciences, 576, 743-755. DOI: https://DOI.org/10.1016/j.ins.2021.08.020

Bejani et al. (2020). Adaptive low-rank factorization to regularize shallow and deep neural networks, arXiv preprint arXiv: 2005.01995. DOI: https://DOI.org/10.48550/arXiv.2005.01995

Bengio et al. (2013). Representation learning: a review and new perspectives. IEEE transactions on pattern analysis and machine intelligence, 35(8), 1798-1828. **DOI:** 10.1109/TPAMI.2013.50

Bergmann et al. (2020). Uninformed students: student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4183-4192).

Beygelzimer et al. (2015). Online gradient boosting. Advances in neural information processing systems, 28.

Bhujun et al. (2017). Biodiversity, drug discovery, and the future of global health: introducing the biodiversity to biomedicine consortium, a call to action. J glob health, 7(2), 020304. DOI:10.7189/jogh.07.020304

Binder et al. (2016). Layer-wise relevance propagation for neural networks with local renormalization layers. In Artificial Neural Networks and Machine Learning–ICANN 2016: 25th International Conference on Artificial Neural Networks, Barcelona, Spain, September 6-9, 2016, Proceedings, Part II 25 (pp. 63-71). Springer International Publishing. DOI: https://DOI.org/10.1007/978-3-319-44781-0_8

Bohdal et al. (2020). Flexible dataset distillation: learn labels instead of images. arXiv preprint arXiv:2006.08572.DOI: https://DOI.org/10.48550/arXiv.2006.08572

Borman et al. (2022). Classification of medicinal wild plants using radial basis function neural network with least mean square. In 2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS) (pp. 141-146). IEEE. DOI: 10.1109/ICE3IS56585.2022.10010072

Brahimi et al. (2017). Deep learning for tomato diseases: classification and symptoms visualization. Applied artificial intelligence, 31(4), 299-315. DOI:https://DOI.org/10.1080/08839514.2017.1315516

Breiman. (1996). Bagging predictors. Machine learning, 24, 123-140. DOI: https://DOI.org/10.1007/BF00058655

Brutzkus et al. (2019). Why do larger models generalize better? A theoretical perspective via the xor problem. In International Conference on Machine Learning (pp. 822-830). PMLR.

Bucilua et al. (2006). Model compression, in proceedings of the 12 th acm sigkdd international conference on knowledge discovery and data mining. New York, NY, USA, 3.

Bussmann et al. (2021). Clutia abyssinica jaub. & spach. P eraceae. Ethnobotany of the Mountain Regions of Africa, 305-308. DOI: https://DOI.org/10.1007/978-3-030-38386-2_43

Chan et al. (2015). Deep-plants: plants identification with convolutional neural networks. In 2015 IEEE international conference on image processing (ICIP) (pp. 452-456). IEEE. DOI: 10.1109/ICIP.2015.7350839

Chang et al. (2020). Improved deep learning-based approach for real-time plants species recognition on the farm. In 2020 12th International Symposium on Communication

Systems, Networks and Digital Signal Processing (CSNDSP) (pp. 1-5). IEEE. DOI: 10.1109/CSNDSP49049.2020.9249558

Charters et al. (2014). Eagle: a novel descriptor for identifying plants species using leaf lamina vascular features. In 2014 IEEE international conference on multimedia and expo workshops (ICMEW) (pp. 1-6). IEEE. DOI: 10.1109/ICMEW.2014.6890557

Chattopadhay et al. (2018). Grad-cam++: generalized gradient-based visual explanations for deep convolutional networks. In 2018 IEEE winter conference on applications of computer vision (WACV) (pp. 839-847). IEEE. DOI: 10.1109/WACV.2018.00097.

Chen et al. (2022). Knowledge distillation with the reused teacher classifier. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11933-11942).

Chen et al. (2021). Cross-layer distillation with semantic calibration. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 8, pp. 7028-7036). DOI: https://DOI.org/10.1609/aaai.v35i8.16865

Chen et al. (2019). Differences in rural and urban health information access and use. The Journal of Rural Health, 35(3), 405-417. DOI: https://DOI.org/10.1111/jrh.12335

Chen et al. (2018). Learning to explain: an information-theoretic perspective on model interpretation. Paper presented at the international conference on machine learning. In International conference on machine learning (pp. 883-892). PMLR

Chen et al. (2018). Deep boosting for image denoising. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 3-18)

Chen et al. (2019). Real-world image denoising with deep boosting. 3071-3087. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(12), 3071-3087.

Chen et al. (2017). Ensemble application of convolutional and recurrent neural networks for multi-label text categorization. In 2017 International joint conference on neural networks (IJCNN) (pp. 2377-2383). IEEE. DOI: 10.1109/IJCNN.2017.7966144

Chen et al. (2016). Conservation and sustainable use of medicinal plants: problems, progress, and prospects. Chin med, 11, 37. DOI:10.1186/s13020-016-0108-7

Cheng et al. (2020). Explaining knowledge distillation by quantifying the knowledge.

Cheng et al. (2017). A survey of model compression and acceleration for deep neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12925-12935).

Cheng et al. (2018). Model compression and acceleration for deep neural networks: the principles, progress, and challenges. 35(1), 126-136. DOI: 10.1109/MSP.2017.2765695

Cheng et al. (2021). Relation-based knowledge distillation for anomaly detection. In Pattern Recognition and Computer Vision: 4th Chinese Conference, PRCV 2021, Beijing, China, October 29–November 1, 2021, Proceedings, Part I 4 (pp. 105-116). Springer International Publishing. DOI: https://DOI.org/10.1007/978-3-030-88004-0_9

Cho et al. (2019). On the efficacy of knowledge distillation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 3967-3976).

Chollet. (2017). Xception: deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

Chollet. (2017). Xception: deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

Choromanska et al. (2015). The loss surfaces of multilayer networks. In Artificial intelligence and statistics (pp. 192-204). PMLR.

Chung et al. (2023). Single classifier vs. Ensemble machine learning approaches for mental health prediction. Brain informatics, 10(1), 1. DOI https://DOI.org/10.1186/s40708-022-00180-6

Cope et al. (2012). Plants species identification using digital morphometrics: a review. Expert Systems with Applications, 39(8), 7562-7573. DOI: https://DOI.org/10.1016/j.eswa.2012.01.073.

Cope et al. (2010). Plants texture classification using gabor co-occurrences. In Advances in Visual Computing: 6th International Symposium, ISVC 2010, Las Vegas, NV, USA, November 29–December 1, 2010, Proceedings, Part II 6 (pp. 669-677). Springer Berlin Heidelberg.

Cortes et al. (2017). Adanet: adaptive structural learning of artificial neural networks. In International conference on machine learning (pp. 874-883). PMLR

Cortes et al. (2014). Deep boosting. In International conference on machine learning (pp. 1179-1187). PMLR.

Daba et al. (2020). Ethnobotanical study on the medicinal value of selected five species in gullele botanic garden and its surroundings. Tropical Plants Research, 7(2), 285-295. DOI: 10.22271/tpr.2020.v7.i2.034

Dagnew et al. (2021). Ensemble learning-based classification of microarray cancer data on tree-based features. Cognitive Computation and Systems, 3(1), 48-60.DOI: https://DOI.org/10.1049/ccs2.12003

Dai et al. (2020). Parameters sharing in residual neural networks. Neural Processing Letters, 51(2), 1393-1410. DOI: https://DOI.org/10.1007/s11063-019-10143-4

Dao et al. (2021). Knowledge distillation as semiparametric inference. arXiv preprint arXiv:2104.09732. DOI: https://DOI.org/10.48550/arXiv.2104.09732

Davani et al. (2022). Dealing with disagreements: looking beyond the majority vote in subjective annotations. Transactions of the Association for Computational Linguistics, 10, 92-110.DOI: https://DOI.org/10.1162/tacl_a_00449

De luna et al. (2019). Size classification of tomato fruit using thresholding, machine learning, and deep learning techniques. AGRIVITA Journal of Agricultural Science, 41(3), 586-596.DOI: DOI: http://DOI.org/10.17503/agrivita.v41i3.2435

Demie et al. (2018). Ethnobotanical study of medicinal plants used by indigenous people in and around dirre sheikh hussein heritage site of south-eastern ethiopia. Journal of Ethnopharmacology 220 (2018): 87-93. DOI: https://DOI.org/10.1016/j.jep.2018.03.033

Deng et al. (2011). Deep convex net: a scalable architecture for speech pattern classification. In Twelfth annual conference of the international speech communication association.

Denton et al. (2014). Exploiting linear structure within convolutional networks for efficient evaluation. Advances in neural information processing systems, 27.

Dhurandhar et al. (2018). Explanations based on the missing: towards contrastive explanations with pertinent negatives. Towards contrastive explanations with pertinent negatives. Advances in neural information processing systems, 31.

Dieber et al. (2020). Why model why? Assessing the strengths and limitations of lime. arXiv preprint arXiv:2012.00093. DOI: https://DOI.org/10.48550/arXiv.2012.00093

Dileep et al. (2019). Ayurleaf: a deep learning approach for classification of medicinal plants. In TENCON 2019-2019 IEEE Region 10 Conference (TENCON) (pp. 321-325). IEEE. DOI: 10.1109/TENCON.2019.8929394.

Ding et al. (2019). Adaptive regularization of labels. Adaptive regularization of labels. arXiv preprint arXiv:1908.05474. DOI: https://DOI.org/10.48550/arXiv.1908.05474.

Diwedi et al. (2023). Cnn-based medicinal plants identification and classification using optimized svm. Knowledge-Based Systems, 112147. DOI: https://DOI.org/10.1016/j.knosys.2024.112147

Dong et al. (2020). A survey on ensemble learning. Front Comput Sci 14: 241–258.

Doshi-velez et al. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.DOI: https://DOI.org/10.48550/arXiv.1702.08608

Duong trung et al. (2019). A combination of transfer learning and deep learning for medicinal plants classification. In Proceedings of the 2019 4th International Conference on Intelligent Information Technology (pp. 83-90). DOI: https://DOI.org/10.1145/3321454.332146

Edeh et al. (2022). Artificial intelligence-based ensemble learning model for prediction of hepatitis c disease. 10, 892371. Frontiers in Public Health, 10, 892371. DOI: https://DOI.org/10.3389/fpubh.2022.892371

Ekanayake et al. (2022). A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using shapley additive explanations (shap). 16, e01059. Case Studies in Construction Materials, 16, e01059. DOI: https://DOI.org/10.1016/j.cscm.2022.e01059

El sheikha. (2017). Medicinal plants: ethno-uses to biotechnology era. 1-38 Biotechnology and production of anti-cancer compounds, 1-38. DOI: https://DOI.org/10.1007/978-3-319-53880-8_1

Esteva et al. (2021). Deep learning-enabled medical computer vision. NPJ digital medicine. DOI: https://DOI.org/10.1038/s41746-020-00376-2

Ferentinos. (2018). Deep learning models for plants disease detection and diagnosis. DOI: https://DOI.org/10.1016/j.compag.2018.01.009.

Feyssa et al. (2011). Wild edible fruits of importance for human nutrition in semi-arid parts of east shewa zone, ethiopia: associated indigenous knowledge and implications to food security. Pakistan Journal of Nutrition, 10(1), 40-50.

Friedman. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232. DOI: https://www.jstor.org/stable/2699986.

Gale et al. (2019). The state of sparsity in deep neural networks. arXiv preprint arXiv:1902.09574. DOI: https://DOI.org/10.48550/arXiv.1902.09574.

Ganaie et al. (2022). Ensemble deep learning: a review. A review. Engineering Applications of Artificial Intelligence, 115, 105151. https://DOI.org/10.1016/j.engappai.2022.105151.

Ganguly et al. (2022). Bleafnet: a bonferroni mean operator based fusion of cnn models for plants identification using leaf image classification. Ecological Informatics, 69, 101585.DOI: https://DOI.org/10.1016/j.ecoinf.2022.101585

Gao et al. (2021). Rethinking logits-level knowledge distillation. In Proceedings of the 2021 10th International Conference on Computing and Pattern Recognition (pp. 283-289). DOI: https://DOI.org/10.1145/3497623.3497669

Gao et al. (2021). Interpretable deep learning model for building energy consumption prediction based on attention mechanism. Energy and Buildings, 252, 111379. DOI: https://DOI.org/10.1016/j.enbuild.2021.111379.

Garreau et al. (2020). Explaining the explainer: a first theoretical analysis of lime. In International conference on artificial intelligence and statistics (pp. 1287-1296). PMLR.

Garreau et al. (2021). What does lime really see in images? In International conference on machine learning (pp. 3620-3629). PMLR.

Getnet et al. (2016). Studies on traditional medicinal plants in ambagiorgis area of wogera district, amhara regional state, ethiopia. Int J Pure Appl Biosci, 4, 38-45. DOI: http://dx.DOI.org/10.18782/2320-7051.2240

Ghosh et al. (2023). Recognition of sunflower diseases using hybrid deep learning and its explainability with ai. Mathematics, 11(10), 2241. DOI: https://DOI.org/10.3390/math11102241

Giday et al. (2003). An ethnobotanical study of medicinal plants used by the zay people in ethiopia. Journal of ethnopharmacology, 85(1), 43-52. DOI: https://DOI.org/10.1016/S0378-8741(02)00359-8.

Giday et al. (2013). Ethnobotanical study of plants used in management of livestock health problems by afar people of ada'ar district, afar regional state, ethiopia. ournal of Ethnobiology and Ethnomedicine, 9, 1-10. DOI: https://DOI.org/10.1186/1746-4269-9-8

Gilpin et al. (2018). Explaining explanations: an overview of interpretability of machine learning. In 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA) (pp. 80-89). IEEE. DOI: 10.1109/DSAA.2018.00018

Goldstein et al. (2015). Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. journal of Computational and Graphical Statistics, 24(1), 44-65.DOI: https://DOI.org/10.1080/10618600.2014.907095.

Gordon et al. (2019). Explaining sequence-level knowledge distillation as data-augmentation for neural machine translation. arXiv preprint arXiv:1912.03334. DOI: https://DOI.org/10.48550/arXiv.1912.03334..

Gou et al. (2021). Knowledge distillation: a survey. International Journal of Computer Vision, 129(6), 1789-1819. DOI: https://DOI.org/10.1007/s11263-021-01453-z.

Guillen et al. (2023). Gradient tree boosting and the estimation of production frontiers. Expert Systems with Applications, 214, 119134. https://DOI.org/10.1016/j.eswa.2022.119134

Guo et al. (2016). Human protein subcellular localization with integrated source and multi-label ensemble classifier. Scientific Reports, 6(1), 28087. https://DOI.org/10.1038/srep28087.

Gurumoorthy et al. (2019). Efficient data representation by selecting prototypes with importance weights. In 2019 IEEE International Conference on Data Mining (ICDM) (pp. 260-269). IEEE. DOI: 10.1109/ICDM.2019.00036

Haile et al. (2022). Detection and classification of gastrointestinal disease using convolutional neural network and svm. Cogent engineering, 9(1), 2084878. https://DOI.org/10.1080/23311916.2022.2084878

Hajam et al. (2023). An effective ensemble convolutional learning model with fine-tuning for medicinal plants leaf identification. Information, 14(11), 618. DOI: https://DOI.org/10.3390/info14110618

Hall et al. (2015). Evaluation of features for leaf classification in challenging conditions. In 2015 IEEE Winter conference on applications of computer vision (pp. 797-804). IEEE. DOI: 10.1109/WACV.2015.111.

Ham et al. (2023). Cosine similarity knowledge distillation for individual class information transfer. Information Transfer. arXiv preprint arXiv:2311.14307. DOI: https://DOI.org/10.48550/arXiv.2311.14307

Han et al. (2016). Incremental boosting convolutional neural network for facial action unit recognition. Advances in neural information processing systems, 29.

Hang et al. (2023). Deep stacked least square support matrix machine with adaptive multi-layer transfer for eeg classification. Biomedical Signal Processing and Control, 82, 104579. DOI: https://DOI.org/10.1016/j.bspc.2023.104579

He et al. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Heo et al. (2019). Knowledge transfer via distillation of activation boundaries formed by hidden neurons. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, No. 01, pp. 3779-3787). DOI: https://DOI.org/10.1609/aaai.v33i01.33013779

Hevner et al. (2008). Design science in information systems research. Management Information Systems Quarterly, 28(1), 6.

Hinton et al. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531. DOI: https://DOI.org/10.48550/arXiv.1503.02531

Ho et al. (2020). Utilizing knowledge distillation in deep learning for classification of chest x-ray abnormalities. EEE access, 8, 160749-160761. DOI: 10.1109/ACCESS.2020.3020802

Hoefler et al. (2021). Sparsity in deep learning: pruning and growth for efficient inference and training in neural networks. Journal of Machine Learning Research, 22(241), 1-124.

Høye et al. (2021). Deep learning and computer vision will transform entomology. Proceedings of the National Academy of Sciences, 118(2), e2002545117. DOI: https://DOI.org/10.1073/pnas.200254511

Hridoy et al. (2022). Deep neural networks-based recognition of betel plants diseases by leaf image classification. Computational intelligence and neuroscience, 2016(1), 3289801. DOI: https://DOI.org/10.1155/2016/3289801

Huang et al. (2018). Learning deep resnet blocks sequentially using boosting theory. In International Conference on Machine Learning (pp. 2058-2067). PMLR.

Huang et al. (2013). Random features for kernel deep convex network. In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 3143-3147). IEEE. DOI: 10.1109/ICASSP.2013.6638237.

Huang et al. (2017). Like what you like: knowledge distill via neuron selectivity transfer. arXiv preprint arXiv:1707.01219. DOI: https://DOI.org/10.48550/arXiv.1707.01219.

Hutchinson et al. (2012). Tensor deep stacking networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1944-1957. DOI: 10.1109/TPAMI.2012.268.

Islam et al. (2021). Explainable artificial intelligence approaches: a survey. arXiv preprint arXiv:2101.09429. DOI: https://DOI.org/10.48550/arXiv.2101.09429.

Ismail et al. (2021). Improving deep learning interpretability by saliency guided training. Advances in Neural Information Processing Systems, 34, 26726-26739.

Ji et al. (2020). Kullback–leibler divergence metric learning. transactions on cybernetics, 52(4), 2047-2058. DOI: 10.1109/TCYB.2020.3008248

Ji et al. (2020). Knowledge distillation in wide neural networks: risk bound, data efficiency and imperfect teacher. Advances in Neural Information Processing Systems, 33, 20823-20833.

Joshi et al. (2021). Progressive transfer learning approach for identifying the leaf type by optimizing network parameters. 53(5). Neural Processing Letters, 53(5), 3653-3676. DOI: https://DOI.org/10.1007/s11063-021-10521-x

Ju et al. (2018). The relative performance of ensemble methods with deep convolutional neural networks for image classification. Journal of applied statistics, 45(15), 2800-2818. DOI: https://DOI.org/10.1080/02664763.2018.1441383.

Junsongduang et al. (2014). Karen and lawa medicinal plants use: uniformity or ethnic divergence. Journal of ethnopharmacology, 151(1),517-527.

Kale et al. (2023). Identification of ayurvedic leaves using deep learning. In 2023 International Conference on Communication, Circuits, and Systems (IC3S) (pp. 1-6). IEEE.DOI: 10.1109/IC3S57698.2023.10169388

Kalyoncu et al. (2015). Geometric leaf classification. Computer Vision and Image Understanding, 133, 102-109. DOI: https://DOI.org/10.1016/j.cviu.2014.11.001.

Kamilaris et al. (2018). Deep learning in agriculture: a survey. Computers and electronics in agriculture, 147, 70-90. DOI: https://DOI.org/10.1016/j.compag.2018.02.016

Kang et al. (2020). A novel deep learning model by stacking conditional restricted boltzmann machine and deep neural network. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 1316-1324). DOI: https://DOI.org/10.1145/3394486.3403184.

Kang et al. (2020). Ensemble learning of lightweight deep learning models using knowledge distillation for image classification. Mathematics, 8(10), 1652. DOI: https://DOI.org/10.3390/math8101652

Kavitha et al. (2023). Medicinal plants identification in real-time using deep learning model. SN Computer Science, 5(1), 73. DOI: https://DOI.org/10.1007/s42979-023-02398-5

Kibebew. (2001). The status and availability of oral and written knowledge on traditional health care on traditional health care in ethiopia. In Conservation and sustainable use of medicinal plants in Ethiopia, Proceedings of the National workshop (Vol. 28, pp. 107-119).

Kidane et al. (2014). Use and management of traditional medicinal plants by maale and ari ethnic communities in southern ethiopia. Journal of ethnobiology and ethnomedicine, 10, 1-15. DOI: https://DOI.org/10.1186/1746-4269-10-46

Kim et al. (2016). Examples are not enough, learn to criticize! Criticism for interpretability. Advances in neural information processing systems, 29.

Kim et al. (2017). Transferring knowledge to smaller network with class-distance loss. Workshop track - ICLR 2017

Kim et al. (2021). Comparing kullback-leibler divergence and mean squared error loss in knowledge distillation. arXiv preprint arXiv:2105.08919.DOI: https://DOI.org/10.48550/arXiv.2105.08919

Kim et al. (2018). Paraphrasing complex network: network compression via factor transfer. Advances in neural information processing systems, 3.

Kindermans et al. (2017). Learning how to explain neural networks: patternnet and patternattribution. arXiv preprint arXiv:1705.05598. DOI: https://DOI.org/10.48550/arXiv.1705.05598

Komodakis et al. (2017). Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer. arXiv preprint arXiv:1612.03928. DOI: https://DOI.org/10.48550/arXiv.1612.03928

Kostrikov et al. (2020). Image augmentation is all you need: regularizing deep reinforcement learning from pixels. arXiv preprint arXiv:2004.13649.DOI: https://DOI.org/10.48550/arXiv.2004.13649

Krizhevsky et al. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

Kumar et al. (2012). Leafsnap: a computer vision system for automatic plants species identification. Paper presented at the european conference on computer vision. In Computer Vision–ECCV 2012: 12th European Conference on Computer Vision,

Florence, Italy, October 7-13, 2012, Proceedings, Part II 12 (pp. 502-516). Springer Berlin Heidelberg. DOI:https://DOI.org/10.1007/978-3-642-33709-3_36

Kumar et al. (2021). Role of traditional ethnobotanical knowledge and indigenous communities in achieving sustainable development goals. Sustainability, 13(6), 3062. DOI: https://DOI.org/10.3390/su13063062

Kümmerer et al. (2014). Deep gaze i: boosting saliency prediction with feature maps trained on imagenet. arXiv preprint arXiv:1411.1045. DOI: https://DOI.org/10.48550/arXiv.1411.1045

Kunapuli. (2023). Ensemble methods for machine learning: simon and schuster.

Kuncheva et al. (2003). Limits on the majority vote accuracy in classifier fusion. Pattern Analysis & Applications, 6, 22-31. DOI: https://DOI.org/10.1007/s10044-002-0173-7

Landolt et al. (2021). A taxonomy for deep learning in natural language processing. In HICSS (pp. 1-10).

Larese et al. (2014). Automatic classification of legumes using leaf vein image features. Pattern Recognition, 47(1), 158-168.DOI: https://DOI.org/10.1016/j.patcog.2013.06.012

Lasri et al. (2023). Facial emotion recognition of deaf and hard-of-hearing students for engagement detection using deep learning. Education and Information Technologies, 28(4), 4069-4092. DOI: https://DOI.org/10.1007/s10639-022-11370-4.

Lauriola et al. (2022). An introduction to deep learning in natural language processing: models, techniques, and tools. Neurocomputing, 470, 443-456. DOI: https://DOI.org/10.1016/j.neucom.2021.05.103

Lee et al. (2017). How deep learning extracts and learns leaf features for plants classification. Pattern recognition, 71, 1-13. DOI: https://DOI.org/10.1016/j.patcog.2017.05.015

Lee et al. (2020). Self-supervised label augmentation via input transformations. In International conference on machine learning (pp. 5714-5724). PMLR.

Lee et al. (2018). Self-supervised knowledge distillation using singular value decomposition. In Proceedings of the European conference on computer vision (ECCV) (pp. 335-350).

Lee et al. (2021). Data preprocessing. Deep Learning for Hydrometeorology and Environmental Science, 21-25. DOI: https://DOI.org/10.1007/978-3-030-64777-3_3

Lee et al. (2019). Graph-based knowledge distillation by multi-head attention network. arXiv preprint arXiv:1907.02226. DOI: https://DOI.org/10.48550/arXiv.1907.02226

Li et al. (2015). Sparse deep stacking network for image classification. In Proceedings of the AAAI conference on artificial intelligence (Vol. 29, No. 1).DOI: https://DOI.org/10.1609/aaai.v29i1.9786.

Li et al. (2023). Rethinking feature-based knowledge distillation for face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 20156-20165).

Li et al. (2013). Multi-label ensemble based on variable pairwise constraint projection. Information Sciences, 222, 269-281. DOI: https://DOI.org/10.1016/j.ins.2012.07.066

Li et al. (2022). Distilling a powerful student model via online knowledge distillation. IEEE transactions on neural networks and learning systems, 34(11), 8743-8752. DOI: 10.1109/TNNLS.2022.3152732

Li et al. (2017). Semi-supervised ensemble dnn acoustic model training. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5270-5274). IEEE. DOI: 10.1109/ICASSP.2017.7953162

Li et al. (2023). Scconv: spatial and channel reconstruction convolution for feature redundancy. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6153-6162).

Li et al. (2019). Semi-supervised deep coupled ensemble learning with classification landmark exploration. IEEE Transactions on Image Processing, 29, 538-550.DOI: 10.1109/TIP.2019.2933724.

Li et al. (2022). Interpretable deep learning: interpretation, interpretability, trustworthiness, and beyond. Knowledge and information systems, 64(12), 3197-3234. DOI:10.1007/s10115-022-01756-8

Li et al. (2022). Interpretable deep learning: interpretation, interpretability, trustworthiness, and beyond. nformation Systems, 64(12), 3197-3234. DOI: https://DOI.org/10.1007/s10115-022-01756-8.

Li et al. (2019). A novel deep stacking least squares support vector machine for rolling bearing fault diagnosis. Computers in Industry, 110, 36-47. DOI: https://DOI.org/10.1016/j.compind.2019.05.005

Lin et al. (2019). Transfer learning based traffic sign recognition using inception-v3 model. Periodica Polytechnica Transportation Engineering, 47(3), 242-250. DOI: https://DOI.org/10.3311/PPtr.11480

Linardatos et al. (2020). Explainable ai: a review of machine learning interpretability methods. Entropy, 23(1), 18. DOI: https://DOI.org/10.3390/e23010018

Lingappa et al. (2023). Deep learning-based active contour technique with bagging and boosting algorithms hybrid approach for detecting bone cancer from mri scan images. Multimedia Tools and Applications, 82(23), 36363-36377. DOI: https://DOI.org/10.1007/s11042-023-14811-5

Lipton. (2018). The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. Queue, 16(3), 31-57.

Liu et al. (2019). Knowledge distillation via instance relationship graph. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7096-7104).

Liu et al. (2018). Ssel-ade: a semi-supervised ensemble learning framework for extracting adverse drug events from social media. Artificial intelligence in medicine, 84, 34-49. DOI: https://DOI.org/10.1016/j.artmed.2017.10.003.

Lopez-paz et al. (2015). Unifying distillation and privileged information. arXiv preprint arXiv:1511.03643. DOI: https://DOI.org/10.48550/arXiv.1511.03643.

Low et al. (2019). Stacking-based deep neural network: deep analytic network for pattern classification. IEEE Transactions on Cybernetics, 50(12), 5021-5034. DOI: 10.1109/TCYB.2019.2908387

Lu et al. (2020). Learning the relation between interested objects and aesthetic region for image cropping. IEEE Transactions on Multimedia, 23, 3618-3630. DOI: 10.1109/TMM.2020.3029882

Lulekal et al. (2008). An ethnobotanical study of medicinal plants in mana angetu district, southeastern ethiopia. Journal of ethnobiology and Ethnomedicine, 4, 1-10. DOI: https://DOI.org/10.1186/1746-4269-4-10

Lundberg et al. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

Luss et al. (2019). Generating contrastive explanations with monotonic attribute functions. arXiv preprint arXiv:1905.12698, 3.

Malik et al. (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), 1952. DOI: https://DOI.org/10.3390/plants11151952

Maroyi et al. (2018). A review of the ethnomedicinal uses, phytochemistry and pharmacological properties of ekebergia capensis sparr. Asian Journal of Pharmaceutical and Clinical Research, 11(10), 59-68.

Mengisti berihu et al. (2024). Explainable ai: leaf-based medicinal plants classification using knowledge distillation. In 44. GIL-Jahrestagung, Biodiversität fördern durch digitale Landwirtschaft (pp. 23-34). Gesellschaft für Informatik eV.

Mhawi et al. (2022). Advanced feature-selection-based hybrid ensemble learning algorithms for network intrusion detection systems. Symmetry, 14(7), 1461. DOI: https://DOI.org/10.3390/sym1407146

Mikołajczyk et al. (2018). Data augmentation for improving deep learning in image classification problem. In 2018 international interdisciplinary PhD workshop (IIPhDW) (pp. 117-122). IEEE.DOI: 10.1109/IIPHDW.2018.8388338

Miller. (2019). Explanation in artificial intelligence: insights from the social sciences. Artificial intelligence, 267, 1-38.DOI: https://DOI.org/10.1016/j.artint.2018.07.007

Mirzadeh et al. (2020). Improved knowledge distillation via teacher assistant. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 04, pp. 5191-5198). DOI: https://DOI.org/10.1609/aaai.v34i04.5963

Moghimi et al. (2016). Boosted convolutional neural networks. In BMVC (Vol. 5, p. 6).

Mohammed et al. (2023). A comprehensive review on ensemble deep learning: opportunities and challenges. Journal of King Saud University-Computer and Information Sciences, 35(2), 757-774. DOI: https://DOI.org/10.1016/j.jksuci.2023.01.014

Montavon et al. (2017). Explaining nonlinear classification decisions with deep taylor decomposition. Pattern recognition, 65, 211-222. DOI: https://DOI.org/10.1016/j.patcog.2016.11.008

Mosca et al. (2017). Deep incremental boosting. arXiv preprint arXiv:1708.03704. DOI: DOI: https://DOI.org/10.48550/arXiv.1708.03704

Mouine et al. (2012). Advanced shape context for plants species identification using leaf image retrieval. In Proceedings of the 2nd ACM international conference on multimedia retrieval (pp. 1-8). DOI: https://DOI.org/10.1145/2324796.2324853

Mounce et al. (2017). Ex situ conservation of plants diversity in the world's botanic gardens. Nature Plants, 3(10), 795-802. DOI: https://DOI.org/10.1038/s41477-017-0019-3

Mudrakarta et al. (2018). Did the model understand the question. arXiv preprint arXiv:1805.05492. DOI: https://DOI.org/10.48550/arXiv.1805.05492

Müller et al. (2019). When does label smoothing help? , Advances in neural information processing systems, 32.

Mulugeta et al. (2024). Deep learning for medicinal plants species classification and recognition: a systematic review. Frontiers in Plants Science, 14, 1286088. DOI: https://DOI.org/10.3389/fpls.2023.1286088

Murdoch et al. (2019). Definitions, methods, and applications in interpretable machine learning. Proceedings of the National Academy of Sciences, 116(44), 22071-22080. DOI: https://DOI.org/10.1073/pnas.1900654116

Naeem et al. (2021). The classification of medicinal plants leaves based on multispectral and texture feature using machine learning approach. Agronomy, 11(2), 263. DOI: https://DOI.org/10.3390/agronomy11020263

Nakach et al. (2023). Deep hybrid bagging ensembles for classifying histopathological breast cancer images. In ICAART (2) (pp. 289-300).

Naresh et al. (2016). Classification of medicinal plants: an approach using modified lbp with symbolic representation. Neurocomputing, 173, 1789-1797. DOI: https://DOI.org/10.1016/j.neucom.2015.08.090

Natesan ramamurthy et al. (2020). Model agnostic multilevel explanations. Advances in neural information processing systems, 33, 5968-5979.

Neto et al. (2006). Plants species identification using elliptic fourier leaf shape analysis. Computers and electronics in agriculture, 50(2), 121-134. DOI: https://DOI.org/10.1016/j.compag.2005.09.004

Nguyen et al. (2021). Evaluation of explainable artificial intelligence: shap, lime, and cam. In Proceedings of the FPT AI Conference (pp. 1-6).

Nguyen et al. (2021). Stochasticity and skip connection improve knowledge transfer. In 2020 28th European Signal Processing Conference (EUSIPCO) (pp. 1537-1541). IEEE. DOI: 10.23919/Eusipco47968.2020.9287227

Nigussie amsalu et al. (2018). Use and conservation of medicinal plants by indigenous people of gozamin wereda, east gojjam zone of amhara region, ethiopia: an ethnobotanical approach. DOI: 10.1155/2018/2973513

Nikam et al. (2022). Explainable approach for species identification using lime. In 2022 IEEE Bombay Section Signature Conference (IBSSC) (pp. 1-6). IEEE. DOI: 10.1109/IBSSC56953.2022.10037417

Opitz et al. (2017). Bier-boosting independent embeddings robustly. In Proceedings of the IEEE international conference on computer vision (pp. 5189-5198).

Pacifico et al. (2019). Automatic classification of medicinal plants species based on color and texture features. In 2019 8th Brazilian Conference on Intelligent Systems (BRACIS) (pp. 741-746). IEEE. DOI: 10.1109/BRACIS.2019.00133

Palangi et al. (2014). Recurrent deep-stacking networks for sequence classification. In 2014 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP) (pp. 510-514). IEEE. DOI: 10.1109/ChinaSIP.2014.6889295

Papernot et al. (2016). Semi-supervised knowledge transfer for deep learning from private training data. arXiv preprint arXiv:1610.05755. DOI: https://DOI.org/10.48550/arXiv.1610.05755.

Papernot et al. (2016). Distillation as a defense to adversarial perturbations against deep neural networks. In 2016 IEEE symposium on security and privacy (SP) (pp. 582-597). IEEE.DOI: 10.1109/SP.2016.41

Park et al. (2021). Learning student-friendly teacher networks for knowledge distillation. 34, 13292-13303. In 2016 IEEE symposium on security and privacy (SP) (pp. 582-597). IEEE.

Park et al. (2019). Relational knowledge distillation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 3967-3976).

Passalis et al. (2020). Heterogeneous knowledge distillation using information flow modeling. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2339-2348).

Passban et al. (2021). Alp-kd: attention-based layer projection for knowledge distillation. In Proceedings of the AAAI Conference on artificial intelligence (Vol. 35, No. 15, pp. 13657-13665). DOI: https://DOI.org/10.1609/aaai.v35i15.17610

Pathak et al. (2022). Deep transfer learning based classification model for covid-19 disease. Irbm, 43(2), 87-92. DOI: https://DOI.org/10.1016/j.irbm.2020.05.003

Paulson et al. (2020). Ai based indigenous medicinal plants identification. In 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA) (pp. 57-63). IEEE. DOI: 10.1109/ACCTHPA49271.2020.9213224

Peffers et al. (2020). Design science research process: a model for producing and presenting information systems research. arXiv preprint arXiv:2006.02763. DOI: https://doi.org/10.48550/arXiv.2006.02763

Peng et al. (2019). Correlation congruence for knowledge distillation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 5007-5016).

Peng et al. (2023). Mbfquant: a multiplier-bitwidth-fixed, mixed-precision quantization method for mobile cnn-based applications. IEEE Transactions on Image Processing, 32, 2438-2453. DOI: 0.1109/TIP.2023.3268562

Perez et al. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621. DOI: https://doi.org/10.48550/arXiv.1712.04621

Petch et al. (2022). Opening the black box: the promise and limitations of explainable machine learning in cardiology. Canadian Journal of Cardiology, 38(2), 204-213. DOI:https://doi.org/10.1016/j.cjca.2021.09.004

Petsiuk et al. (2018). Rise: randomized input sampling for explanation of black-box models. arXiv preprint arXiv:1806.07421. DOI: https://doi.org/10.48550/arXiv.1806.07421

Phuong et al. (2019). Towards understanding knowledge distillation. Paper presented at the international conference on machine learning. In Proceedings of the 36th International Conference on Machine Learning (Vol. 97)

Pio et al. (2014). Integrating microrna target predictions for the discovery of gene regulatory networks: a semi-supervised ensemble learning approach. BMC bioinformatics, 15, 1-17. DOI: https://doi.org/10.1186/1471-2105-15-S1-S4

Polson et al. (2020). Deep learning: computational aspects. Wiley Interdisciplinary Reviews: Computational Statistics, 12(5), e1500. DOI: https://doi.org/10.1002/wics.1500.

Preuer et al. (2019). Interpretable deep learning in drug discovery. Explainable AI: interpreting, explaining and visualizing deep learning, 331-345. DOI: https://doi.org/10.1007/978-3-030-28954-6_18

Pudaruth et al. (2021). Medicplants: a mobile application for the recognition of medicinal plants from the republic of mauritius using deep learning in real-time. IAES International Journal of Artificial Intelligence, 10(4), 938. DOI: 10.11591/ijai.v10.i4.pp938-94

Pukhrambam et al. (2022). Advanced medicinal plants classification and bioactivity identification based on dense net architecture. preservation, 13(6).

Rajaraman et al. (2021). Novel loss functions for ensemble-based medical image classification. Plos one, 16(12), e0261307. DOI: https://doi.org/10.1371/journal.pone.0261307

Ramcharan et al. (2017). Deep learning for image-based cassava disease detection. Frontiers in plants science, 8, 1852. DOI: https://doi.org/10.3389/fpls.2017.01852

Read et al. (2011). Classifier chains for multi-label classification. Machine learning, 85, 333-359. DOI: https://doi.org/10.1007/s10994-011-5256-5

Reza. (2023). Deep learning for resource constraint devices.Digital Communicaitons , Prairie View A&M University 8-2023.

Ribeiro et al. (2018). Anchors: high-precision model-agnostic explanations. In Proceedings of the AAAI conference on artificial intelligence (Vol. 32, No. 1). DOI: https://doi.org/10.1609/aaai.v32i1.11491

Ribeiro et al. (2016). " why should i trust you?" explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144). DOI: https://doi.org/10.1145/2939672.293977

Romero et al. (2015). "Fitnets: Hints for thin deep nets." arXiv preprint arXiv:1412.6550 (2014). DOI: https://doi.org/10.48550/arXiv.1412.6550

Romero et al. (2015). Experimental study of event based pid controllers with different sampling strategies. Application to brushless dc motor networked control system. In 2015 XXV international conference on information, communication and automation technologies (ICAT) (pp. 1-6). IEEE. DOI: 10.1109/ICAT.2015.7340515

Roopashree et al. (2021). Deepherb: a vision based system for medicinal plants using xception features. Ieee Access, 9, 135927-135941. DOI: 10.1109/ACCESS.2021.3116207

Roth. (1988). The shapley value: essays in honor of lloyd s. Shapley: cambridge university press. Cambridge University Press

Russakovsky et al. (2015). Imagenet large scale visual recognition challenge. International journal of computer vision, 115, 211-252. DOI: https://doi.org/10.1007/s11263-015-0816-y

Saberian et al. (2019). Multiclass boosting: margins, codewords, losses, and algorithms. Journal of Machine Learning Research, 20(137), 1-68.

Sachar et al. (2022). Deep ensemble learning for automatic medicinal leaf identification. International Journal of Information Technology, 14(6), 3089-3097. DOI: https://doi.org/10.1007/s41870-022-01055-z

Sakib et al. (2019). An overview of convolutional neural network: its architecture and applications. Preprints. DOI: https://doi.org/10.20944/preprints201811.0546.v4

Sandler et al. (2018). Mobilenetv2: inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520). DOI:https://doi.org/10.48550/arXiv.1801.04381

Scetbon et al. (2021). Low-rank sinkhorn factorization. In International Conference on Machine Learning (pp. 9344-9354). PMLR.

Schietgat et al. (2010). Predicting gene function using hierarchical multi-label decision tree ensembles. BMC bioinformatics, 11, 1-14. DOI:https://doi.org/10.1186/1471-2105-11-2

Selvaraju et al. (2017). Grad-cam: visual explanations from deep networks via gradient-based localization. International journal of computer vision, 128, 336-359. DOI: https://doi.org/10.1007/s11263-019-01228-7

Seta et al. (2021). Gullele botanic garden, addis ababa (ethiopia): current status, challenges and opportunities. DOI: 10.24823/Sibbaldia.2021.313

Seta et al. (2022). Botanic garden profile gullele botanic garden, addis ababa (ethiopia): current status, challenges and opportunities. Sibbaldia: the International Journal of Botanic Garden Horticulture, (21), 13-34.

Sharma et al. (2023). Analysis and prediction of covid-19 multivariate data using deep ensemble learning methods. International Journal of Environmental Research and Public Health, 20(11), 5943. DOI: https://doi.org/10.3390/ijerph20115943

Sheng et al. (2024). Cosine similarity knowledge distillation for surface anomaly detection. Scientific Reports, 14(1), 8150. DOI: https://doi.org/10.1038/s41598-024-58409-9

Shi et al. (2011). Multi-label ensemble learning. In Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2011, Athens, Greece, September 5-9, 2011, Proceedings, Part III 22 (pp. 223-239). DOI: https://doi.org/10.1007/978-3-642-23808-6_15

Shrikumar et al. (2017). Learning important features through propagating activation differences. In International conference on machine learning (pp. 3145-3153). PMlR.

Shyaula et al. (2020). Osyris quadripartita salzm. Ex decne. Santalaceae. In ethnobotany of the himalayas (pp. 1-8): springer. DOI: https://doi.org/10.1007/978-3-030-45597-2_170-1

Simonyan et al. (2013). Deep inside convolutional networks: visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034. DOI: https://doi.org/10.48550/arXiv.1312.6034

Simonyan et al. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. DOI: https://doi.org/10.48550/arXiv.1409.1556

Singh et al. (2020). Explainable deep learning models in medical image analysis. Journal of imaging, 6(6), 52. DOI: https://doi.org/10.3390/jimaging6060052

Siu. (2019). Residual networks behave like boosting algorithms. In 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA) (pp. 31-40). IEEE. DOI: 10.1109/DSAA.2019.00017

Smaida et al. (2020). Bagging of convolutional neural networks for diagnostic of eye diseases. Paper presented at the colins. In COLINS (pp. 715-729)

Smilkov et al. (2017). Smoothgrad: removing noise by adding noise. arXiv preprint arXiv:1706.03825. DOI: https://doi.org/10.48550/arXiv.1706.03825

Smith-hall et al. (2012). People, plants and health: a conceptual framework for assessing changes in medicinal plants consumption. Journal of ethnobiology and ethnomedicine, 8, 1-11.DOI: DOI: https://doi.org/10.1186/1746-4269-8-43

Springenberg et al. (2014). Striving for simplicity: the all convolutional net. arXiv preprint arXiv:1412.6806. DOI: https://doi.org/10.48550/arXiv.1412.6806

Srivastav et al. (2024). Deep learning based automatic facial emotion recognition. Intelligent Data Analysis, (Preprint), 1-31. DOI: 10.3233/IDA-237366

Ssenku et al. (2022). Medicinal plants use, conservation, and the associated traditional knowledge in rural communities in eastern uganda. Tropical medicine and health, 50(1), 39. DOI:10.1186/s41182-022-00428-1

Su et al. (2019). Deep-resp-forest: a deep forest model to predict anti-cancer drug response. Methods, 166, 91-102. DOI: https://doi.org/10.1016/j.ymeth.2019.02.009

Sun et al. (2018). Sparse deep stacking network for fault diagnosis of motor. IEEE Transactions on Industrial Informatics, 14(7), 3261-3270. DOI: 10.1109/TII.2018.2819674

Sun et al. (2024). Logit standardization in knowledge distillation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 15731-15740). DOI: https://doi.org/10.48550/arXiv.2403.01427

Sun et al. (2017). Revisiting unreasonable effectiveness of data in deep learning era. In Proceedings of the IEEE international conference on computer vision (pp. 843-852). DOI:https://doi.org/10.48550/arXiv.1707.02968

Sun et al. (2023). An image object detection model based on mixed attention mechanism optimized yolov5. Electronics, 12(7), 1515. DOI: https://doi.org/10.3390/electronics12071515

Sundararajan et al. (2017). Axiomatic attribution for deep networks. Paper presented at the international conference on machine learning. In International conference on machine learning (pp. 3319-3328). PMLR.

Swaminathan et al. (2020). Sparse low rank factorization for deep neural network compression. Neurocomputing, 398, 185-196.DOI: https://doi.org/10.1016/j.neucom.2020.02.035

Szegedy et al. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI conference on artificial intelligence (Vol. 31, No. 1). DOI: https://doi.org/10.1609/aaai.v31i1.11231

Szegedy et al. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

Szegedy et al. (2014). Scalable, high-quality object detection. arXiv preprint arXiv:1412.1441. DOI: https://doi.org/10.48550/arXiv.1412.1441

Szegedy et al. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

Tabernik et al. (2020). Spatially-adaptive filter units for compact and efficient deep neural networks. International Journal of Computer Vision, 128(8), 2049-2067. DOI: https://doi.org/10.1007/s11263-019-01282-1

Tadesse et al. (2009). Ovicidal and larvicidal activity of crude extracts of maesa lanceolata and plectranthus punctatus against haemonchus contortus. Journal of ethnopharmacology, 122(2), 240-244. DOI: https://doi.org/10.1016/j.jep.2009.01.014

Takahashi et al. (2019). Data augmentation using random image cropping and patching for deep cnns. IEEE Transactions on Circuits and Systems for Video Technology, 30(9), 2917-2931. DOI: 10.1109/TCSVT.2019.2935128

Takamoto et al. (2020). An efficient method of training small models for regression problems with knowledge distillation. In 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 67-72). IEEE. DOI: 10.1109/MIPR49039.2020.00021

Talebi et al. (2021). Learning to resize images for computer vision tasks. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 497-506). DOI: https://doi.org/10.48550/arXiv.2103.09950

Tan et al. (2020). Deep learning for plants species classification using leaf vein morphometric. Ieee/acm trans comput biol bioinform, 17(1), 82-90. DOI:10.1109/tcbb.2018.2848653

Tan et al. (2019). Efficientnet: rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.

Tang et al. (2021). Deep stacking network for intrusion detection. Sensors, 22(1), 25. DOI: https://doi.org/10.3390/s22010025

Tang et al. (2020). Understanding and improving knowledge distillation. arXiv preprint arXiv:2002.03532. DOI: https://doi.org/10.48550/arXiv.2002.03532

Tang et al. (2015). A local binary pattern based texture descriptors for classification of tea leaves. Neurocomputing, 168, 1011-1023. DOI: https://doi.org/10.1016/j.neucom.2015.05.024

Tarvainen et al. (2017). Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems, 30.

Tefera et al. (2019). Ethnobotanical study of medicinal plants in the hawassa zuria district, sidama zone, southern ethiopia. Journal of ethnobiology and ethnomedicine, 15, 1-21. DOI: https://doi.org/10.1186/s13002-019-0302-7

Teka et al. (2020). Medicinal plants use practice in four ethnic communities (gurage, mareqo, qebena, and silti), south central ethiopia. Journal of ethnobiology and ethnomedicine, 16, 1-12. DOI: https://doi.org/10.1186/s13002-020-00377-1

Thompson et al. (2020). The computational limits of deep learning. arXiv preprint arXiv:2007.05558, 10.

Thomson et al. (2020). Efficient and compact convolutional neural network architectures for non-temporal real-time fire detection. In 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 136-141). DOI: IEEE. 10.1109/ICMLA51294.2020.00030

Tian et al. (2019). Contrastive representation distillation. arXiv preprint arXiv:1910.10699. DOI: https://doi.org/10.48550/arXiv.1910.10699

Tian et al. (2023). Recent advances in stochastic gradient descent in deep learning. Mathematics, 11(3), 682. DOI: https://doi.org/10.3390/math11030682

Torfi et al. (2020). Natural language processing advancements by deep learning: a survey. arXiv preprint arXiv:2003.01200. DOI :https://doi.org/10.48550/arXiv.2003.01200

Ts et al. (2021). Identification of indian medicinal plants from leaves using transfer learning approach. In 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 980-987). IEEE. DOI: 10.1109/ICOEI51242.2021.9452917

Tsoumakas et al. (2007). Random k-labelsets: an ensemble method for multilabel classification. Paper presented at the european conference on machine learning. In European conference on machine learning (pp. 406-417). Berlin, Heidelberg: Springer Berlin Heidelberg. DOI: https://doi.org/10.1007/978-3-540-74958-5_38

Tu et al. (2019). Understanding generalization in recurrent neural networks. In International Conference on Learning Representations.

Tuasha et al. (2018). Plants used as anticancer agents in the ethiopian traditional medical practices: a systematic review. Evidence-Based Complementary and Alternative Medicine, 2018(1), 6274021. DOI: https://doi.org/10.1155/2018/6274021.

Tung et al. (2019). Similarity-preserving knowledge distillation. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1365-1374).

Uddin et al. (2023). Deep-learning-based classification of bangladeshi medicinal plants using neural ensemble models. Mathematics, 11(16), 3504. DOI: https://doi.org/10.3390/math11163504

Urban et al. (2016). Do deep convolutional nets really need to be deep and convolutional? arXiv preprint arXiv:1603.05691. DOI:https://doi.org/10.48550/arXiv.1603.05691

Urner et al. (2011). Access to unlabeled data can speed up prediction time. In Proceedings of the 28th International Conference on Machine Learning (ICML-11) (pp. 641-648).

Van hieu et al. (2020). Automatic plants image identification of vietnamese species using deep learning models. arXiv preprint arXiv:2005.02832. DOI: https://doi.org/10.48550/arXiv.2005.02832

Van looveren et al. (2021). Interpretable counterfactual explanations guided by prototypes. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 650-665). Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-86520-7_40

Vapnik et al. (2015). Learning using privileged information: similarity control and knowledge transfer. J. Mach. Learn. Res., 16(1), 2023-2049.

Wachter et al. (2017). Counterfactual explanations without opening the black box: automated decisions and the gdpr. Harv. JL & Tech., 31, 841.

Waltner et al. (2019). Hibster: hierarchical boosted deep metric learning for image retrieval. In 2019 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 599-608). IEEE. DOI: 10.1109/WACV.2019.00069

Wang. (2020). Enaet: a self-trained framework for semi-supervised and supervised learning with ensemble transformations. IEEE Transactions on Image Processing, 30, 1639-1647. DOI: 10.1109/TIP.2020.3044220

Wang et al. (2020). Revisiting parameter sharing for automatic neural channel number search. Advances in Neural Information Processing Systems, 33, 5991-6002.

Wang et al. (2019). Private model compression via knowledge distillation. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 01, pp. 1190-1197). DOI: https://doi.org/10.1609/aaai.v33i01.33011190

Wang et al. (2020). Exclusivity-consistency regularized knowledge distillation for face recognition. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK,

August 23–28, 2020, Proceedings, Part XXIV 16 (pp. 325-342). Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-58586-0_20

Wang et al. (2020). Particle swarm optimisation for evolving deep neural networks for image classification by evolving and stacking transferable blocks. In 2020 IEEE Congress on Evolutionary Computation (CEC) (pp. 1-8). IEEE. DOI: 10.1109/CEC48606.2020.9185541.

Wang et al. (2021). Knowledge distillation and student-teacher learning for visual intelligence: a review and new outlooks. IEEE transactions on pattern analysis and machine intelligence, 44(6), 3048-3068. DOI: 10.1109/TPAMI.2021.3055564

Wang et al. (2017). Deep additive least squares support vector machines for classification with model transfer. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 49(7), 1527-1540. DOI: 10.1109/TSMC.2017.2759090

Wang et al. (2018). Dataset distillation. arXiv preprint arXiv:1811.10959. DOI: https://doi.org/10.48550/arXiv.1811.10959

Welchowski et al. (2016). A framework for parameter estimation and model selection in kernel deep stacking networks. Artificial intelligence in medicine, 70, 31-40. DOI: https://doi.org/10.1016/j.artmed.2016.04.002

Who. (2019). Who global report on traditional and complementary medicine 2019: world health organization.

Woldeamanuel et al. (2022). Ethnobotanical study of endemic and non-endemic medicinal plants used by indigenous people in environs of gullele botanical garden addis ababa, central ethiopia: a major focus on asteraceae family. Frontiers in Pharmacology, 13, 1020097. DOI: https://doi.org/10.3389/fphar.2022.1020097

Woldegerima et al. (2017). Ecosystem services assessment of the urban forests of addis ababa, ethiopia. Urban Ecosystems, 20, 683-699. DOI: https://doi.org/10.1007/s11252-016-0624-3

Wu et al. (2022). Effect of transfer learning on the performance of vggnet-16 and resnet-50 for the classification of organic and residual waste. Frontiers in Environmental Science, 10, 1043843. DOI: https://doi.org/10.3389/fenvs.2022.1043843

Xia et al. (2021). Multi-label classification with weighted classifier selection and stacked ensemble. Information Sciences, 557, 421-442. DOI: https://doi.org/10.1016/j.ins.2020.06.017

Xiang et al. (2019). Fruit image classification based on mobilenetv2 with transfer learning technique. In Proceedings of the 3rd international conference on computer science and application engineering (pp. 1-7). DOI: https://doi.org/10.1145/3331453.3361658

Xiao et al. (2010). Hog-based approach for leaf classification. Paper presented at the international conference on intelligent computing. DOI: https://doi.org/10.1007/978-3-642-14932-0_19

Xie et al. (2020). Self-training with noisy student improves imagenet classification. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10687-10698). DOI: https://doi.org/10.48550/arXiv.1911.04252

Xu et al. (2023). Teacher-student collaborative knowledge distillation for image classification. Applied Intelligence, 53(2), 1997-2009. DOI:https://doi.org/10.1007/s10489-022-03486-4

Xu et al. (2020). Knowledge distillation meets self-supervision. In European conference on computer vision (pp. 588-604). Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-58545-7_34

Xue et al. (2013). Restructuring of deep neural network acoustic models with singular value decomposition. In Interspeech (pp. 2365-2369). DOI: 10.21437/Interspeech.2013-552

Xue et al. (2021). Multimodal knowledge expansion. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1295-1304). DOI: https://doi.org/10.48550/arXiv.2106.10598

Yang et al. (2023). A survey on ensemble learning under the era of deep learning. Artificial Intelligence Review, DOI: https://doi.org/10.1007/s10462-022-10283-5.

Yang et al. (2020). Model compression with two-stage multi-teacher knowledge distillation for web question answering system. In Proceedings of the 13th International Conference on Web Search and Data Mining (pp. 690-698). DOI: https://doi.org/10.1145/3336191.3371792

Yang et al. (2015). Convolutional channel features. In Proceedings of the IEEE international conference on computer vision (pp. 3074-3082).

Yang et al. (2023). From knowledge distillation to self-knowledge distillation: a unified approach with normalized loss and customized soft labels. DOI: https://doi.org/10.48550/arXiv.2303.13005

Yaniv. (2014). Introduction: medicinal plants in ancient traditions. Medicinal and Aromatic Plants of the Middle-East, 1-7. DOI: https://doi.org/10.1007/978-94-017-9276-9_1.

Yeom et al. (2021). Pruning by explaining: a novel criterion for deep neural network pruning. Pattern Recognition, 115, 107899. DOI: https://doi.org/10.1016/j.patcog.2021.107899

Yim et al. (2017). A gift from knowledge distillation: fast optimization, network minimization and transfer learning. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4133-4141).

Yirgu et al. (2019). Useful medicinal tree species of ethiopia: comprehensive review. South African Journal of Botany, 122, 291-300. DOI: https://doi.org/10.1016/j.sajb.2019.03.026

Yosinski et al. (2015). Understanding neural networks through deep visualization. arXiv preprint arXiv:1506.06579. DOI: https://doi.org/10.48550/arXiv.1506.06579

You et al. (2017). Learning from multiple teacher networks. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1285-1294).

Yu et al. (2012). Transductive multi-label ensemble classification for protein function prediction. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1077-1085). DOI: https://doi.org/10.1145/2339530.2339700

Yu et al. (2019). Learning metrics from teachers: compact networks for image embedding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2907-2916).

Yu et al. (2015). Learning deep representations via extreme learning machines. Neurocomputing, 149, 308-315. DOI: https://doi.org/10.1016/j.neucom.2014.03.077

Yuan et al. (2019). Revisit knowledge distillation: a teacher-free framework.

Zafar et al. (2019). Dlime: a deterministic local interpretable model-agnostic explanations approach for computer-aided diagnosis systems. arXiv preprint arXiv:1906.10263. DOI: https://doi.org/10.48550/arXiv.1906.10263.

Zagoruyko et al. (2016). Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer. arXiv preprint arXiv:1612.03928. DOI: https://doi.org/10.48550/arXiv.1612.03928

Zeiler et al. (2014). Visualizing and understanding convolutional networks. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13 (pp. 818-833). Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-10590-1_53

Zeiler et al. (2011). Adaptive deconvolutional networks for mid and high level feature learning. In 2011 international conference on computer vision (pp. 2018-2025). IEEE. **DOI:** 10.1109/ICCV.2011.6126474

Zeng et al. (2023). Abs-cam: a gradient optimization interpretable approach for explanation of convolutional neural networks. Signal, Image and Video Processing, 17(4), 1069-1076. DOI: https://doi.org/10.1007/s11760-022-02313-0

Zhai et al. (2019). Lifelong gan: continual learning for conditional image generation. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 2759-2768).

Zhang et al. (2022). Stack operation of tensor networks. Frontiers in Physics, 10, 906399. DOI: https://doi.org/10.3389/fphy.2022.906399

Zhang et al. (2021). A bagging dynamic deep learning network for diagnosing covid-19. Scientific Reports, 11(1), 16280. DOI: https://doi.org/10.1038/s41598-021-95537-y

Zhang et al. (2019). Deep stacked hierarchical multi-patch network for image deblurring. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5978-5986).

Zhang et al. (2020). Snapshot boosting: a fast ensemble framework for deep neural networks. Science China Information Sciences, 63(1), 112102. DOI: https://doi.org/10.1007/s11432-018-9944-x

Zhang et al. (2018). An information-theoretic view for deep learning. arXiv preprint arXiv:1804.09060. DOI: https://doi.org/10.48550/arXiv.1804.09060

Zhang et al. (2020). Improve object detection with feature-based knowledge distillation: towards accurate and efficient detectors. In International Conference on Learning Representations.

Zhang et al. (2018). Better and faster: knowledge transfer from multiple self-supervised learning tasks via graph distillation for video classification. arXiv preprint arXiv:1804.10069. DOI: https://doi.org/10.48550/arXiv.1804.10069.

Zhang et al. (2020). Interpretable deep learning under fire. In 29th security symposium.

Zhang et al. (2018). Deep mutual learning. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4320-4328). DOI: 10.48550/arXiv.1706.00384.

Zhang et al. (2020). Grasp for stacking via deep reinforcement learning. In 2020 IEEE International Conference on Robotics and Automation (ICRA) (pp. 2543-2549). IEEE. DOI: 10.1109/ICRA40945.2020.9197508

Zhao et al. (2015). Saliency detection by multi-context deep learning. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1265-1274).

Zhou. (2012). Ensemble methods: foundations and algorithms: crc press.

Zhou et al. (2019). Deep forest. National science review, 6(1), 74-86. DOI: 10.1093/nsr/nwy108

Zhou et al. (2014). Stacked extreme learning machines. IEEE transactions on cybernetics, 45(9), 2013-2025. DOI: 10.1109/TCYB.2014.2363492

Zhou et al. (2016). Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921-2929).

Zhou et al. (2019). Understanding knowledge distillation in non-autoregressive machine translation. arXiv preprint arXiv:1911.02727. DOI: 10.48550/arXiv.1911.02727

Zhou et al. (2021). Rethinking soft labels for knowledge distillation: a bias-variance tradeoff perspective. arXiv preprint arXiv:2102.00650. DOI: 10.48550/arXiv.1911.02727

Zhou et al. (2021). Bert learns to teach: knowledge distillation with meta learning. arXiv preprint arXiv:2106.04570. DOI: https://doi.org/10.48550/arXiv.2106.04570

Zhu et al. (2017). To prune, or not to prune: exploring the efficacy of pruning for model compression. arXiv preprint arXiv:1710.01878. DOI: 10.48550/arXiv.1710.01878.

Zong et al. (2022). Better teacher better student: dynamic prior knowledge for knowledge distillation. arXiv preprint arXiv: 2206.06067. DOI: 10.48550/arXiv.2206.06067.

Zuchniak. (2023). Multi-teacher knowledge distillation as an effective method for compressing ensembles of neural networks. arXiv preprint arXiv:2302.07215. DOI: 10.48550/arXiv.2302.07215

Appendix-A: Sample List of reviewed Ethiopian medicinal plants used for various traditional disease treatments with their parts.

No	Sample Image	Scientific Name	Local Name	Parts used	Description(Traditional Usage)
1		*Aloe monticola* Reynolds	Eret	Leaves and Roots	For also curing anthrax by pounding the root and mixing it with cold water and local alcohol(Awulachew &Plants, 2021)
2		*Achyranthes aspera L.*	Telenj	leaves	boils, asthma, in facilitating delivery, bleeding, bronchitis, debility, dropsy, cold, colic, cough, dog bite, snake bite, scorpion bite, dysentery, earache, headache, leukoderma, renal complications, pneumonia, and skin diseases
4		*Buddleja polystachya* Fresen.	Anfar	leaves	For treating the cattle eye diseases by chewing and Spitting on the affected area(Awulachew &Plants, 2021).
5		*Bersama abyssinica* Fresen.	Azamr	Leaves&Stem	For treating wound by squeezing the leaves and creaming on the wound(Awulachew &Plants, 2021)
6		*Carissa spinarum L*	Agam	Roots	Used for preventing evil eye by inhaling the smoke of pounded roots. It is also used for treating wounds via applying the powder of the roots(Awulachew &Plants, 2021)
7		*Clematis simensis Fresen.*	Hareg	leaves	Used for curing wound and stomachache(Awulachew &Plants, 2021)

8		*Clutia abyssinica* Jaub. & Spach.	Fiyele Fejj	whole parts	Roots used for liver problems, colds, fever, headache, malaria, stomachache, flu, and indigestion. The leaves are applied to treat tonsillitis. The woods are used to treat menstrul pains (Bussmann *et al.*, 2021).
9		*Croton macrostachyus* Del.	Bissana	whole parts and its seed	Whole plants against tapeworm, treating aphasia, eye disease, ringworm, constipation, for induction of abortion and as a purgative, allergy and wounds(Getnet *et al.*, 2016).
10		*Dovyalis abyssinica* (A. Rich.) Warb.	Koshim	whole parts	Used for treating malaria, asthma, tapeworm, cough, and stomach ulcer(Alemneh, 2021).
11		*Ekebergia capensis* Sparrm.	Meliaceae	Bark	Used for Fever and malaria Gastrointestinal problems (diarrhea, dysentery, gastritis, and stomach ache) (Maroyi &Research, 2018)
12		*Echinops kebericho* Mesfin .	Kebericho	Root	For treating toothache, vomiting, and headache (Awulachew &Plants, 2021).
13		*Ficus sur Forssk.*	Shola	fruit,Bark	Dysentery(Amsalu *et al.*, 2018)
14		*Hagenia abyssinica* (Bruce) J.F.Gmelin	Koso	Seed	Used for treating tapeworm(Amsalu *et al.*, 2018).

15		*Jasminum abyssinicum* Hochst. Ex Dc.	Tembelel	leaves	Used for treating tapeworm(Amsalu *et al.*, 2018)
16		*Laggera tomentosa* (Sch. Bip. ex A. Rich.) Oliv. & Hiern	Koskoso	leaves	It is used as a fumigant. it is also used as the treatments of tonsil (Daba &Asfaw, 2020)
		Aloe Ankoberensis Gilbert & Sebsebe	Merrarie	Leaves	treatment of wounds and skin complaints, malaria, microbial infections, and complaints of the digestive system(Woldeamanuel *et al.*, 2022)
17		*Leonotis ocymifolia* (Burml. f.) Iwarsson var. raineriana	Ras Kemer	Leaves and Roots	Used for treating Diarrhea (Amsalu *et al.*, 2018)
18		*Leucas abyssinica* (Benth.) Briq.	Chimida	Leaves	Stomach ache, Amoebiasis, Stomach bloating, Head ache, Food poisoning, Vomiting(Kidane *et al.*, 2014)
19		*Lippia adoensis* Hochst. ex Walp.	Kosseret	Leaves	Gastritis (Amsalu *et al.*, 2018)

20		*Lobelia rhynchopetalum* Hemsl.	Jebera	Leaves and roots	Evil eye(Amsalu *et al.*, 2018)
21		*Millettia ferruginea* (Hochst.) Bak.	Birbira	Roots and leaves	Skin infection , Goiter (Amsalu *et al.*, 2018)
22		*Maesa lanceolata* Forssk.	Qelewa	Leaves, fruits	Used as a vermifuge, against tapeworm, against ascaris. treat sore throat, tapeworms, hepatitis and cholera(Tadesse *et al.*, 2009).
23		*Osyris quadripartita* Decn.	Qeret	Leaves	Used for treating malaria, quaqucha (Shyaula *et al.*, 2020)
24		*Phytolacca dodecandra* L 'Hérit.	Indod	Roots, leaves	Gonorrhea, Stomach bloating(Kidane *et al.*, 2014)
25		Plantsago lanceolata L.	Gorteb	Leaves	Wound (Bussmann *et al.*, 2021).
26		Rumex abyssinicus Jacq.	Mekmeko	Roots	Malaria, Hypertension (Awulachew &Plants, 2021)

27		*Rumex nervosus* Vahl (Polygonaceae)	Embwacho	Root,leaves	Burn(Amsalu *et al.*, 2018)
28		*Solanecio gigas* (Vatke) C. Jeffrey	Shekoko Gomen	Root and Leaves	The treatment of colic, diarrhea, gout, otitis media, typhoid fever(Asres *et al.*, 2007).
29		*Stephania abyssinica* (Dillon & A. Rich.) Walp	Walp.(Etse Eyesus, Nech-Hareg)	Leaves	For treating external tumor/ cancer and Stomachache (Awulachew &Plants, 2021)
30		*Thymus schimperi* Ron.	Tosgn	Leaves	Eveil Eye, Asthma (Bussmann *et al.*, 2021)
31		Urtica simensis Stedel	Sama	Leaves	Gastritis(Amsalu *et al.*, 2018)
32		Verbascum sinaiticum Benth	Yeahiya Joro	Roots, leaves	For treating heart disease, cancer, trypanosomiasis(Awulachew &Plants, 2021)

33		Vernonia amygdalina Del.	Grawa	Leaves	for treating ascariasis (Awulachew &Plants, 2021)
34		*Vernonia leopoldi* (Sch. Bip. ex Walp.) Vatke	Merara kitel	leaves	Used for treating Tumor(Tuasha *et al.*, 2018)
35		*Inula confertiflora* A. Rich	Weynagft	Leaves	Infected eye(Amsalu *et al.*, 2018)

Appendix-B: List of Publications

Mulugeta Adibaru Kiflie., DP Sharma & Mesfin Abebe Haile (2024). **Deep learning for medicinal plants species classification and recognition: a systematic review.** Frontiers in Plants Science, 14, 1286088.

Mulugeta Adibaru Kiflie., DP Sharma, Mesfin Abebe Haile & Ramasamy Srinivasagan (2024). **EfficientNet Ensemble Learning: Identifying Ethiopian Medicinal Plants Species and Traditional Uses by Integrating Modern Technology with Ethnobotanical Wisdom.** Computers, 13(2), 38.

Mulugeta Adibaru Kiflie., DP Sharma & Mesfin Abebe Haile (2024). **Deep Learning for Ethiopian Indigenous Medicinal Plants Species Identification and Classification**. JAIM (15), Article accepted for publication on 17 May 2024. Track Your Article (elsevier.com), (JAIM_100987, Kiflie)

Adibaru, M. (2023). **Ethiopian Indigenous Medicinal Plants Dataset**. DOI: 10.6084/m9.figshare.24137802.v1.

Appendix-C: Sample Snapped Code (Interpretable Deep Learning)

```
# Define Distiller class

class Distiller(keras.Model):
    def __init__(self, student, teachers):
        super(Distiller, self).__init__()
        self.student = student
        self.teachers = teachers
    # Define compile method
    def compile(
        self,
        optimizer,
        metrics,
        student_loss_fn,
        distillation_loss_fn,
        alpha=0.1,
        temperature=3,
    ):
        super(Distiller, self).compile(optimizer=optimizer, metrics=metrics)
        self.student_loss_fn = student_loss_fn
        self.distillation_loss_fn = distillation_loss_fn
        self.alpha = alpha
        self.temperature = temperature
    # Define call method
    def call(self, inputs, training=False):
        return self.student(inputs, training=training)
    # Define train_step method
    def train_step(self, data):
        x, y = data
        # Forward pass of teacher models
        teacher_predictions = [teacher(x, training=False) for teacher in self.teachers]
        with tf.GradientTape() as tape:
            student_predictions = self.student(x, training=True)
```

```
            student_loss = self.student_loss_fn(y, student_predictions)
            distillation_loss = 0
            num_teacher_predictions = len(teacher_predictions)
            for teacher_prediction in teacher_predictions:
                distillation_loss += self.distillation_loss_fn(
                    tf.nn.softmax(teacher_prediction / self.temperature, axis=1),
                    tf.nn.softmax(student_predictions / self.temperature, axis=1),
                )
            distillation_loss /= num_teacher_predictions
            total_loss = self.alpha * student_loss + (1 - self.alpha) * distillation_loss
        trainable_vars = self.student.trainable_variables
        gradients = tape.gradient(total_loss, trainable_vars)
        self.optimizer.apply_gradients(zip(gradients, trainable_vars))
        self.compiled_metrics.update_state(y, student_predictions)
        results = {m.name: m.result() for m in self.metrics
        results.update(
            {"student_loss": student_loss, "distillation_loss": distillation_loss}
        )
        return results
# Train teacher models with fine-tuning and learning rate scheduling
for teacher in [teacher1, teacher2, teacher3]:
    teacher.compile(
        optimizer=keras.optimizers.Adam(learning_rate=1e-5),
        loss='categorical_crossentropy',
        metrics=['accuracy']
    )
    for layer in teacher.layers[-20:]:
        layer.trainable = True

    history = teacher.fit(train_generator,
                          steps_per_epoch=train_generator.n // batch_size,
                          validation_data=val_generator,
                          validation_steps=val_generator.n // batch_size,
                          epochs=ep,
                          callbacks=callbacks)
# Initialize and compile the distiller
distiller = Distiller(student=student, teachers=[teacher1, teacher2, teacher3])
# Define initial distiller weights
initial_distiller_weights = distiller.get_weights()
```

```
# Define function to evaluate cosine similarity
def evaluate_cosine_similarity(model, data_generator):
    # Explicitly convert the model prediction to a numpy array
    predictions = np.array(model.predict(data_generator))
    student_predictions = np.mean(predictions, axis=0)
    # Calculate teacher predictions separately
    teacher_predictions = []
    for teacher in [teacher1, teacher2, teacher3]:
        # Explicitly convert the teacher prediction to a numpy array
        teacher_pred = np.array(teacher.predict(data_generator))
        teacher_predictions.append(np.mean(teacher_pred, axis=0))
    # Ensure both arrays have the same number of features
    num_features = min(student_predictions.shape[0], *map(lambda x: x.shape[0], teacher_predictions))
    student_predictions = student_predictions[:num_features]
    teacher_predictions = [pred[:num_features] for pred in teacher_predictions]
    cosine_similarity_scores = [cosine_similarity(teacher_pred.reshape(1, -1), student_predictions.reshape(1, -
1)) for teacher_pred in teacher_predictions]
    return np.mean(cosine_similarity_scores)
# Define the hyperparameter grid
param_grid = {
    'temperature': [4, 5],  # Distillation temperature
    'alpha': [0.3, 0.5],  # Weight for distillation loss
    'learning_rate': [1e-4, 5e-5],  # Learning rate for student model
}
# Generate all possible combinations of hyperparameters
param_combinations = list(ParameterGrid(param_grid))
best_cosine_similarity = -1  # Initialize best cosine similarity score
best_hyperparameters = None  # Initialize best hyperparameters
# Iterate over all hyperparameter combinations
for params in param_combinations:
    # Reset distiller weights
    distiller.set_weights(initial_distiller_weights)
    # Initialize and compile the distiller with current hyperparameters
    distiller.compile(
        optimizer=keras.optimizers.Adam(learning_rate=params['learning_rate']),
        metrics=['accuracy'],
        student_loss_fn=keras.losses.CategoricalCrossentropy(),
        distillation_loss_fn=keras.losses.KLDivergence(),
        alpha=params['alpha'],
        temperature=params['temperature'],
    )
    # Train the distiller model
    distiller.fit(train_generator, epochs=5)
    # Evaluate cosine similarity on validation data
    cosine_similarity_score = evaluate_cosine_similarity(distiller, val_generator)
```

```
    # Update best cosine similarity and hyperparameters if current score is better
    if cosine_similarity_score > best_cosine_similarity:
        best_cosine_similarity = cosine_similarity_score
        best_hyperparameters = params
# Compile and train the student model
callbacks_student = [
    EarlyStopping(monitor='val_loss', patience=5, restore_best_weights=True),
    ReduceLROnPlateau(monitor='val_loss', factor=0.1, patience=3, verbose=1, min_lr=1e-6)]
student.compile(optimizer=keras.optimizers.Adam(learning_rate=1e-4), loss='categorical_crossentropy',
metrics=['accuracy'])
history = student.fit(train_generator, epochs=20, validation_data=val_generator, callbacks=callbacks_student)
# Fine-tune the base layers by unfreezing some of them
for layer in student.layers[-20:]:
    layer.trainable = True
# Compile the model again for fine-tuning
student.compile(
    optimizer=tf.keras.optimizers.Adam(learning_rate=1e-5),
    loss='categorical_crossentropy',
    metrics=['accuracy']
)
# Train the entire model with fine-tuning
history_fine_tune = student.fit(
    train_generator,
    epochs=20,
    validation_data=val_generator,
    callbacks=callbacks_student
# Define predict_image function
def predict_image(image_path, model):
    img = load_img(image_path, target_size=(224, 224))
    img_array = img_to_array(img) / 255.0
    img_array = np.expand_dims(img_array, axis=0)
    pred_probabilities = model.predict(img_array)
    predicted_class_index = np.argmax(pred_probabilities)
    class_info = class_names[predicted_class_index]
    pred_prob = pred_probabilities[0][predicted_class_index]
    return class_info, pred_prob, img_array
# Define generate_semantic_lime_explanations function
def generate_semantic_lime_explanations(image_array, image_path,
teacher_models, student_model, cosine_similarity_score):
    explainer = LimeImageExplainer()
    teacher_pred_probs = [teacher_model.predict(image_array) for
teacher_model in teacher_models]
    explanation = explainer.explain_instance(
        image_array[0].astype(float),
```

```
        student_model.predict,
        top_labels=3,
        num_samples=1000,
        random_seed=42,
    )
    predicted_class_index = np.argmax(student_model.predict(image_array))
    class_info = class_names[predicted_class_index]
    pred_prob =
student_model.predict(image_array)[0][predicted_class_index]

    cosine_similarities = []
    mse_scores = []
    for i, teacher_pred_prob in enumerate(teacher_pred_probs):
        teacher_pred_prob = teacher_pred_prob[0]
        teacher_pred_prob = np.asarray(teacher_pred_prob).reshape(1, -1)
        pred_prob = np.asarray(pred_prob).reshape(1, -1)

        if teacher_pred_prob.shape[1] != pred_prob.shape[1]:
            pred_prob = np.pad(pred_prob, ((0, 0), (0,
teacher_pred_prob.shape[1] - pred_prob.shape[1])))
        mse_score = mean_squared_error(teacher_pred_prob.flatten(),
pred_prob.flatten())
        mse_scores.append(mse_score)
    temp, mask = explanation.get_image_and_mask(
        explanation.top_labels[0],
        positive_only=True,
        num_features=15,
        hide_rest=False,
    )
    semantic_img = mark_boundaries(temp, mask)
    additional_info_text = (
        f'Scientific Name: {class_info.get("scientific")n'
        f'Local Name: {class_info.get("local")}\n'
        f'Parts Used: {class_info.get("parts")}\n'
        f'Uses/Treatment: {class_info.get("uses")}\n'
        f'Confidence: {float(pred_prob[0][0]):.2f}'
    )
    fig, axs = plt.subplots(1, 2, figsize=(16, 6))
    img_original = Image.open(image_path)
    axs[0].imshow(img_original)
    axs[0].set_title("Predicted Class")
    axs[0].axis('off')
    axs[0].text(0.5, -0.1, additional_info_text,
transform=axs[0].transAxes, fontsize=12, ha='center', va='center',
```

```
bbox=dict(facecolor='white', alpha=0.7, edgecolor='white',
boxstyle='round,pad=0.5'))
    axs[1].imshow(semantic_img)
    axs[1].set_title(f'Interpretation: Why the species was predicted as
{class_info["scientific"]}')
    axs[1].axis('off')
    axs[1].text(0.5, -0.1, f'Feature Similarity:
{cosine_similarity_score:.2f}\nMSE: {np.mean(mse_scores) * 100:.4f}',
transform=axs[1].transAxes, fontsize=12, ha='center', va='center',
bbox=dict(facecolor='white', alpha=0.7, edgecolor='white',
boxstyle='round,pad=0.5'))
    plt.show()
for image_path in image_paths:
    img = load_img(image_path, target_size=(224, 224))
    img_array = img_to_array(img) / 255.0
    img_array = np.expand_dims(img_array, axis=0)
    class_info, pred_prob, _ = predict_image(image_path, student)  # Pass
the student model as the second argument
    # Calculate cosine similarity score for this image
    cosine_similarity_score = evaluate_cosine_similarity(distiller,
val_generator)
    import random
    # Calculate cosine similarity score for this image
    cosine_similarity_score = evaluate_cosine_similarity(distiller,
val_generator)
    generate_semantic_lime_explanations(img_array, image_path, [teacher1,
teacher2, teacher3], student, cosine_similarity_score)
def evaluate_student_model(student, test_generator):
    # Evaluate the student model on the test data
    test_accuracy = student.evaluate(test_generator, verbose=0)
    print(f"Student Model Test Accuracy: {test_accuracy[1] * 100:.2f}%")
```