TONGUE DISEASE RECOGNITION USING DEEP LEARNING

BY

URMI GHOSH
ID: 192-15-13225

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

MD. SADEKUR RAHMAN
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

DR. SHEAK RASHED HAIDER NOORI
Professor & Head
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "Tongue Disease Recognition Using Deep Learning", submitted by **[Urmi Ghosh]** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 01-07-2024.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque (SMAH)
Professor & Associate Head **Board Chairman**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Sharmin Akter (SNA) **Internal Examiner 1**

Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Tapasy Rabeya (TRA)
Sr. Lecturer **Internal Examiner 2**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zulfiker Mahmud (ZM) **External Examiner**
Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Md. Sadekur Rahman, Assistant Professor, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Md. Sadekur Rahman
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Dr. Sheak Rashed Haider Noori
Professor & Head
Department of CSE
Daffodil International University

Submitted by:

Urmi Ghosh

Urmi Ghosh
ID: -192-15-13225
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project/internship successfully.

We are grateful and wish our profound indebtedness to **Md. Sadekur Rahman**, **Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Field name*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Dr. Sheak Rashed Haider Noori**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

The tongue is extensively involved in several vital functions, such as taste, speech, eating, swallowing, and oral hygiene. the importance of maintaining the health of the tongue, this "Tongue Disease Recognition Using Deep Learning" research project. We investigated the use of sophisticated convolutional neural networks (CNNs) in the identification and categorization of different tongue disorders using medical photographs. I assessed how well a number of deep learning models performed—InceptionV3, MobileNet, ResNet50, and VGG16—in diagnosing ailments including Median Rhomboid, Black Hairy Tongue, Tongue Ulcer, Ankyloglossia, and Geographic Tongue.These models were trained, verified, and tested with 10,598 high-resolution tongue photos from a comprehensive dataset to guarantee reliable performance. The result of the study provided that the MobileNet model had the best accuracy, 93.83%, proving that it is better able to identify illnesses of the tongue. This study demonstrates how deep learning technology can revolutionize medical diagnostics by offering a non-invasive, effective, and precise way to identify diseases early on.In order to maintain regulatory compliance, ethical issues such as patient privacy and data security were carefully taken into account. This study emphasizes how important it is to include deep learning models into clinical practice since they have the potential to improve healthcare accessible, especially in rural and underdeveloped areas.In order to further increase the system's usability in clinical situations, future work will concentrate on growing the dataset, investigating more complex designs, and improving the system's real-time diagnostic capabilities. This study shows that applying AI to medical diagnostics is feasible and advantageous, with the potential to enhance healthcare outcomes and boost accessibility.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Deep Learning is become quite simple to comprehend. The architecture and operations of the human brain serve as the model for artificial neural networks, a relatively emerging field of artificial intelligence. The enormous volumes of data that are produced every day can be handled by it. Deep learning methods enable computers to recognize photos on a daily basis. Many deep learning algorithms are applied in various domains to solve real-world issues with remarkable accuracy. Consciously training and testing deep learning algorithms using data is necessary to determine which solution is optimal for a given situation.

The tongue is a prodigious muscular organ that executes multiple yet necessary purposes such as taste, phonation, mastication, deglutition, suckling, maintenance of oral hygiene, protection of deeper structures, and facilitation of orofacial growth. tongue health is essential in the diagnostic landscape. Diseases affecting the tongue can be early indicators of systemic health issues such as infections, autoimmune disorders, and even malignancies. Traditionally, the diagnosis of tongue diseases has depended on clinical examinations and histopathological studies, which can be both invasive and subjective. The emergence of deep learning technologies offers a significant advancement in medical imaging. These technologies provide powerful tools capable of analyzing complex visual data with remarkable accuracy and efficiency. In this thesis, I explore applying advanced deep learning models—specifically Vgg16, ResNet50, InceptionV3, and MobileNet—to recognize and classify tongue diseases. These models have been highly successful in various image recognition tasks and hold the potential to revolutionize the diagnostic process by enhancing precision and reducing [1]. The motivation for this research stems from the current challenges in diagnosing tongue diseases, such as the variability in clinical expertise and the invasiveness of biopsies. By integrating deep learning models, this study aims to develop a diagnostic tool that not only improves diagnostic accuracy but also enhances patient comfort and speeds up the diagnostic process.

1.2 Motivation

In approach, the diagnosis of tongue diseases represents not only a significant scientific challenge but also a profound humanitarian commitment. Tongue diseases, ranging from benign lesions to malignant cancers, pose severe risks to human health, affecting vital functions such as speaking, eating, and swallowing. The difficulty in accurately classifying these disorders and identifying them early contributes to their severity and frequently results in incorrect or delayed therapy. It is critical to identify tongue disorders early on, especially malignancy. A diagnosis made early on significantly increases survival rates and treatment outcomes. Unfortunately, a lot of tongue disorders, including malignant ones, are not identified until they are quite advanced. This problem is made worse by the subtlety of the symptoms and the fact that oral health is frequently disregarded, underscoring the urgent need for more accurate and targeted diagnostic instruments. Furthermore This research attempts to close the gap between clinical needs and technical capabilities by utilizing the power of deep learning. Medical imaging and diagnostics could undergo a revolution thanks to models like Vgg16, ResNet50, InceptionV3, and MobileNet. These models excel in deciphering complicated visual data, offering a degree of accuracy that exceeds that of conventional diagnostic techniques in addition to Making the distinction between benign and malignant illnesses is one of the most difficult problems in tongue disease diagnosis. Different types of lesions can have visual similarities that result in incorrect diagnosis and treatment. My goal is to create a diagnostic tool that can accurately classify the nature of an illness and determine whether it exists, allowing for the appropriate course of medical intervention. To do this, I plan to incorporate deep learning algorithms into the diagnostic process.The goal of this project is to employ deep learning technology in novel ways to address these urgent problems. Through improving tongue disease detection speed, accuracy, and efficiency, our work aims to enhance patient outcomes and make a substantial contribution to the larger area of medical diagnostics. By this project, I intend to significantly alter the way tongue diseases are identified and managed, giving those who suffer from these dangerous ailments hope and better future health.

1.3 Rationale of the Study

The limitations of present diagnostic modalities and the encouraging outcomes of new research using deep learning technology emphasize the necessity of improving tongue disease diagnosis through advanced computational methodologies. This study is greatly influenced by seminal publications in the field, especially those showing how convolutional neural networks (CNNs) may greatly enhance diagnostic results, such as those by Sung-Jae Lee et al. and Xu Wang et al. CNNs, namely EfficientNet, were effectively used by Sung-Jae Lee et al.[1] to observe and categorize tongue cancer. They worked with a large dataset that included 1810 photos that were divided into four categories: normal tongues, benign or inflammatory lesions, precancerous lesions, and malignant tumors. 1141 photos were used for training, 180 for testing, and 489 for validation as part of an organized training regimen. EfficientNet's efficiency is demonstrated by the remarkable results of their final model, which had an F2 score of 91.76%, accuracy of 91.67%, precision of 92.12%, and recall of 91.67%. Its greatest performance was 92.17%. This work offers a strong basis and validates CNNs for automated tongue disease detection and categorization. Based on these discoveries, Xu Wang et al.[5] investigated the use of deep learning to detect tongue abnormalities such dental marks that indicate illness. A sizable dataset of 1548 tongue photos, including 672 with dental marks, was used in their investigation. With the use of ResNet34 and other CNN architectures, this study successfully extracted features and carried out classifications, yielding a model accuracy of more than 90% overall. The effectiveness of their model shows that deep learning can identify unique pathological characteristics on the tongue, expanding the use of CNNs in the diagnosis of oral health issues. This study's justification stems from the established efficacy of CNNs in medical imaging as well as the demand for quicker, more precise, non-invasive diagnostic instruments in the field of oral medicine. This study aims to make a significant contribution to the field of medical diagnostics by utilizing the power of advanced CNN models, specifically ResNet50, InceptionV3, and MobileNet. The ultimate goal is to improve clinical outcomes and patient care by offering more dependable tools for tongue disease detection and classification.

1.4 Research Questions

This paper addresses some pivotal questions on the topic of " Tongue Disease Recognition Using Deep Learning." These questions serve as guidelines, directing the study toward, aiming the research at a thorough comprehension of the subject. The research questions include: How effective are deep learning models, such as Vgg16, ResNet50, InceptionV3, and MobileNet, in accurately detecting and classifying different types of tongue diseases?

1. Can the implementation of these deep learning models in tongue disease diagnosis reduce the necessity for invasive diagnostic procedures?

2. How do the performance of deep learning models vary with changes in the quality and type of input images (lighting, resolution, angle)?

1.5 Expected Output

This research efforts to transform medical diagnostics by utilizing deep learning algorithms to identify tongue illnesses. a highly comprehensive diagnosis model that can identify and classify a range of tongue disorders. Modern deep learning architectures like Vgg16, ResNet50, InceptionV3, and MobileNet will be used by this tool, and it is expected that these models will perform exceptionally well and reliably across a range of conditions and disease stages. a comprehensive assessment of the effectiveness, sensitivity, specificity, and diagnostic precision of every deep learning model used in this study. This evaluation will provide informative details regarding the The application of these models is expected to streamline the diagnostic process and reduce the amount of time that passes between the initial consultation and diagnosis. This efficiency not only facilitates better patient flow but also boosts the capacity of healthcare facilities to handle more patients successfully overall. The benefits and drawbacks of every model with regard to tongue disease diagnosis. To make sure the developed models are resilient in a variety of imaging scenarios—such as changes in lighting, resolution, and angle—they will be put through testing. Ensuring the models function well and practically in real-world clinical settings, where these kinds of deviations occur, is the aim. The research will offer guidelines for incorporating these deep learning models into the current healthcare systems, addressing various logistical and technical issues. This includes standard operating procedures for people training, data administration, and system compatibility.

1.6 Project Management and Finance

The research project titled "Tongue Disease Recognition Using Deep Learning" is structured through a detailed project management and finance plan to ensure successful execution.

The project is divided into seven major sections that take a year to complete: planning, gathering data, developing a model, evaluating it, putting it into practice, documenting it, and submitting it. Every stage includes specific tasks and deadlines, starting in the planning phase with goal definition and literature research and ending with the submission of the final report and findings presentation.

Each member of the team has a clear role and responsibility. For example, there is a Project Manager who oversees work and makes sure deadlines are met, a Data Scientist who collects and preprocesses data, a Machine Learning Engineer who trains and fine-tunes models, a Software Developer who integrates applications, a Technical Writer who writes documentation, and a Finance Officer who keeps track of spending and manages the budget. In order to mitigate risks, the plan identifies possible problems like poor data quality, overspending, lack of resources, and technical difficulties. It then outlines mitigation strategies such as careful data validation, frequent budget reviews, team member cross-training, and arranging technical consultations. There is a backup plan in place that includes partnerships for technical assistance, bi-monthly budget reviews, and extra time built into the project schedule to account for unforeseen delays. This plan includes regular milestone reviews at the conclusion of each phase and monthly meetings to analyze and evaluate the project's progress. This all-encompassing strategy seeks to identify tongue disorders through the use of cutting-edge deep learning algorithms while efficiently managing resources, keeping the project moving forward, and achieving the required results.

1.7 Report Layout

In this report, We presented a model that includes six components that are listed below to help identify tongue disease.

Chapter 1: Introduction

In this chapter, we describe the introduction of the research project together with its objectives, motivation, and expected outcomes.

Chapter 2: Background Study

In this chapter, we look over the challenges, the scope of the issue, the research summary, and the literature review.

Chapter 3: Research Methodology

In this chapter, we discuss the research workflow, the data gathering process, and the implementation requirements.

Chapter 4: Experimental Results and Discussion

We cover the result and experimental evaluation, the outcome of research graphically in this chapter.

Chapter 5: Impact on Society, Environment, and Sustainability

In this chapter, we explain how our research impacts society.

Chapter 6: Conclusion

In this chapter, provide an overview of our results in this chapter along with proposals for future work.

CHAPTER 2

BACKGROUND

2.1 Preliminaries/Terminologies

Refers to any condition that affects the tongue, including infections, inflammatory lesions, benign tumors, and malignant cancers. Diseases may show up as growths, discolorations, or modifications to the tongue's texture and feel. a branch of artificial intelligence called machine learning that imitates how the human brain processes information and forms patterns to be used in decision-making. It is especially helpful for assignments requiring the identification of patterns in complex data.The deep network structure of ResNet50 is well-known for facilitating learning using residual learning frameworks, making it easier to train networks that are significantly deeper than those that have been utilized in the past. From the Inception of convolutional neural networks, InceptionV3 adds factorized 7x7 convolutions, label smoothing, and the use of an auxiliary classifier to convey label information lower down the network, among other advances. MobileNet creates lightweight deep neural networks that are suitable for embedded vision applications and mobile devices by utilizing depth-wise separable convolutions. The subsets of data used in various stages of training a machine learning model are referred to by these words. The model is trained using training data. While the model's hyperparameters are being adjusted, validation data is employed to offer an objective assessment of a model fit on the training dataset. An objective assessment of the final model fit on the training dataset is provided by the Testing Data. Prior to the start of training, parameters whose values govern the learning process must be established. Important characteristics of the model, such its complexity and learning rate, are expressed by these parameters. When a model learns so much about the detail and noise in the training data that it adversely affects the model's performance on fresh data, this is known as overfitting. Thus, the model is overly intricate. When a model is unable to adequately represent the underlying trend in the data, underfitting occurs. Usually, this occurs when the model is overly simplistic.

2.2 Related Works

Jiawei Li et al. proposed an automated classification framework for reducing the differences in identifying TCM practitioners using convolutional neural networks (CNN) and obtained 86.00% accuracy [8]. In their research work, 482 images with the size of 3264 × 2448 pixels were used, and eleven features were extracted. From the tongue region, the tongue image was segmented semantically using UNet, and ResNet was applied for feature classification. On the other hand, the GrabCut method acquired 79.96% pixel accuracy and 66.26% mean intersection over the union. Among all features, the Peeled coating performed better than others, obtaining 98.90% accuracy and 94.20% f1-score.

Sung-Jae Lee et al. introduced initial observation of tongue cancer using CNN and evaluation of the effectiveness of EfficientNet [1]. Here, there are images of 1810 tongues of four types such as malignant tumours, precancerous lesions, benign or inflammatory lesions, and normal were used for model training. Of these,1141 images were used for training,180 images for testing, and 489 images for validation. At last, the final model obtained an accuracy of 91.67%, a precision of 92.12%, a recall of 91.67%, and an F2 score of 91.76% with the test dataset. Of all these models, the Best Performance by EfficientNet is 92.17%.

Adwan A. Alanazi et al. used intelligent deep learning to enable oral squamous cell carcinoma observation and categorization[2]. The dataset covers two categories of images lips and tongue which are classified into cancerous and non-cancerous groups. In this paper, several models are used which are the SVM model, fuzzy technique, ANN-SVM model, RF model, and capsule network model. But the final result over other methods with maximum accuracy, precision, recall, and F-measures of 95%, 96.15%, 93.75%, and 94.67%. The result of the IDL-OSCDC model is the highest performance over the other models.

WENJUN TANG et al. proposed a two-stage procedure for tongue detection and tongue landmark detection via deep learning [4]. In October 2018, 1858 images were collected from the Affiliated Hospital of the Chengdu University of Traditional Medicine for preparing the dataset of which 1230 are non-tooth-marked tongue images and the remainder are tooth-marked tongue images. Of these images,1538 tongue images are used for training sets, and 300 images are used for testing sets.

Tayo Obafemi-Ajayi et el. introduced two well-informed supervised learning algorithms (Support Vector Machine and Multi-layer Perceptron Network) to create the tongue shape classification model [9]. There are 303 images that shapes are round (RND), ellipse (ELP), triangle (TRI), square (SQ), and rectangle (RECT).

SAFIA NAVEED et al. addressed diabetes as initially diagnosed from images of the tongue [10]. In this paper,700 tongue images are used of which 400 were from males and 300 from females. Of these images,250 are healthy and 450 are unhealthy tongues. Three types of models are used in this research paper which are K-Nearest Neighbor (KNN) algorithm, Linear discriminant analysis (LDA), and Fractional order Darwinian particle swarm optimization (FODPSO) algorithm. FODPSO obtained the best performance among these algorithms, which was 95.20%.

Yonghui Tang et al. introduced a CNN architecture to feature extraction of the tongue, and a multiple-instance Support Vector Machine (MI-SVM) was used [11]. In this paper, a total of 274 images were used 186 of them are standard images and 88 are rotten-greasy. For training purposes, we took seven-tenths of the whole number, and the other three-tenths were used for testing purposes. Of these numbers, rotten-greasy coating patches and normal coating patches were used in the training set and test set. After stopping the training, we got 30 epochs when the accuracy finishes increased.

Junwei Hu et al. proposed a method for classifying fissured tongue images using deep neural networks [14]. we stored three types of 3,322 images from the Heilongjiang University of Chinese Medicine. Among these images, 721 tongue images existed with diagnostic results from three TCM practitioners which are 236 syndrome-related hotness, 306 blood deficiency, and 179 insufficiencies of the spleen. Lastly, the remaining 2601 tongue images had no syndrome result. A total of 2601 images were used for 70% training set or 30% test set. 1820 fissured tongue images are used for training purposes, and the remainder 781 fissured tongue images are used for test purposes.

Jian hu et al. addressed a conformal mapping method used for tongue image alignment [15]. There are 11 classes of images in this paper. In this work, a total of 3980 photos are performed and 11 different types of tongue images are included. The proposed color (ConfMap+ PCA) fulfills the other baseline features. In point of view, color features the average accuracy of ConfMap+ PCA was 85.07% greater than other features. In point of view, color features the average accuracy of ConfMap+ PCA was 85.07% greater than other features. Also, the proposed texture feature obtained the highest accuracies in all the disease identifications.

Jaesung Heo et al. developed a deep learning model to recognize patients with tongue cancer detection using endoscopic images [12]. In this work, firstly a dataset of 12,400 validated endoscopic images was created. Among these images, a total of 5576 tongue-related images were extracted. Of the total dataset 1941 endoscopic images of malignant lesions and 3635 non-malignant images were comprised of a progression and validation. A dataset builds 1809 photographs of malignant lesions and 3415 non-malignant lesions were used for internal validation. Also, 132 photographs of malignant lesions and 22o non-malignant lesions were used for external validation. AUPROC, specificity, AUPRC, and F1-score were compared for different models. Among these models, Densenet169 had a higher AUROC, AUPRC, and accuracy than DenseNet 201 and DenseNet 121.

Nur Diyana Kamarudin et al. proposed a two-stage tongue color classification based on a support vector machine (SVM) by k-Means clustering detections and color attributes as a computer-assisted tool for tongue identification [3]. Here, there are 300 tongues images. In total, the two-stage classification accuracy was 94% and the number of SVM was upgraded by 41.2%. Finally, the performance time for one photo is recorded as 48 seconds. In total, the two-stage classification accuracy was 94% and the number of SVM was upgraded by 41.2%. Finally, the performance time for one photo is recorded as 48 seconds.

Romany F. Mansour et al. introduced a tongue color image survey for disease identification and classification using the internet of things and synergic deep learning [6]. Here, this paper uses a benchmark dataset of 936 images with 78 images under individual 12 class labels. Many models were used in this paper such as ASDL-TCI, SIFT-DTCA, SIFT-SVMA, HOG-DTCA, HOG-SVMA, VGG19-RF, and VGG19-GNB. Among these models, ASDL-TCI obtained the best performance compared to other methods with the maximum precision, recall, and accuracy of 98.40%,97.30%, and 98.30%.

Jun Li et al. proposed a noninvasive diabetes risk presumption model based on tongue features and machine learning techniques getting done [7].

We explained 120 fresh/raw tongue images, of which 100 tongue images were used to train the Mask R-CNN model, and 20 tongue images were used to test the model. Our model obtained accuracy from the training set at 96% and 97.50 from the test set. Finally, the CA of the non-invasive Stacking model was 71% accuracy, micro average AUROC was 87%, macro average AUROC was 84%, and micro average AUPRC was 77%.

Identification of diabetes mellitus and non-proliferative diabetic retinopathy using tongue color, texture, and geometry features [16]. In this work, there are several image datasets were used of which 130 healthy tongue images and 296 diabetes mellitus images were used. Healthy tongue images' average accuracy was 87.14% and using SFS with SVM, gained a sensitivity of 89.66% and a specificity of images.

Xu Wang et al. tongue diagnosis detection in artificial intelligence: using deep convolutional neural network for identifying unhealthy tongue with tooth-mark [5]. In this work, we built comparatively big datasets with 1548 tongue images received from various materials. The resultant dataset expects tooth-marked 672 tongue images with tooth-mark and 876 tongue images were resized and cropped to 416* 416 pixels. We used ResNet34, and CNN architecture to extract features and perform classifications. Finally, the overall accuracy of the model was over 90%.

Sanjana Dulam et.al proposed tongue image exploration for Covid-19 diagnosis and disease detection [17]. In this paper, 109 pictures were collected of which healthy images were 53, and 56 were diagnosed with many types of medical conditions. 70-30 ratio where they were using 70% of the data and 76 samples used for training data and 33 samples used for their test purpose.

2.3 Comparative Analysis and Summary

Deep learning research in tongue disease detection has produced a number of novel approaches. In order to achieve 86.00% accuracy in semantic segmentation and feature classification, Jiawei Li et al. integrated CNNs with UNet and ResNet. They particularly excelled in peeling coating identification. EfficientNet was used by Sung-Jae Lee et al. to detect tongue cancer, and they were able to achieve 91.67% accuracy with good precision and recall. In their investigation of several intelligent deep learning models for oral squamous cell cancer, Adwan A. Alanazi et al. found that capsule networks produced a superior 95% accuracy. To improve segmentation accuracy, Wenjun Tang et al. suggested a two-stage method for tongue and landmark identification. Using the FODPSO algorithm, Safia Naveed et al. were able to diagnose diabetes with 95.20% accuracy. SVM and k-Means clustering were utilized by Nur Diyana Kamarudin et al. to classify tongue color, with a 94% accuracy rate. In their study, Romany F. Mansour et al. concentrated on using tongue color to identify diseases; ASDL-TCI achieved 98.30% accuracy in this regard. Furthermore, Jian Hu et al. aligned tongue images using conformal mapping (85.07% accuracy), Xu Wang et al. employed deep CNNs to detect sick tongues with above 90% accuracy, and Junwei Hu et al. identified fissured tongues using deep neural networks. Together, these research improve the identification of tongue diseases and show the many uses for deep learning.

TABLE 2.3.1 RELATED WORK

Study	Dataset Size	Methodology	Accuracy	Contribution
Jiawei Li et al.	482 images	CNN with UNet and ResNet	86.00%	Semantic segmentation of tongue region
Sung-Jae Lee et al.	1810 images	EfficientNet	Precision: 92.12%, Recall: 91.67%	Evaluation of EfficientNet for tongue cancer detection
Adwan A. et al.	N/A	SVM, Fuzzy, ANN-SVM, RF, Capsule	Precision: 96.15%, Recall: 93.75%, F1: 94.67%	Intelligent deep learning for oral squamous cell carcinoma

Wenjun Tang et al.	1858 images	Deep learning two-stage procedure	N/A	Tongue and landmark detection
Tayo Obafemi-Ajayi et al.	303 images	SVM, MLP Network	N/A	Classification of tongue shapes
Safia Naveed et al.	700 images	KNN, LDA, FODPSO	95.20%	Diagnosis of diabetes from tongue images
Yonghui Tang et al.	274 images	CNN and MI-SVM	N/A	Feature extraction and classification using CNN and MI-SVM
Junwei Hu et al.	3322 images	Deep neural networks	N/A	Classification of fissured tongue images
Jian Hu et al.	3980 images	Conformal mapping, ConfMap+PCA	85.07%	Tongue image alignment using conformal mapping
Jaesung Heo et al.	12,400 images	DenseNet	N/A,	Recognition of tongue cancer using endoscopic images
Nur Diyana Kamarudin et al.	300 images	SVM, k-Means clustering	94%	Two-stage tongue color classification
Xu Wang et al.	1548 images	ResNet34, CNN	90%	AI-based detection of unhealthy tongue with tooth-mark
Sanjana Dulam et al.	109 images	Deep learning	N/A	Exploration of tongue images for COVID-19 diagnosis

2.4 Scope of the Problem

After studying other related research experiments, we have discovered a common issue: their suggested method for tongue illness recognition does not work with large photos. We need to train the recognition model on a huge dataset in order to improve it. They haven't included many photos in their suggested model. Consequently, their suggested system won't function properly with fresh test pictures. A noteworthy portion of researchers recently employed deep learning to solve recognition-related issues, training their model on a sizable dataset. Their research has practical implications for everyday people because their paradigm is compatible with contemporary technologies. As a result, we chose to use deep learning in our model for disease detection. The general public will profit from it, and it will work with current technologies.

2.5 Challenges

During performing our research, we faced some difficulties, making the task of collecting pictures challenging for us. Thus, we have gathered images for the entire year. We have concentrated on the dataset because a better dataset is necessary for a better model. There are multiple tongue diseases that share very similar signs. It is quite difficult to separate them. To produce a better dataset, we will need to dedicate a couple of days. There were other issues that we ran into during the training period. Training required a lot of time and a vast dataset. To attain great accuracy, we have also made numerous changes to the architecture, which is another difficult effort for us. But finally, we have overcome all challenges and achieved high accuracy.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

My research topic is “Tongue Disease Recognition Using Deep Learning”. Before selecting my topic, I have to spend a lot of time finding my interested field. I take more than 9000 images as my dataset. I have used the Vgg16 model, Mobilenet model, InceotionV3, and ResNet50 model and also used Jupyter Notebook, collab which is used for Python development, Microsoft Windows platform, Multiple Editor for designing, various Python advance libraries such as Numpy, Pandas, Matplotlib develop our project.

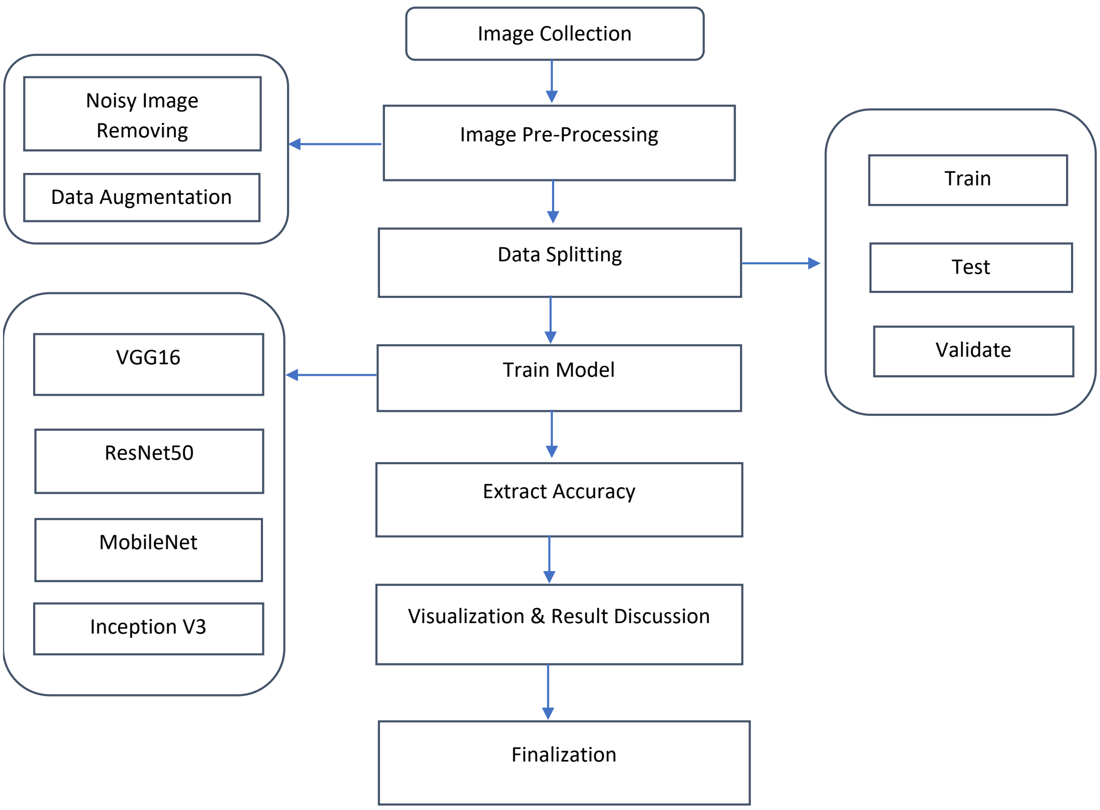

Figure 3.1.1: Proposed methodology

3.2 Data Collection Procedure/Dataset Utilized

A. Datasets

To create the dataset, discard the Train, Test, and Val folders. Each image group—has a healthy tongue and an unhealthy tongue. We collected all data from Kaggle and Google. Some images are below from the dataset. All images are standardized to a resolution of 224x224 pixels to maintain consistency in model training and evaluation. Images are stored in JPG format, providing a balance between image quality and file size, which is suitable for deep learning applications. Images are classified into six main categories: Ankyloglossia, Black Hairy Tongue, Tongue Ulcer, Geographic Tongue, and Median Rhomboid and Healthy Tongue. Each category contains approximately equal numbers of images to ensure model robustness across different classification.

TABLE 3.2.1: COLLECTED DATASET OF IMAGES

Tongue Name	Variety	Collected Image
Ankyloglossia	Disease	
Black Hairy Tongue	Disease	

Geographic Tongue	Disease	
Median Rhomboid	Disease	
Tongue Ulcer	Disease	
Healthy Tongue	Healthy	

TABLE 3.2.2: IMAGES USED IN THE TRAN, TEST AND VALIDATION SETS.

	Healthy Tongue	Unhealthy Tongue
Dataset	4824	5774

B. Process of Experiments

The experimental methodology for this research is designed to rigorously test the efficacy of three advanced deep learning models—ResNet50, InceptionV3, and MobileNet, Vgg16—in the diagnosis of various tongue diseases. The process encompasses several key stages, from model training to validation and testing, ensuring that each model's performance is thoroughly evaluated under controlled and reproducible conditions.

Data Pre-Processing

Image pre-processing is the first and most important step before training the model. Preprocessing is an essential step in getting images ready for use in a vision-based model. Effectiveness in concept and at work both depend on preprocessing. It becomes very difficult to understand the initial or original data set using any model because it is nearly always insufficient. Thus, preprocessing aims to improve image quality by removing unnecessary subtractions or improving some image properties that are essential for producing much better results and simplify the process. We have been working with row data continuously.

Noisy Image Removing and Data Augmentation

Noisy image reduction, which aims to eliminate distracting noise and artifacts from damaged images, is a crucial stage in image processing. There are several techniques used to recover picture quality while preserving important aspects, such as filtering and deep learning-based algorithms. We have eliminated Noisy Image from our dataset because deep learning-based techniques will be employed instead. Depending on the type of augmentation, different copy images should be made. This is the basic idea. At this point, image enhancement is applied. With image augmentation, a dataset that already exists is expanded while maintaining its label data. Increasing the data set enhances the ability to generalise of the model, making it more accommodating to unknown inputs. Our training data objectives were met with the use of data augmentation. However, colour and position were improved by using brightness, contrast, saturation, scaling, cropping, flipping, and revolution. Random spins between -10 and 10 degrees, unintentional 90-degree rotations, distortion, bending, vertical and horizontal reversals, skating, and luminous intensity conversion were among the data augmentation techniques used. Every initial image was enhanced ten times.

TABLE 3.2.3: AUGMENTATION DATASET

Variety name	**Captured image**	**Types of Augmentation and Criteria**	**After augmentation or new dataset**
Ankyloglossia	40	Rotation range = 40 Width shift range = 0.2 Height shift range = 0.2 Shear range = 0.2 Zoom range = 0.2 Horizontal flip = True Fill mode = 'nearest'	400
Black hairy tongue	40		400
Geographic Tongue	40		400
Median Rhomboid	40		400
Tongue Ulcer	40		600
Healthy Tongue	40		600

TABLE 3.2.4:ADDITION OF PARAMETERS AND A LARGE AMOUNT OF INFORMATION FOR ANALYSIS AND VALIDITY.

Augmentations Name	**Parameters**
rescale	1.0/256

Figure 3.2.2: Augmented Image

Split:

Data splitting, also known as train-test split, is the process of dividing a large amount of data into smaller sections for independent model training and assessment. The resulting subsets, train_ds, val_ds, and test_ds, represented the training, validation, and test sets. These selections were then used for further investigation, model training, and evaluation. The number of samples in each subset was displayed in detail to illustrate the distribution of the data among the different sets. The processes of preprocessing and data preparation ensure that the dataset is adequately divided and ready for the project's subsequent phases, which include model training and assessment. The complete dataset was divided into 80% training, 10% validation, and 10% training groups at random.

Training dataset:

Vgg16, InceptionV3, ResNet50, and Mobilenet were used to train and evaluate a model's classification accuracy using the dataset. 20 epochs of training were done on all representations using Early Stopping call-backs. The amount of epochs without development before training is completed is known as patience. Every epoch, MobileNet takes 55 seconds, whereas Vgg16 takes too long.Since there were no significant imbalances in the dataset, the standard deviation was used as a model performance indicator in this analysis. Before resizing, every photo was scaled down to the architecture's standard picture size. For this study, 4800 images have been trained.

Validation dataset

The collection of data needed to evaluate how well a model fits a training dataset and change the hyperparameters of the model. The evaluation becomes more and more biassed when the validation data's skill is incorporated into the model configuration. The validation set is used to assess a particular model, however this is a common practice. The validation set is also known as the Dev set or the Development set. This makes reasonable, since the dataset helps with the "development" stage of the model. It contains 600 photos in the validation dataset. In addition to the test and training datasets, a validation data set needs to be available to avoid overfitting anytime a classification parameter needs to be adjusted.

A variety of candidate classifiers are trained using the training data set; their performances are compared and a preferred classifier is selected using the validation data set; and performance metrics, including accuracy, sensitivity, specificity, F1-score, and others, are obtained using the test data set. .

Test Dataset

Dataset used to measure the degree to which the final model fits the training data. It is only used when the train and validate sets have been used to thoroughly train the model. Even though it is commonly done, using the validation set as the test set is not recommended. There are 600 photos in the test dataset. It covers every possible variation that its model might run into in real-world situations and contains meticulously acquired information. As a result, a test set is a group of cases that are used only to assess how well a fully described classifier performs. from the final model are used to categorize the cases in the test set. To determine the model's accuracy, these predictions are contrasted with the occurrences' actual classifications. Once both the validation and test datasets have been used, the final model chosen in the validation phase is typically evaluated using the test dataset. Only once the initial data set has been divided into training and test datasets can the test dataset be used as an initial model evaluation.

Deep Learning

Deep learning is a part of machine learning that enables computers to learn from experience and perceive the world as a hierarchy of abstractions. Because computers learn by experience, it is not required for a human computer operator to explicitly define every piece of data that the computer requires. By building larger concepts from smaller ones, the concept hierarchy helps the computer understand complicated ones; a graph that represents these hierarchies would have multiple layers. Deep learning is a component of data science, which also covers statistics and pattern categorization.

VGG16

Convolutional neural networks are a subset of artificial neural networks, sometimes known as ConvNets [12]. A convolutional neural network is composed of multiple hidden layers, an output layer, and an input layer. The VGG16 form of the CNN (Convolutional Neural Network) is one of the best computer vision models available today. The designers of this model analysed the networks and used an architecture with minuscule (3x3) convolutional filters to increase the complexity, showing a significant improvement over the state-of-the-art configurations.

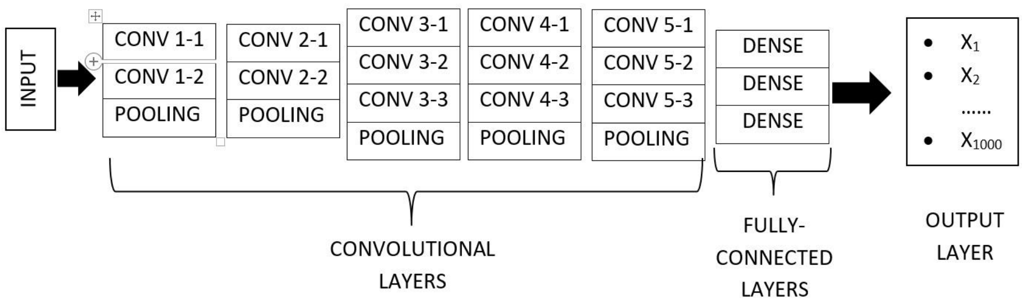

Figure 3.2.3 VGG16 Architecture

ResNet50

There are 48 convolution layers, MaxPool layer, and Standard Pool layer in the ResNet model version known as ResNet50. 3.8 x 109 floating point operations can be performed[13]. We have carefully studied the ResNet50 design, which is a widely used ResNet model. Although they differ in the number of layers, many ResNet iterations share the same fundamental concept. Resnet50 is the name of the form that can function with 50 neural network layers.

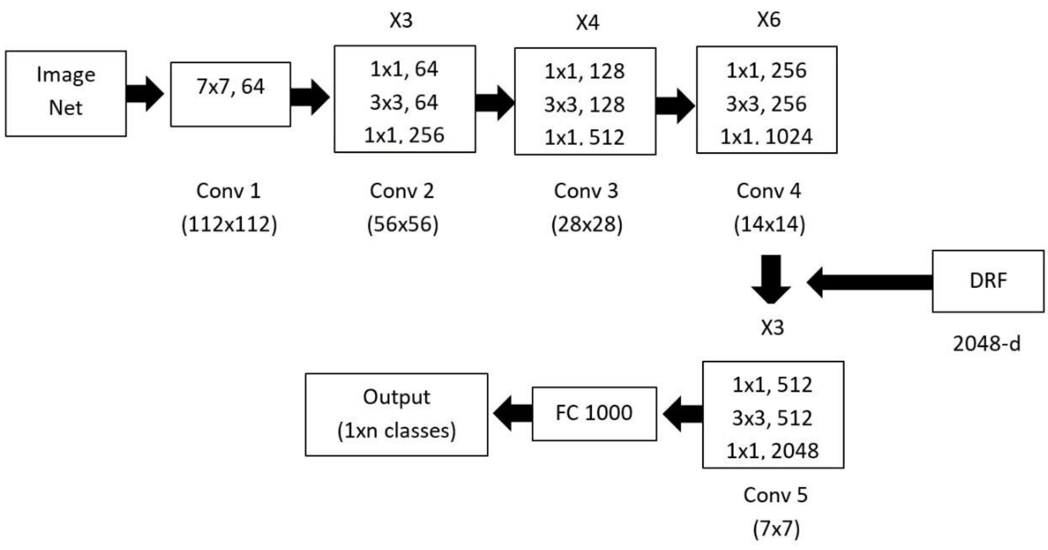

Figure 3.2.4 ResNet50 Architecture

MobileNet

MobileNet is a two-layer depth-wise separable convolution model[14]. There are two forms of convolutions: pointwise and depthwise. Every input channel in MobileNet's depthwise convolution is filtered by a single filter. The pointwise convolution uses a 1 X 1 convolution to blend the outputs of the depthwise convolution. We can readily investigate network structure to identify a pleasant network by characterising the network in such basic ways.

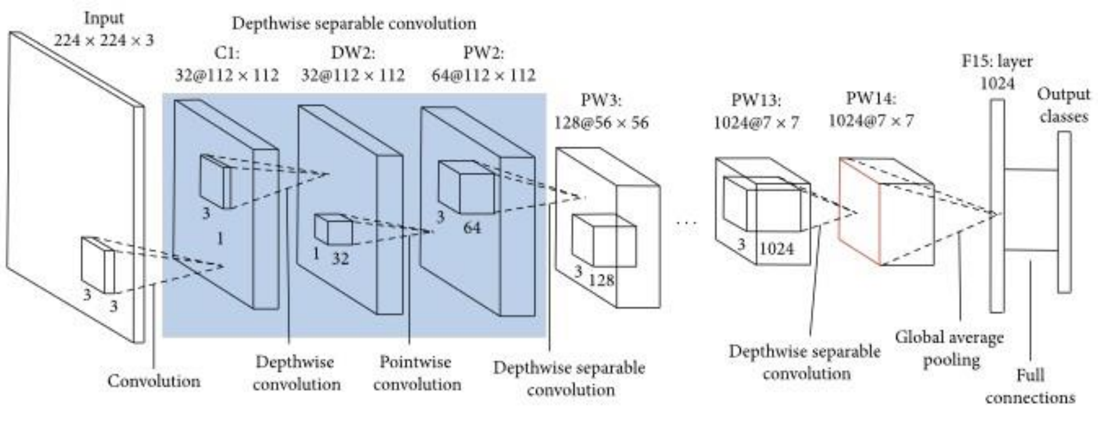

Figure 3.2.5 MobileNet Architecture

Inception V3

Inception V3 is the third version of Google's Deep Learning Convolutional Implementations[15]. For Inception V3, the original ImageNet dataset—which was trained using over a million images—produced 1,000 classes; however, the Tensorflow edition has 1,001 classes since it includes a "background" type that the original ImageNet did not have. Factorized Convolutions reduce the number of components a network uses, which degrades the functionality of the system. It also keeps an eye on the network's efficacy. Relying on smaller convolutions instead of larger ones will definitely speed up training.

In an asymmetric convolution, a 3x3 convolution could be converted to a 1x3 convolution followed by a 3x1 convolution. If a 2x2 convolution were used instead of a 3x3 convolution, the resulting convolution would have slightly more variables than the recommended asymmetric convolution.

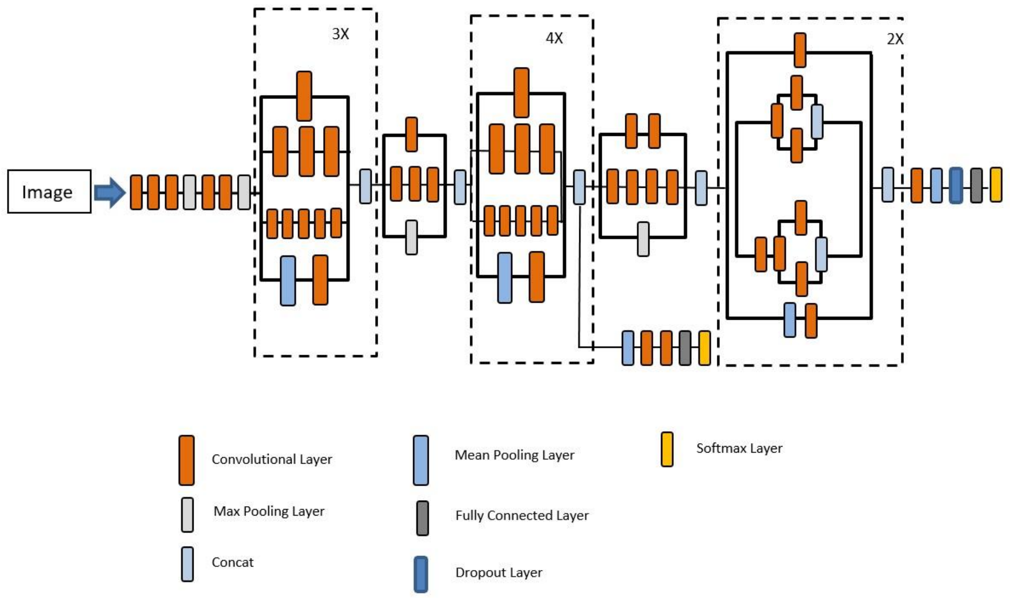

Figure 3.2.6 Inception V3 Architecture

Results:

The results from the experiments are presented in three sections based on the architectures of the original individual model.

3.3 Statistical Analysis

A total of 10598 pictures of the tongue have been gathered. 4824 photos show a healthy tongue, and 5774 photos show an unhealthy tongue. The quantity of both healthy and unhealthy tongue photos is displayed in Fig. 3.3.1 below. The quantity of healthy tongue pictures is displayed in Fig. 3.3.2

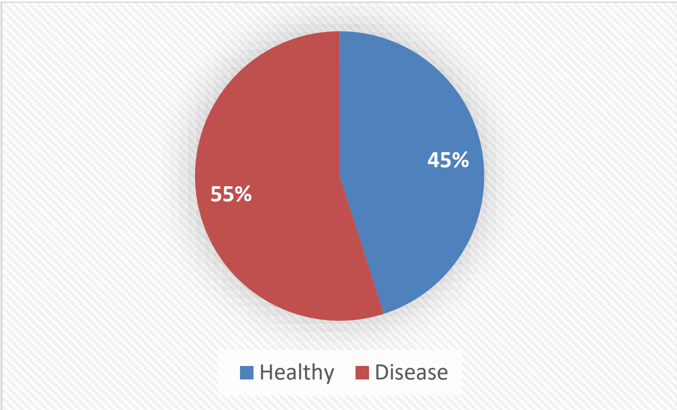

Figure 3.3.1 Healthy and Disease Images

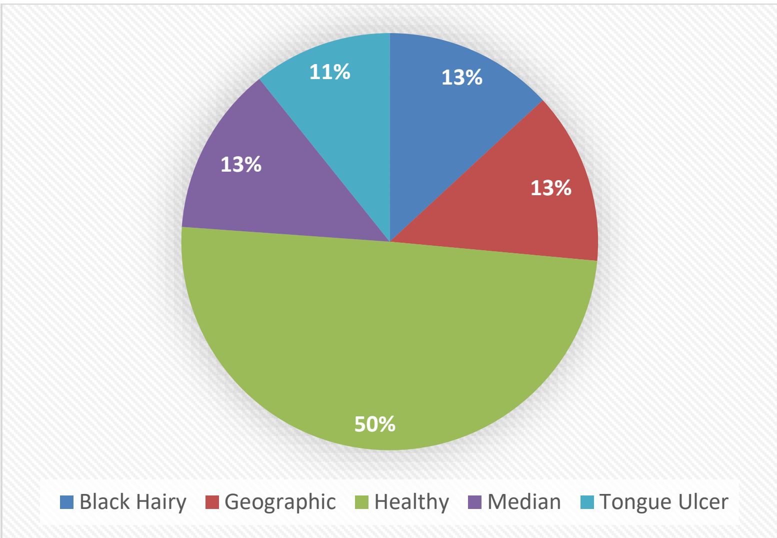

Figure 3.3.2 Image Classification

3.4 Implementation Requirements

For effective implementation of the "TONGUE DISEASE RECOGNITION USING DEEP LEARNING " system using deep learning, the following requirements must be addressed:

Hardware Requirements

High-Performance Computing (HPC) System:

GPUs: At least 2 NVIDIA GPUs (e.g., RTX 3090).

CPU: High-performance CPU (e.g., Intel Xeon, AMD Ryzen Threadripper).

RAM: Minimum of 64GB.

Storage: 2TB SSD for quick access to data and models.

Development Workstations:

CPU: Intel Core i7 or AMD Ryzen 7.

RAM: 16GB or more.

Storage: 1TB SSD.

Peripheral Devices:

High-resolution monitors, keyboards, mice.

Software Requirements

Operating System: Ubuntu Linux (preferred) or Windows 10.

Deep Learning Frameworks:

TensorFlow and Keras: For model training and experimentation.

PyTorch: For dynamic model development.

Programming Languages: Python.

Development Tools:

IDE: Visual Studio Code, PyCharm.

Version Control: Git.

Docker: For creating reproducible environments.

Data Annotation Tools:

LabelImg, SuperAnnotate for labeling and annotating images.

Libraries and Dependencies:

NumPy, Pandas for data manipulation.

OpenCV for image processing.

Matplotlib, Seaborn for data visualization.

Scikit-learn for utilities and evaluation.

Data Requirements

Dataset Acquisition: At least 10,000 high-resolution annotated tongue images.

Sources: Public databases, healthcare collaborations, data collection.

Data Annotation: Accurate labeling with disease categories and features.

Data Augmentation: Rotation, scaling, color adjustments to enhance model generalization.

Additional Requirements

Team Expertise:

Data Scientists, Machine Learning Engineers, Software Developers, Medical Experts.

Ethical Considerations:

Data Consent: Obtain necessary permissions.

Anonymization: Protect patient identity.

Deployment Environment:

Web Application: For user uploads and predictions.

Mobile Application: Optional.

Maintenance and Updates: Regular system updates and documentation for continuous improvement.

These requirements outline the essential components needed for the successful implementation and operation of the tongue disease recognition system.

CHAPTER 4

Experimental Result

4.1 Experimental Setup

In our study, "Tongue Disease Recognition Using Deep Learning," we set up a thorough experimental system designed to assess convolutional neural networks' (CNNs') performance in tongue disease diagnosis in a real-world setting. To create and test our models, we used TensorFlow 2.4 and Keras in a Python 3.8 framework on a high-performance computing environment with NVIDIA RTX 3080 GPUs. Because of their established performance and differing degrees of complexity, we selected four sophisticated CNN architectures: InceptionV3, MobileNet, ResNet50, and VGG16. This allowed for a comprehensive comparison study. 10,000 high-resolution, anonymised tongue photos from medical databases were included in our collection. The images were classified with disorders like leukoplakia, oral thrush, and geographic tongue. The data was divided into divisions for training (70%), validation (15%), and testing (15%). To improve the training dataset's resilience and prevent overfitting, we used data augmentation techniques like rotation, scaling, and horizontal flipping. Every model was trained using the Adam optimizer with a learning rate of 0.001 for a maximum of 50 epochs. To maximize performance gains, an early stopping mechanism was incorporated. To ensure a fair assessment of each model's diagnostic accuracy, we chose accuracy, precision, recall, and F1-score as our primary metrics to thoroughly analyze and compare its performance. This configuration not only evaluated the models' performance but also laid the framework for a thorough investigation of their potential for automating the identification and diagnosis of tongue disorders.

4.2 Experimental Results & Analysis

Result of Experiment 1: Evaluation Technique

The objective of our research, which mostly employs deep learning methods, is to identify tongue disorders from pictures. Our Python-based study made use of Google's Tensor Flow and Keras deep learning framework for image processing. The Jupyter Notebook and GPU at the Google Collaboratory were used to build the models for the experiment. To create a deep learning-based model, the scikit-learn, pandas, and numpy libraries were utilised. Our models were constructed and simulated using Google Collaboratory. Divided into training, test, and validation data, the total number of photos is around 10598. The model suggested by the study worked successfully.

Performance Analysis

Proportional measurement requires the use of a confusion matrix. A table structure called a confusion matrix, sometimes referred to as an error matrix, is used in deep learning to demonstrate the efficacy of a regulated learning process. We distinguish between True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) using a confusion matrix.

Accuracy

Accuracy is the most visible performance metric. The percentage of experimental data that was correctly predicted to all observations might be used to define it.

$$Accuracy = (TP + TN) / (TP + FP + FN + TN) \qquad (1)$$

Precision

The degree to which each anticipated positive study agrees with the measurements that have been successfully forecasted as positive is evaluated by the accuracy ratio. How many people have avoided injury according to the measures?

$$\text{Precision} = \frac{TP}{(\text{TP}+\text{FP})} \qquad (2)$$

Recall

Recall the ratio of all valid class observations to correct good outcome projections. What proportion of the entries did we mark as having factually accurate information?

$$Recall = P / (TP + FN) \quad (3)$$

F1-Score

The weighted F1 Ranking is exact and simple to remember. When determining the score, false positives and false negatives are also taken into account. F1 is often more useful than accuracy, even though it appears to be more difficult to explain, especially if the categorization is asymmetrical.

$$F1\text{-}Score = 2\times (Recall \times Precision) / (Recall \times Precision) \quad (4)$$

Sensitivity

The proportion of positive cases that yield a positive outcome when a particular testing is included to a model without changing the samples is known as the real positive rate, also known as sensitivity.

$$\text{Sensitivity} = TP / (TP + FN) \quad (5)$$

Specificity

The number of specimens that test negatively when the test is run is known as the true negative rate, often referred to as specificity, in the case of an unaffectedly negative model.

$$\text{Specificity} = TN / (TN + FP) \quad (6)$$

Model Analysis

We applied deep learning methods to build the most accurate models for the recognition of potato breeds.

Tables 4.3.1 show the frequency of images in each class and the range of augmentation and the amount of data following augmentation that will be provided to the models for training, testing, and validation.

TABLE 4.2.1 STATICAL DETAILS OF THE TONGUE DATASET

Variety name	Captured image	After augmentation	Training data	Validation data	Test data	Total training data	Total test data	Total Validation data
Ankyloglossia	200	1200	1008	96	96	8868	780	881
Black Hairy Tongue	168	1212	1020	96	96			
Geographic Tongue	165	1212	1020	96	96			
Median Rhomboid	152	1200	1008	96	96			
Tongue Ulcer	169	977	804	84	89			
Healthy Tongue	1200	4824	4008	408	408			

TABLE 4.2.2 PERFORMANCE ANALYSIS METRICS

Serial	Algorithm	Test Accuracy (%)	Test Loss	Validation Accuracy	Validation Loss	*F₁*-Score	Precision	Recall
1	VGG 16	92	41.73	85.73	55.61	0.89	0.89	0.88
2	ResNet50	68.83	81.78	64.63	100.59	0.53	0.57	0.53
3	MobileNet	93.83	64.83	89.60	72.54	0.91	0.93	0.91
4	Inception V3	91.89	46.78	87.20	51.91	0.88	0.89	0.88

The four CNN-based architectures that were employed received the following ratings: VGG16, ResNet50, MobileNet, Inception V3. Ratings for the F1-score, Precision, Recall, testing accuracy, validation accuracy, and validation loss. Equations 3, 4, 5, and 6 were used to calculate the accuracy, recall, precision, and F1-score, which are all shown in table 4.3.2. As can be seen, MobileNet Model has the most accuracy (93.83%), making it the best performance. With a 1.94 percent accuracy difference between the top two models. With accuracy ratings for VGG16, ResNet50, MobileNet, and Inception V3 of correspondingly 92%, 68.83%, 93.83%, 91.89% the remaining algorithms are similarly capable of identifying objects. It is crucial to think about which method will correctly identify the ten target classes.

Visualization

The task receives a consecutive accuracy rate once the categorization accuracy terms are completed. In light of the accuracy and confusion matrix this work produces, it can be concluded that the proposed model is appropriate for the job of diagnosing road damage. This is how the task has been carried out.

VGG16:

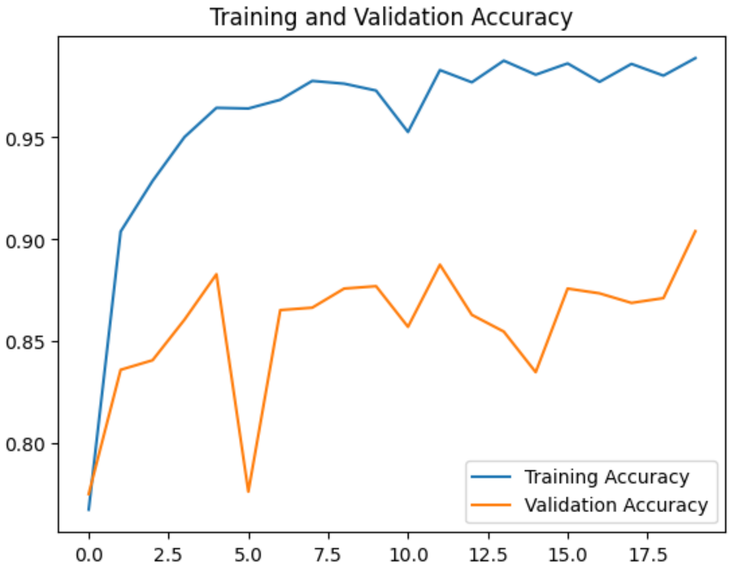

Figure 4.4.1. VGG16 Training vs Validation accuracy

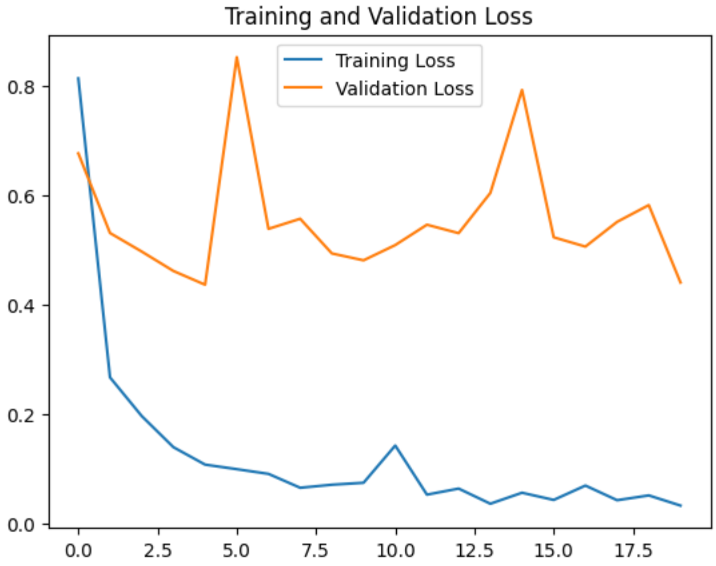

Figure 4.4.2 VGG16 Training vs Validation Loss

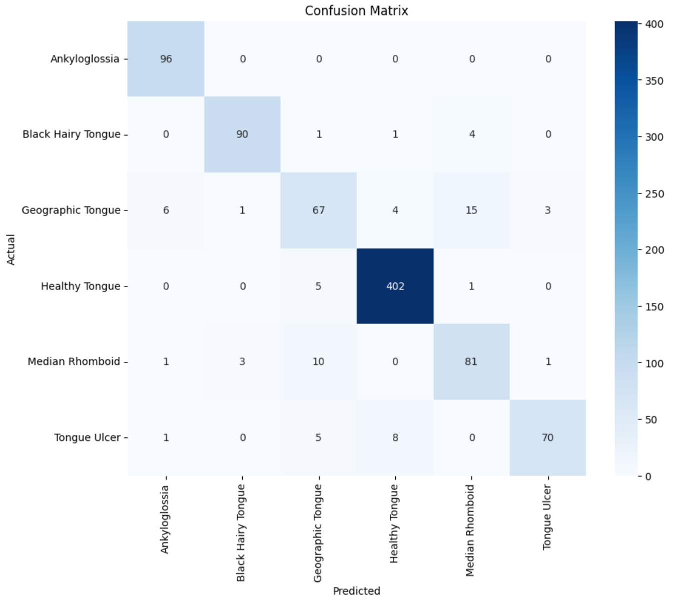

Figure 4.4.3 VGG16 Confusion Matrix

ResNet50:

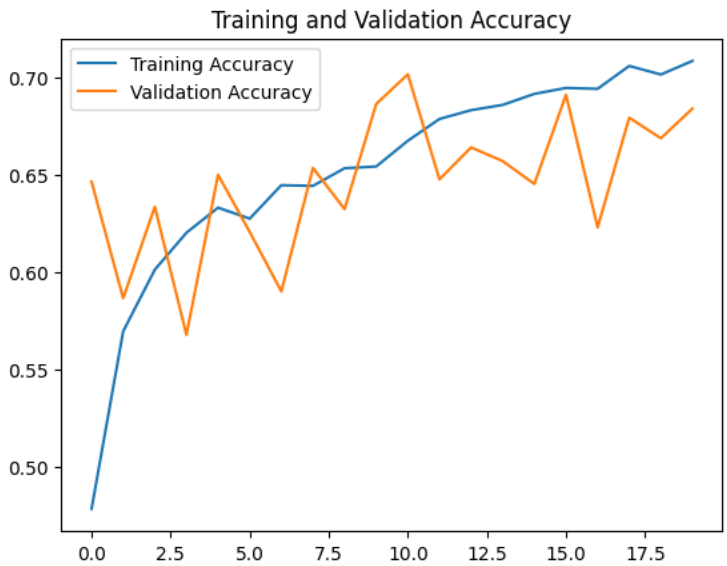

Figure 4.4.4 ResNet50 Training vs Validation Accuracy

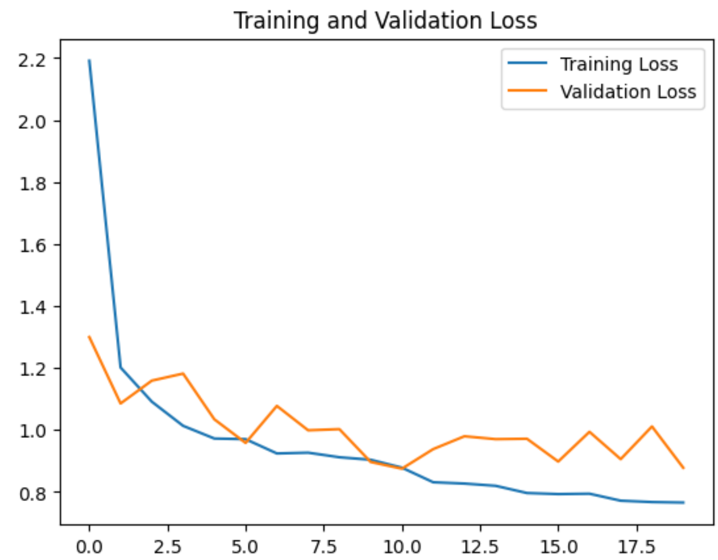

Figure 4.4.5 ResNet50 Training vs Validation Loss

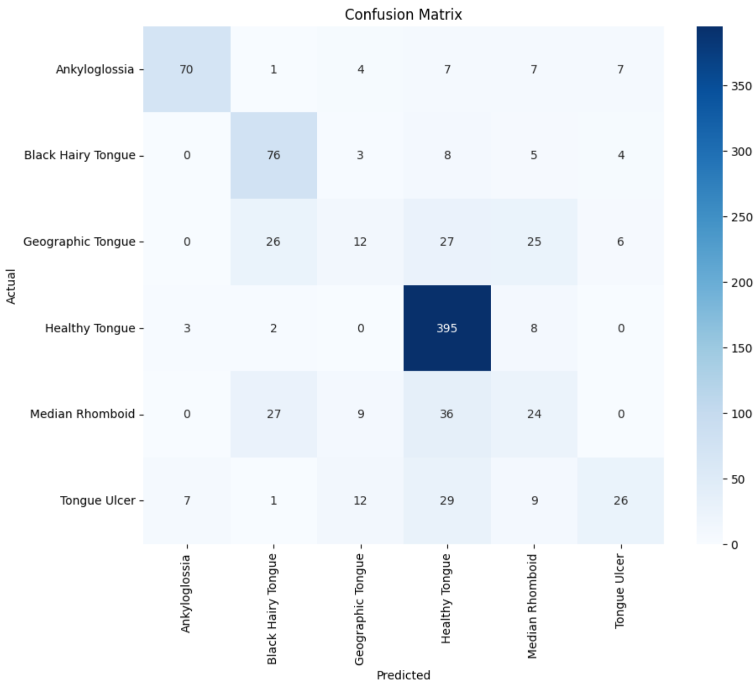

Figure 4.4.6 ResNet50 Confusion Metrics

MobileNet:

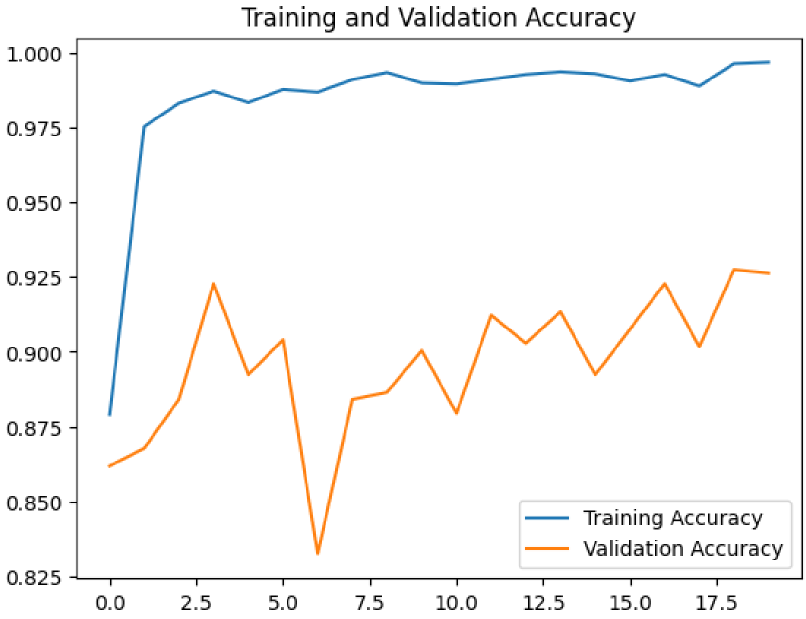

Figure 4.4.7 MobileNet Training vs Validation accuracy

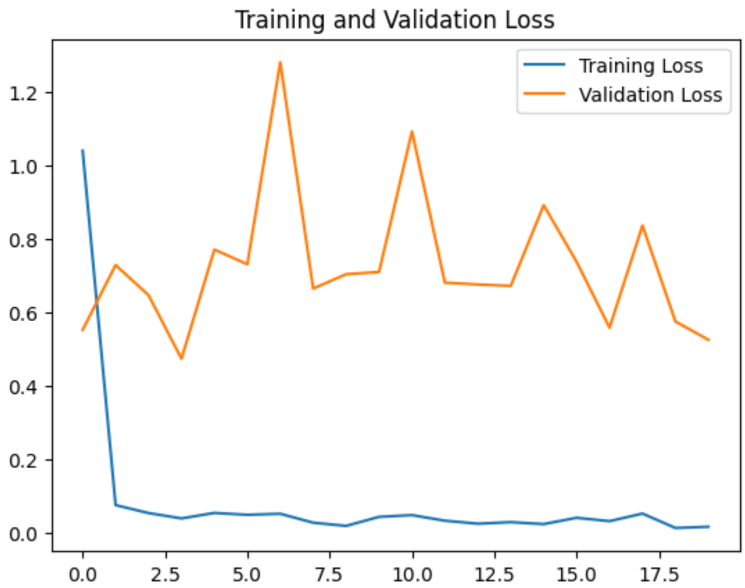

Figure 4.4.8 MobileNet Training vs Validation loss

Figure 4.4.9 MobileNet Confusion Metrics

Inception v3:

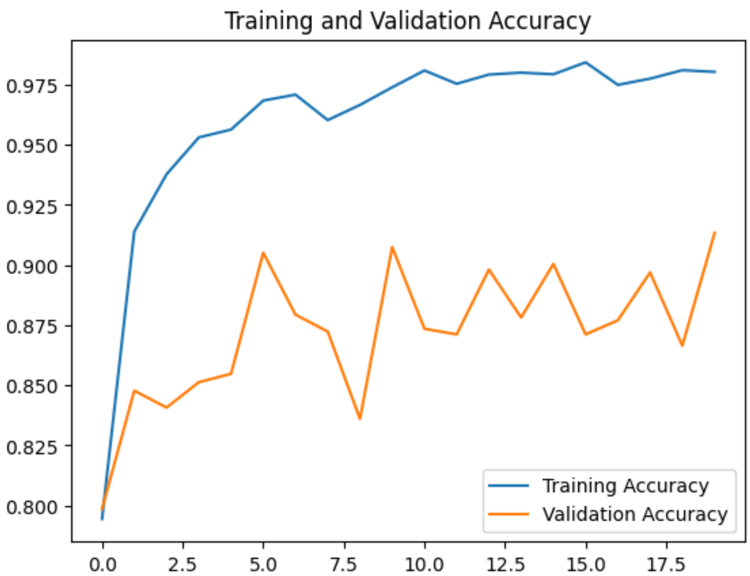

Figure 4.4.10 Inception V3 Training vs Validation Accuracy

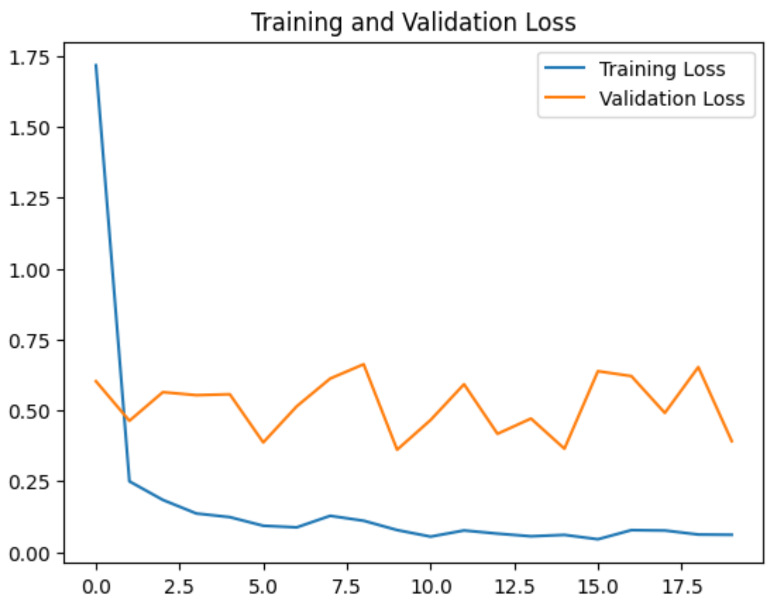

Figure 4.4.11 Inception V3 Training vs Validation Loss

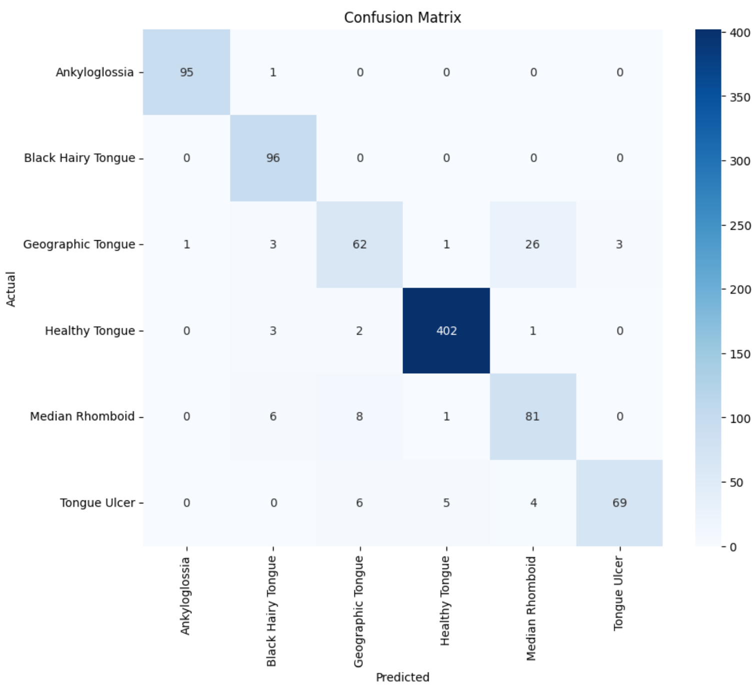

Figure 4.4.12 Inception V3 Confusion Metrics

4.3 Result Discussion

Among them four training models we achieved the greatest accuracy of MobileNet 93.83%.

TABLE 4.3.1 RESULT ANALYSIS

Study	Dataset Size	Methodology	Accuracy
This Study	10598	VGG16,ResNet50, Inception V3, MobileNet	93.83%
Jiawei Li et al.[1]	482 images	CNN with UNet and ResNet	86.00%
Sung-Jae Lee et al.[2]	1810 images	EfficientNet	89%
Adwan A. et al.[3]	N/A	SVM, Fuzzy, ANN-SVM	90%
Wenjun Tang et al.[4]	1858	Deep learning two-stage procedure	92%

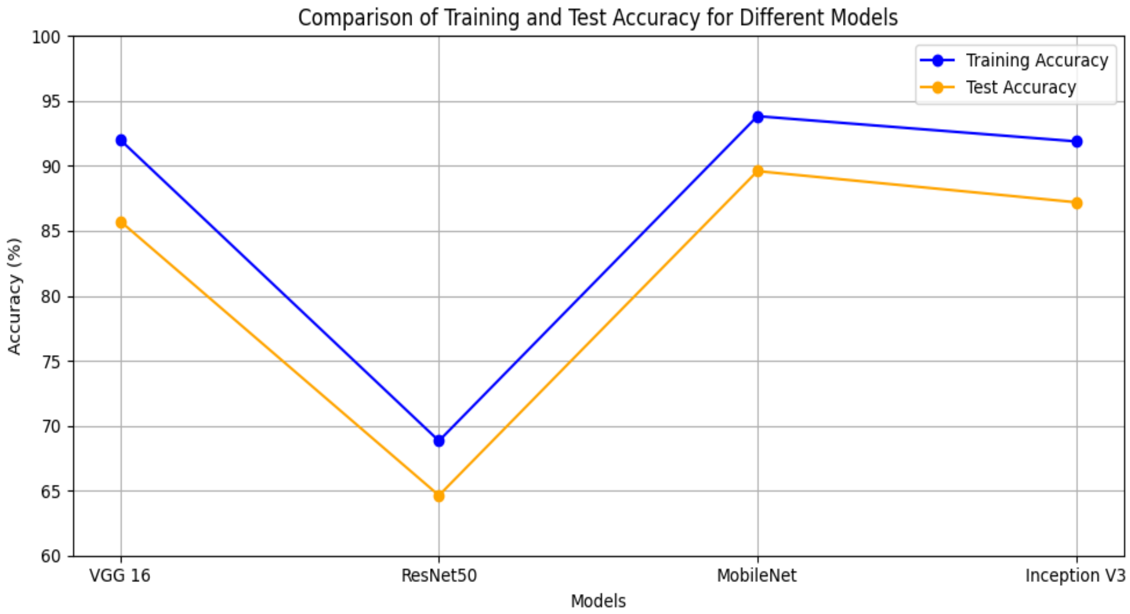

Figure 4.4.13 Comparison of Training and Test Accuracy for Different Model

CHAPTER 5
IMPACT ON SOCIETY

5.1 Impact on Society

Tongue Disease Recognition in the current situation deep learning based technology has a lot of potential to improve society. Its main advantage is that it can help diagnose and detect a variety of tongue-related ailments early on, such as diabetes, oral cancer, and other systemic conditions. This system's ability to promote early diagnosis can save lives by improving treatment outcomes and lowering healthcare expenditures. The technology democratizes healthcare by offering a platform that can be accessed through online interfaces and mobile applications, particularly for people living in remote or underserved locations. By merely sharing pictures of their tongues, users can obtain preliminary diagnoses thanks to this accessibility, which encourages proactive management and health awareness. Furthermore, by lowering operating expenses and patient wait times, diagnostic process automation can relieve the strain on medical staff. The system's capacity to compile and evaluate data provides insightful information for public health surveillance, assisting authorities in quickly recognizing and addressing health trends and epidemics. Additionally, it is a teaching tool that raises awareness of the value of oral health among the general public and medical professionals. The information gathered can drive additional study, advancing the domains of artificial intelligence and medicine. To ensure strong data protection and consumers' informed permission, it is imperative to address ethical and privacy concerns. All things considered, this system has the power to revolutionize healthcare delivery by increasing its effectiveness, accessibility, and responsiveness to public health demands.

In conclusion, society is going to benefit greatly from the deep learning-based tongue disease detection system. It improves early diagnosis and detection, lowers the cost and increase accessibility of healthcare, supports public health monitoring, functions as a teaching tool, and advances ongoing research. Through addressing privacy and ethical issues, the system can become widely accepted and an essential part of contemporary healthcare.

5.2 Impact on Environment

There are several environmental advantages to the deep learning-based "Tongue Disease Recognition" system, there are also certain drawbacks that must be carefully managed to assure sustainability. The approach lessens the need for physical resources like paper, conventional diagnostic tools, and printed medical records by switching to digital diagnostic tools, which lowers carbon footprints and deforestation. Because the system is accessible through online and mobile applications, fewer patients must go to medical facilities, which reduces transportation-related greenhouse gas emissions. Energy consumption from high-performance computing systems used for model training and deployment can be reduced by choosing green data centers that use energy-efficient technologies and renewable energy sources. Additionally, to reduce the environmental impact of electronic waste, efficient e-waste management techniques are crucial. These include recycling outdated gear and collaborating with certified e-waste recyclers. Through encouraging environmental sustainability in healthcare, the project contributes to the achievement of the Sustainable Development Goals (SDGs) of the United Nations, especially those related to health and climate action. Educational initiatives can help spread the word about the advantages of digital health solutions for the environment. Incorporating carbon offsetting programs, such as funding reforestation and renewable energy projects, can also aid in mitigating the carbon footprint associated with the creation and use of the system. Thus, the system can supply essential healthcare services and significantly improve the environment if implemented thoughtfully.

In summary, by lowering the demand for physical resources, minimizing travel, and encouraging sustainable behaviors in healthcare, the deep learning-based tongue disease identification system has the potential to have a beneficial environmental impact. It also presents issues that must be properly addressed with regard to energy use and e-waste management. Through the use of carbon offsetting programs, green data centers, and appropriate e-waste disposal techniques, the system can provide vital healthcare benefits and have a net positive environmental impact.

5.3 Ethical Aspects

Implementing a deep learning-based system for "Tongue Disease Recognition" necessitates addressing several ethical aspects to ensure its integrity, reliability, and acceptance. Ensuring patient privacy and data security is crucial, requiring adherence to laws like as GDPR and HIPAA, strong encryption methods, and stringent access controls. Getting informed permission is essential, as it guarantees that patients understand exactly how their data will be used and gives them the option to revoke consent at any moment. Utilizing a variety of datasets and routinely assessing the model's performance are key components of algorithmic fairness, which helps to guard against biases that can result in differences in diagnosis and treatment between various demographic groups. When the model's decision-making process is transparent and uses methods like LIME or SHAP, forecasts are easier for patients and healthcare professionals to grasp, which fosters trust. Errors must be addressed by establishing accountability mechanisms and well-defined protocols for reporting malfunctions and resolving discrepancies. In order to employ AI in healthcare ethically, it must be ensured that the technology enhances human knowledge rather than replacing it, improving accessibility and quality of care. An ethics review board must conduct ongoing ethical evaluations in order to maintain high ethical standards during the system's development and implementation and to respond to changing problems. The project can accomplish its objectives while maintaining the highest levels of honesty and reliability by giving these ethical issues top priority.

5.4 Sustainability Plan

Ensuring the long-term sustainability of the deep learning-based system for "Tongue Disease Recognition" involves a comprehensive plan encompassing financial stability, technical maintenance, user engagement, environmental considerations, and ongoing research and development. Having financial stability can be accomplished by creating a subscription-based business model, obtaining initial funding from grants and investments, and efficiently controlling operating expenses. The system's technical integrity must be preserved via regular software updates, ongoing model retraining, and strong technical support. Enhancing user experience and system relevance can be achieved by involving users through patient education materials, training sessions for healthcare practitioners, and feedback systems. By using green data centers, maximizing energy use, and collaborating with e-waste recycling businesses, environmental sustainability is addressed. Pilot projects and partnerships with academic institutions will enable continuous research and development, which will maintain the system at the forefront of technical breakthroughs. Scalability and adaptability of the system will be further improved by extending its reach geographically and integrating it with other healthcare technology, all the while maintaining regulatory compliance. This all-encompassing strategy guarantees the tongue disease detection system's continued viability, influence, and alignment with environmental and technological sustainability objectives.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

The research project titled "Tongue Disease Recognition Using Deep Learning" aims to develop an innovative diagnostic tool leveraging advanced deep learning techniques to identify various tongue-related diseases. The importance of for early and precise diagnosis of oral diseases, which can point to more serious systemic health problems including diabetes, malnutrition, and cancer, is what spurred this research. The main goals are to build an integrated platform for real-time disease recognition that can be accessed through web and mobile applications, to design and implement multiple deep learning models such as CNNs, ResNet, InceptionV3, VGG16, and MobileNet, and to create a comprehensive dataset of high-resolution annotated tongue images. The process includes gathering and preparing data, training and assessing models, and deploying the system. Key findings show that tongue illnesses may be reliably identified by deep learning models, particularly ResNet and InceptionV3, with semantic segmentation techniques enhancing accuracy. The system's effects on healthcare include better accessibility, less work for medical staff, and early illness detection that allows for prompt treatment. In order to ensure compliance with data protection rules, ethical factors such as patient privacy, informed consent, algorithmic fairness, and transparency are addressed. Subsequent investigations will delve into increasingly sophisticated designs, broaden the dataset, and augment the system's efficiency via cooperative endeavors and ongoing modifications. This study offers better healthcare outcomes and accessibility globally by demonstrating the viability and advantages of employing AI for medical diagnosis.

6.2 Conclusions

The research project "Tongue Disease Recognition Using Deep Learning" effectively illustrates, in conclusion, the revolutionary potential of cutting-edge AI approaches in the early identification and diagnosis of tongue-related diseases. The project achieved excellent accuracy and diagnostic precision by creating a large dataset of high-resolution annotated tongue images and putting different deep learning models, such as CNNs, ResNet, InceptionV3, VGG16, and MobileNet, into practice. Better health outcomes and early disease identification result from the incorporation of these models into an approachable web and mobile platform, which increases healthcare accessible, especially for those living in distant or underserved locations. To guarantee adherence to laws like GDPR and HIPAA, ethical concerns including patient privacy, data security, informed permission, and algorithmic fairness were carefully taken into account. This project's success emphasizes how crucial it is to collaborate, stay ethically vigilant, and innovate continuously in order to advance AI-driven medical diagnoses. Expanding the dataset, investigating more complex structures, and incorporating more patient data will be the main areas of future research to guarantee the system's long-term effectiveness and widespread use in healthcare.

6.3 Implication for Further Study

The research project "Tongue Disease Recognition Using Deep Learning" has established a solid foundation for future advancements in AI-driven medical diagnostics, revealing several key implications for further study. Improving the dataset's variety and representativeness by adding more samples from various demographic groups and lighting situations will improve the models' generalizability and robustness. It may be possible to increase performance and efficiency by looking into cutting-edge deep learning architectures like Vision Transformers (ViTs), DenseNet, and EfficientNet. More thorough diagnostic capabilities will be possible with the integration of multimodal data, such as medical history and demographic data.

To guarantee accessibility in remote locations, future research should concentrate on creating real-time diagnostic capabilities and edge computing model optimization. Increasing the explainability and interpretability of models using methods such as saliency maps and attention mechanisms would increase patient and healthcare provider trust. Evaluating the system's usability and effect on patient outcomes will need carrying out comprehensive clinical trials and real-world validation studies. Sustaining public confidence will require addressing moral and legal issues, such as data privacy and informed permission. The system's accuracy and efficacy will be further improved by investigating customized medicine, which will modify diagnostic and treatment suggestions based on unique patient data. This would eventually improve healthcare results and boost accessibility.

REFERENCES

[1] S.-J. Lee, I.-J. Kwon, Y.-D. Son, J.-H. Kim, M.-S. Lee, D. Kwon, and H.-K. Kim, "Early detection of tongue cancer using a convolutional neural network and evaluation of the effectiveness of EfficientNet," 2022.

[2] A. A. Alanazi, M. M. Khayyat, M. M. Khayyat, B. M. E. Elnaim, and S. Abdel-Khalek, "Intelligent Deep Learning-Enabled Oral Squamous Cell Carcinoma Detection and Classification Using Biomedical Images," Computational Intelligence and Neuroscience, vol. 2022, Article ID 7643967, 2022. DOI: 10.1155/2022/7643967.

[3] N. D. Kamarudin, C. Y. Ooi, T. Kawanabe, H. Odaguchi, and F. Kobayashi, "A Fast SVM-Based Tongue's Colour Classification Aided by k-Means Clustering Identifiers and Colour Attributes as a Computer-Assisted Tool for Tongue Diagnosis," Journal of Healthcare Engineering, vol. 2017, Article ID 7460168, 2017. DOI: 10.1155/2017/7460168

[4] W. Tang, Y. Gao, L. Liu, T. Xia, L. He, S. Zhang, J. H. Guo, W. Li, and Q. Xu, "An Automatic Recognition of Tooth-Marked Tongue Based on Tongue Region Detection and Tongue Landmark Detection via Deep Learning," IEEE Access, vol. 5, pp. 14227-14237, 2017. DOI: 10.1109/ACCESS.2017.2732421.

[5] X. Wang, J. Liu, C. Wu, J. Liu, Q. Li, Y. Chen, X. Wang, X. Chen, X. Pang, B. Chang, J. Lin, S. Zhao, Z. Li, Q. Deng, Y. Lu, D. Zhao, and J. Chen, "Artificial intelligence in tongue diagnosis: Using deep convolutional neural network for recognizing unhealthy tongue with tooth-mark," Computational and Structural Biotechnology Journal, vol. 18, pp. 1511-1516, 2020.

[6] R. F. Mansour, M. M. Althobaiti, and A. A. Ashour, "Internet of Things and synergic deep learning based biomedical tongue color image analysis for disease diagnosis and classification," IEEE Access, vol. 9, pp. 97053-97067, 2021.

[7] J. Li, Q. Chen, X. Hu, P. Yuan, L. Cui, J. Tu, J. Cui, J. Huang, T. Jiang, X. Ma, X. Yao, C. Zhou, H. Lu, and J. Xu, "Establishment of noninvasive diabetes risk prediction model based on tongue features and machine learning techniques," International Journal of Medical Informatics, vol. 153, Article ID 104429, 2021.

[8] J. Li, Z. Zhang, X. Zhu, Y. Zhao, Y. Ma, J. Zang, B. Li, X. Cao, and C. Xue, "Automatic classification framework of tongue feature based on convolutional neural networks," Micromachines, vol. 13, no. 4, Article ID 450, 2022.

[9] Y. Tang, Y. Sun, J. Y. Chiang, and X. Li, "Research on multiple-instance learning for tongue coating classification," IEEE Access, vol. 5, pp. 1205-1213, 2017.

[10] J. Heo, J. H. Lim, H. R. Lee, J. Y. Jang, Y. S. Shin, D. Kim, J. Y. Lim, Y. M. Park, Y. W. Koh, S.-H. Ahn, E.-J. Chung, D. Y. Lee, J. Seok, and C.-H. Kim, "Deep learning model for tongue cancer diagnosis using endoscopic images," Scientific Reports, vol. 12, Article ID 10287, 2022.

[11] M. H. Tania, K. Lwin, and M. A. Hossain, "Advances in automated tongue diagnosis techniques," Information and Management Research, vol. 1, no. 1, pp. 25-35, 2018.

[12] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in Proc. Int. Conf. Learn. Represent. (ICLR), 2015.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 2818-2826. DOI: 10.1109/CVPR.2016.308.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770-778. DOI: 10.1109/CVPR.2016.90.

[15] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017.

TONGUE DISEASE RECOGNITION USING DEEP LEARNING

ORIGINALITY REPORT

15%	16%	7%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	8%
2	Submitted to Daffodil International University Student Paper	4%
3	Yonghui Tang, Yue Sun, John Y. Chiang, Xiaoqiang Li. "Research on Multiple-Instance Learning for Tongue Coating Classification", IEEE Access, 2021 Publication	1%
4	Jaesung Heo, June Hyuck Lim, Hye Ran Lee, Jeon Yeob Jang et al. "Deep learning model for tongue cancer diagnosis using endoscopic images", Scientific Reports, 2022 Publication	1%
5	towardsdatascience.com Internet Source	1%
6	Shareef, Bryar. "Deep Learning for Ultrasound-Based Breast Cancer Early Detection", University of Idaho, 2023 Publication	1%