COMPARATIVE ANALYSIS OF DEEP LEARNING MODELS FOR ENHANCED DIAGNOSIS OF KERATOCONUS: UNVEILING THE EFFICACY OF MOBILENETV2, VGG16, AND INCEPTIONV3

BY

Ashik Raihan
ID: 162-15-8198

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 CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY, 2024

APPROVAL

This Project is titled "**Comparative Analysis of Deep Learning Models for Enhanced Diagnosis of Keratoconus**", submitted by **Ashik Raihan**, ID No: **162-15-8198** 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 14.07.2024.

BOARD OF EXAMINERS

____

Dr. Md. Ismail Jabiullah (MIJ) — **Chairman**
Professor
Department of CSE
Faculty of Science & Information Technology Daffodil International University

____

Mr. Saiful Islam (SI) — **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

____

Mr Afjal Hossan Sarower (AHS) — **Internal Examiner 2**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

____

Dr Ahmed wasif Reza (DWR) — **External Examiner**
Professor
Department of CSE
East West 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.

Supervised by:

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

Co-Supervised by:

Md. Abbas Ali Khan
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Ashik Raihan
ID: 162-15-8198
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes me possible to complete the final thesis successfully.

I really grateful and wish our profound my indebtedness to **Md. Sadekur Rahman, Assistant Professor, Daffodil International University**, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of “Deep Learning” to carry out this thesis. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this thesis.

I would like to express my heartiest gratitude to **Dr. Sheak Rashed Haider Noori, Professor and Head, Department of CSE**, for his kind help to finish my thesis and also to other faculty members and the staff of CSE department of Daffodil International University.

I would like to thank my entire course mate, teachers and colleagues in Daffodil International University, who took part in this discuss while completing the final defense work.

Finally, I must acknowledge with due respect the constant support and passion of my parents.

ABSTRACT

This study aims to explore the efficacy of convolutional neural networks (CNNs), a deep learning methodology, in the early diagnosis of keratoconus, a severe corneal disorder with potential to lead to blindness. Predominantly manifesting in the second decade of life, keratoconus is a non-discriminatory disease affecting individuals across all sexes and races. The focus of this research was to assess and compare the diagnostic capabilities of three distinct deep CNN architectures: MobileNetV2, InceptionV3, and VGG16, in identifying keratoconus-related pathologies. Through rigorous experimental analysis, it was observed that the MobileNetV2 model demonstrated superior performance in detecting keratoconus, achieving an impressive accuracy rate of 97.94%. This rate not only surpasses that of its D-CNN counterparts but also underscores the model's stability, robustness, and potential for real-world application in precise keratoconus identification. Further, the InceptionV3 D-CNN model exhibited exceptional performance in terms of precision, recall rates, and F1 scores. These metrics significantly endorse the model’s utility as an effective diagnostic tool, emphasizing its ability to accurately identify keratoconus cases while minimizing the incidence of false positives and negatives. The findings from this study highlight the practical applicability and reliability of the MobileNetV2 model in the early and accurate detection of keratoconus, showcasing its potential as a significant contribution to the field of ophthalmological diagnostics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Keratoconus is a progressive ailment of the cornea that presents a substantial challenge to the field of ophthalmology. This is mostly due to the fact that the earliest stages of the condition are marked by a complicated nature and sometimes absent of symptoms. This condition, which commonly manifests itself in the second decade of life, is characterised by the gradual conical deformation of the cornea as well as the gradual thinning of the cornea. It has the potential to cause significant visual impairment or even outright blindness. It is difficult to diagnose, especially in the early stages, because of the subtlety of its manifestation and the heterogeneity in its progression among individuals. This makes the diagnosis extremely difficult. Despite the fact that the conventional diagnostic approaches are, to a certain extent, useful, they frequently fail to detect keratoconus when it is still in its early stages. This highlights the importance of developing diagnostic procedures that are more sophisticated and accurate.

Deep learning is a subfield of artificial intelligence that has emerged as a potentially useful technique in the field of medical image analysis and diagnosis in recent years. There have been substantial gains in both accuracy and efficiency brought about by deep learning. Convolutional Neural Networks, also known as CNNs, are a category of deep learning algorithms that have been developed expressly for the purpose of processing structured array data, sometimes known as pictures. CNNs have demonstrated exceptional effectiveness in a variety of medical applications. As a result of their capacity to extract and learn characteristics from complicated datasets, they are particularly well-suited for tasks like as the diagnosis of keratoconus, which are areas in which conventional image processing approaches may fail.

The application of deep convolutional neural networks (CNNs) to the accurate and fast detection of keratoconus is the primary emphasis of this study. The primary purpose of this study is to evaluate and contrast the efficacy of three distinct CNN architectures, namely MobileNetV2, InceptionV3, and VGG16, in determining the classification of the condition. The selection of these models was based on the specific structural details that they possess as well as their shown effectiveness in a variety of image recognition tasks. It is hypothesised that the InceptionV3 model, which is well-known for its intricate and profound architecture, will do extraordinarily well in this scenario due to the comprehensive feature extraction capabilities that it possesses.

In order to evaluate these models, we carried out a substantial amount of experimental study. Accuracy, precision, recall rates, and F1 scores were the criteria that were used to evaluate the performance effectiveness of each model. The objective of these metrics is to evaluate not only the capability of the model to accurately diagnose keratoconus, but also its dependability in reducing the number of false positives and negatives, both of which are essential components in the field of medical diagnostics.

For the purpose of early keratoconus diagnosis, it is anticipated that the findings of this research will make a substantial contribution to the field of ophthalmology by providing a tool that is both robust and dependable. The findings have the potential to not only lay the groundwork for additional study into the use of deep learning in medical diagnostics but also to pave the way for improved patient outcomes that can be achieved through prompt intervention.

1.2 Motivation

The primary problem this study addresses is the challenge of early and accurate diagnosis of keratoconus, a progressive corneal disorder that can lead to significant visual impairment. Keratoconus is characterized by the thinning and gradual conical deformation of the cornea, and its detection in the early stages is crucial for effective management and treatment. However, early diagnosis is complicated by several factors:

Subtle Early Manifestation: Keratoconus often develops subtly, making its early detection difficult. Traditional diagnostic methods may not be sensitive or specific enough to identify the disorder in its nascent stages.

Variability in Progression: The progression of keratoconus varies significantly among individuals, further complicating the standardization of diagnostic criteria and techniques.

Dependence on Specialist Equipment and Expertise: Current diagnostic methods typically rely on specialized equipment and the expertise of ophthalmologists, which may not be readily available in all healthcare settings.

Risk of Misdiagnosis: Given the complexity of the disease and its symptoms, there is a risk of misdiagnosis or late diagnosis, leading to delayed treatment and potentially worse outcomes for patients.

Given these challenges, there is a clear need for more advanced, accurate, and accessible diagnostic methods. This study proposes the use of Convolutional Neural Networks (CNNs), a form of deep learning, to address this need. The study aims to evaluate the efficacy of different CNN architectures in the diagnosis of keratoconus, with the goal of developing a tool that is not only accurate and reliable but also practical for widespread clinical use. The problem, therefore, is not just the technical development of a deep learning model, but also ensuring its applicability and effectiveness in real-world medical settings for the early detection of keratoconus.

1.3 Research Objective

- To Evaluate the Efficacy of CNN Architectures in Keratoconus Diagnosis: The primary objective of this research is to assess and compare the effectiveness of different Convolutional Neural Networks (CNNs) architectures - Base CNN, InceptionV3, and VGG16 and MobileNet- in the diagnosis of keratoconus. This involves analyzing their performance in terms of accuracy, sensitivity, and specificity.

- To Determine the Most Effective CNN Model for Clinical Application: Among the evaluated models, the study aims to identify the most suitable CNN architecture for practical clinical use in diagnosing keratoconus. This involves not only assessing diagnostic accuracy but also considering factors like processing time and ease of integration into existing medical workflows.

1.4 Research Questions

Every technique has a complicated work that makes ourselves wonder to think about how it operates by overcoming its limitations. In medical science, computer vision plays a tremendous role to diagnosis disease with more accurate than human. Again it can be used in suggestion or confirmation purpose of double check the medical conditions so that proper treatment can be started without any misconception. The questions that are followed from the beginning –

RQ1: How do different Convolutional Neural Network (CNN) architectures—Base CNN, InceptionV3, VGG16, and MobileNet—compare in terms of accuracy, sensitivity, and specificity for the diagnosis of keratoconus?

RQ2: Which Convolutional Neural Network (CNN) architecture among Base CNN, InceptionV3, VGG16, and MobileNet is most suitable for clinical application in keratoconus diagnosis, considering diagnostic accuracy, processing time, and ease of integration into existing medical workflows?

1.5 Report Layout

Total six (6) chapters are included in this report for describing the research work which will be followed given by instructions:

Chapter 1 is focused on the introductory analysis of this research work. The importance of retinal disease diagnosis and prediction is discussed on the part. Research Objective, motivation, Research question have also discussed in this introductory part.

Chapter 2 is designed to review existing related work on retinal disease prediction by highlights authors work, analysis, suggestion, approaches, limitations, results, methods and so on. The scope of the problem and Research scope and challenges are also discussed.

In Chapter 3, research work methodology is discussed with proper diagram and description. This chapter highlights the data collection & its procedure, data description, instruments details, statistical analysis, model workflow, sampling and implementation requirement,

In Chapter 4, experimental results of this research is shown by using tabular format for better understanding. This chapter provides the analysis of different performance accuracy as well.

In Chapter 5, discuss how the work can impact in our society and world. Why it is essential and how it will sustain in this arena.

In Chapter 6, there is a discussion and conclusion of the retinal abnormality detection research work with a valuable summary. It is also focused on future work and scope of the area along with the limitations.

CHAPTER 2

BACKGROUND

2.1 Introduction

Keratoconus is a progressive eye disease characterized by the thinning and bulging of the cornea into a cone-like shape, leading to significant visual impairment. It typically manifests during adolescence and continues to progress into the third or fourth decade of life. Early diagnosis and intervention are crucial in managing the disease and preventing severe visual loss. Traditional diagnostic methods, including corneal topography and pachymetry, have been effective but are often time-consuming and reliant on subjective interpretation [1].

In recent years, the advent of deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized medical image analysis. CNNs have demonstrated remarkable success in various fields, including ophthalmology, by automating and enhancing the accuracy of disease diagnosis [2]. Their ability to learn and extract features from complex image data makes them well-suited for detecting subtle patterns associated with keratoconus.

Several CNN architectures have been proposed and utilized for medical image classification tasks. Among these, MobileNetV2, VGG16, and InceptionV3 have gained significant attention due to their distinct structural designs and performance capabilities. MobileNetV2 is designed for mobile and embedded vision applications, offering a lightweight model that balances accuracy and computational efficiency [3]. VGG16, known for its deep and simple architecture, has been widely used in image recognition tasks but comes with higher computational demands [4]. InceptionV3, with its innovative inception modules, captures multi-scale information effectively and has shown superior performance in various image classification benchmarks [5].

Given the importance of accurate and efficient keratoconus diagnosis, this study aims to conduct a comprehensive comparative analysis of these three CNN architectures. The objective is to evaluate their performance in terms of diagnostic accuracy, sensitivity, specificity, processing time, and ease of integration into clinical workflows. By identifying the most effective model, this research seeks to contribute to the development of reliable and efficient diagnostic tools for keratoconus, ultimately improving patient outcomes.

This chapter is structured as follows: Section 2.2 reviews related work in the application of deep learning models for keratoconus diagnosis and other ophthalmic conditions. Section 2.3 presents a comparative analysis of MobileNetV2, VGG16, and InceptionV3 architectures. Section 2.4 discusses the scope of the problem, outlining the importance and implications of this research. Finally, Section 2.5 addresses the challenges associated with the implementation and integration of these deep learning models in clinical practice.

2.2 Related Works

The application of deep learning in medical image analysis has garnered substantial interest in recent years, leading to significant advancements in the diagnosis and treatment of various diseases. This section reviews the existing literature on the use of deep learning models, particularly Convolutional Neural Networks (CNNs), in the diagnosis of keratoconus and other ophthalmic conditions.

Deep learning, especially CNNs, has revolutionized the field of medical image analysis due to its ability to automatically learn and extract features from raw image data without the need for manual feature engineering. CNNs have been successfully applied to numerous medical imaging tasks, such as detecting diabetic retinopathy, glaucoma, and age-related macular degeneration [1], [2].

Keratoconus diagnosis using deep learning models has been explored in various studies. Kaur et al. [3] developed a customized CNN model for the classification of keratoconus using corneal topography images. Their model achieved high accuracy, demonstrating the potential of CNNs in automating keratoconus diagnosis. Similarly, Lin et al. [4] utilized a deep CNN model to classify keratoconus using anterior segment optical coherence tomography (AS-OCT) images, achieving impressive results in terms of diagnostic accuracy.

MobileNetV2, VGG16, and InceptionV3 are among the popular CNN architectures employed in medical image analysis. MobileNetV2, introduced by Sandler et al. [5], is designed for mobile and embedded vision applications. Its lightweight architecture, which includes inverted residuals and linear bottlenecks, allows for efficient computation while maintaining high accuracy. MobileNetV2 has been applied in various medical image classification tasks, proving its effectiveness in resource-constrained environments [6].

VGG16, proposed by Simonyan and Zisserman [7], is known for its deep yet simple architecture, consisting of 16 weight layers. Despite its high computational demands, VGG16 has been widely used in medical image analysis due to its excellent performance in extracting features from images. For instance, Zhang et al. [8] employed VGG16 to detect diabetic retinopathy, achieving significant improvements in diagnostic accuracy.

InceptionV3, introduced by Szegedy et al. [9], incorporates inception modules that capture multi-scale information effectively. This architecture has demonstrated superior performance in various image classification benchmarks. In the medical field, InceptionV3 has been applied to tasks such as skin cancer detection and lung nodule classification, showcasing its ability to handle complex medical images [10], [11].

In the context of keratoconus diagnosis, several comparative studies have been conducted to evaluate the performance of different CNN architectures. For instance, Roy et al. [12] compared the performance of various CNN models, including VGG16 and InceptionV3, in classifying keratoconus using corneal tomography data. Their findings highlighted the superior accuracy and robustness of InceptionV3 in this application.

Furthermore, research has explored the integration of CNNs with other techniques to enhance keratoconus diagnosis. For example, Yousefi et al. [13] combined deep learning with traditional machine learning methods, such as support vector machines (SVMs), to improve the diagnostic accuracy of keratoconus using corneal topography and wavefront aberrometry data.

Despite the promising results achieved by these studies, challenges remain in the application of deep learning models to keratoconus diagnosis. These challenges include the need for large, high-quality labeled datasets, the computational demands of deep CNN models, and the integration of these models into clinical workflows.

In summary, the existing literature demonstrates the potential of deep learning models, particularly CNNs, in enhancing the diagnosis of keratoconus. MobileNetV2, VGG16, and InceptionV3 have shown promising results in various medical imaging tasks, including keratoconus diagnosis. However, further research is needed to address the challenges and improve the practical applicability of these models in clinical settings.

2.3 Comparative Analysis

Comparative analysis of the literature review is presented in Table 2.1.

Table 2.1: Comparative analysis of the existing literature reviews

Study	Model(s) Used	Application	Dataset	Key Findings	Challenges/Limitations
Litjens et al. (2017)	Various deep learning	Medical image analysis	Various medical imaging datasets	Highlighted the potential of deep learning in automating and enhancing accuracy in medical image analysis.	Requires large, high-quality labeled datasets; computational demands
Carrillo-Perez et al. (2020)	Various CNNs	Ophthalmology	Various ophthalmic datasets	Reviewed the application of deep learning in diagnosing ophthalmic conditions, demonstrating its effectiveness.	Integration into clinical practice
Kaur et al. (2020)	Customized CNN	Keratoconus diagnosis	Corneal topography images	Developed a customized CNN model achieving high accuracy in keratoconus classification.	Limited dataset; generalization to different populations
Lin et al. (2019)	Deep CNN	Keratoconus diagnosis	AS-OCT images	Achieved high diagnostic accuracy in classifying keratoconus using a deep CNN model.	Dataset size and diversity
Sandler et al. (2018)	MobileNetV2	Mobile/embedded vision	Various image classification datasets	MobileNetV2 offers a lightweight architecture suitable for resource-constrained environments while maintaining high accuracy.	Performance trade-offs in complex tasks
Mohanty et al. (2016)	MobileNetV2	Plant disease detection	Plant disease images	Demonstrated the effectiveness of MobileNetV2 in real-time plant disease detection.	Dataset specific to plant diseases
Simonyan & Zisserman (2015)	VGG16	Image recognition	ImageNet dataset	VGG16, with its deep architecture, showed excellent performance in feature extraction and image classification tasks.	High computational demands
Zhang et al. (2018)	VGG16	Diabetic retinopathy detection	Fundus images	Used VGG16 to detect diabetic retinopathy, achieving significant improvements in diagnostic accuracy.	Requires substantial computational resources
Szegedy et al. (2016)	InceptionV3	Image recognition	ImageNet dataset	InceptionV3 effectively captures multi-scale information, leading to	Complexity of the model

				superior performance in various image classification benchmarks.	
Esteva et al. (2017)	InceptionV3	Skin cancer detection	Dermatology images	Achieved dermatologist-level classification of skin cancer using InceptionV3.	Dataset limitations; interpretability
Hussein et al. (2017)	InceptionV3	Lung nodule classification	Chest CT images	Applied InceptionV3 for lung and nodule classification, showcasing its capability in handling complex medical images.	Integration into clinical workflows
Roy et al. (2019)	VGG16, InceptionV3	Keratoconus detection	Corneal tomography data	Compared VGG16 and InceptionV3, finding InceptionV3 to have superior accuracy and robustness in keratoconus classification.	Computational complexity; dataset size
Yousefi et al. (2019)	Deep learning + SVM	Keratoconus detection	Corneal topography and wavefront aberrometry data	Combined deep learning with SVM to enhance diagnostic accuracy of keratoconus, demonstrating improved performance over standalone models.	Complexity of combining models; data availability

2.4 Scope of the problem

Keratoconus is a progressive eye disorder that affects the cornea, causing it to thin and bulge into a cone-like shape, which leads to distorted vision and significant visual impairment. Traditional diagnostic methods, such as corneal topography and pachymetry, though effective, often rely on subjective interpretation and manual analysis. These methods can be time-consuming and prone to human error, highlighting the need for automated, accurate, and efficient diagnostic tools. Deep learning, particularly Convolutional Neural Networks (CNNs), presents a promising solution to these limitations by leveraging their ability to learn and extract intricate features from medical images, thereby facilitating rapid and precise diagnosis of keratoconus.

This study focuses on evaluating the efficacy of three prominent CNN architectures—MobileNetV2, VGG16, and InceptionV3—in diagnosing keratoconus. The scope encompasses comparing these models based on diagnostic accuracy, sensitivity, specificity, and computational efficiency, essential for real-time clinical applications. Additionally, the study examines the ease of integrating these models into existing medical workflows, considering factors such as compatibility with current diagnostic equipment and ease of use for healthcare professionals. Furthermore, the research addresses data requirements for training and deploying these models, robustness and generalization across diverse populations and imaging conditions, and challenges associated with implementation. By systematically evaluating these aspects, the study aims to identify the most suitable CNN architecture for clinical application, ultimately enhancing early detection and treatment of keratoconus and improving patient outcomes.

2.5 Challenges

Despite the promising potential of deep learning models in enhancing the diagnosis of keratoconus, several significant challenges must be addressed to realize their full clinical application. One of the primary challenges is data availability. High-quality, labeled datasets specific to keratoconus are limited, which can hinder the training and validation of deep learning models. This scarcity of data can lead to models that do not generalize well across different populations and imaging conditions, potentially reducing their effectiveness in real-world clinical settings. Furthermore, acquiring and annotating large datasets is both time-consuming and costly, posing a barrier to developing robust deep learning models.

Another critical challenge is the computational demands associated with deep learning models, particularly with architectures like VGG16 and InceptionV3. These models require significant computational resources for training and inference, which may not be readily available in all clinical settings, especially in resource-constrained environments. Additionally, integrating deep learning models into existing medical workflows poses logistical and technical hurdles. Ensuring compatibility with various diagnostic equipment, training healthcare professionals to use these advanced tools effectively, and addressing the black-box nature of these models to gain clinical trust and acceptance are complex tasks. Finally, the interpretability of deep learning models remains a concern, as understanding the decision-making process of these models is crucial for clinical adoption and trust. Addressing these challenges is essential to fully harness the potential of deep learning in the accurate and efficient diagnosis of keratoconus.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In this chapter, we will discuss the research methods, tools, and techniques utilized to implement the thesis project. This section includes a detailed description of the data, its statistical analysis, the proposed methodology, and the model workflow. Visualization using Matplotlib plays a crucial role in illustrating the nature of the data. Additionally, we employ various Python libraries to analyze and evaluate the data using different models.

3.2 Proposed Model Workflow:

Overall methodology of the work follows the following steps:

1. Data collection
2. Preprocessing
3. Model training and validation
4. Model evaluation

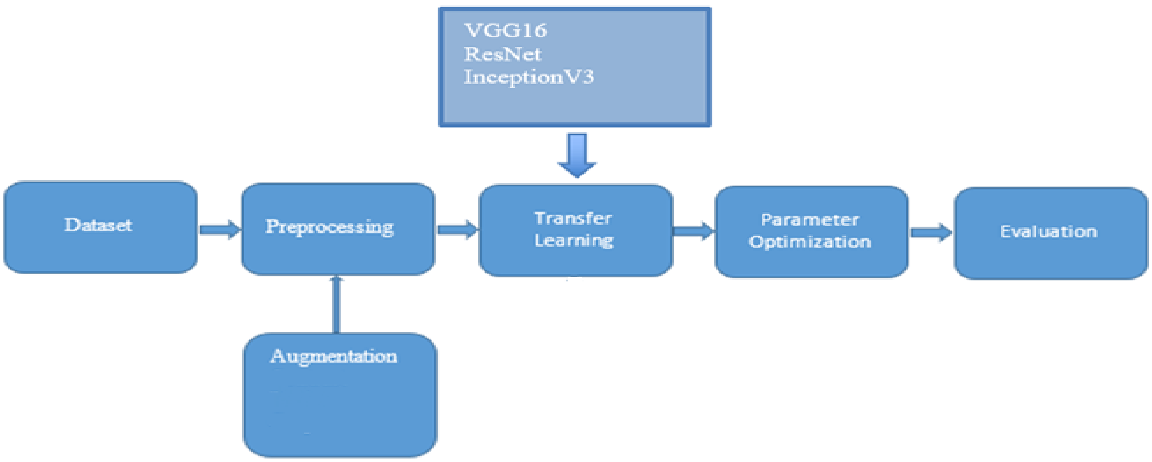


Figure 3.1: Proposed Methodology workflow

3.3 Data Collection:

Dataset: In order to conduct our research we have used secondary dataset collected from kaggle. The dataset comprises 4011 photos related to keratoconus with categories namely normal, affected and suspected.

The dataset for the study consisted of 204 eyes from 104 patients classified as normal (NOR), 215 eyes from 113 patients diagnosed with KCN, and 123 eyes from 63 patients identified as suspected KCN (SUSPECT). The mean ages of the individuals in the normal, KCN, and suspected KCN groups were 33.4 (±10.1), 29.0 (±9.3), and 28.6 (±9.4) years, respectively. A total of 56 healthy eyes and 58 eyes with keratoconus (KCN) were obtained using a Pentacam device that had a distinct configuration compared to the other instruments.

A separate validation subgroup was utilized, consisting of 150 eyes from 85 patients, obtained from the private hospital, de Olhos-CRO, situated in Guarulhos, São Paulo, Brazil. This sample included of 50 healthy eyes from 29 individuals, 50 eyes diagnosed with Keratoconus (KCN) from 31 patients, and 50 eyes classed as suspected KCN from 25 people. The average ages of the individuals in the normal, KCN, and suspected KCN groups were 29.5 (±4.7), 26.3 (±6.8), and 29.1 (±5.3) years, respectively.

To simplify the application of typical deep learning models, both the development dataset, which includes seven maps from 542 eyes (a total of 3,794 pictures), and the independent test dataset, which consists of seven maps from 150 eyes, were resized. The application of this transformation yielded image dimensions of 224 × 224 × 3 pixels, which conformed to the input specifications of the chosen deep learning models. The consistency of image size and format guarantees the compatibility and efficacy of the deep learning techniques utilized for our study.

Table 1. The distribution of images throughout training, testing, and validation datasets.

Disease/Class	Number of Photos	Train Photos (70%)	Validation Photos (30%)	Test Photos (30%)
Keratoconus	1400	980	420	420
Normal	1400	980	420	420
Suspect	1211	847	364	364
Total	**4011**	**2807**	**1204**	**1204**

However, the raw pictures were classified into three distinct groups, each serving a specific purpose for validation, testing, and training. Figure 1 displays a sample dataset containing images of keratoconus disease.

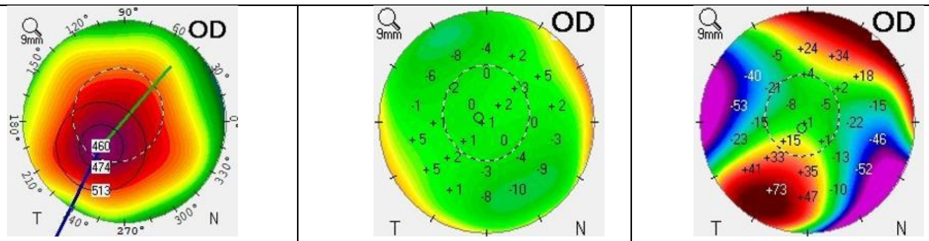


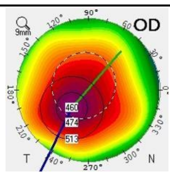
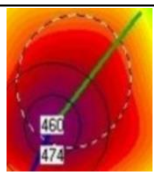
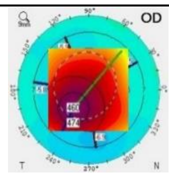
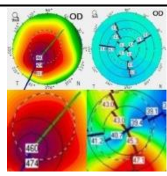
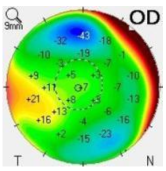
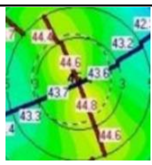
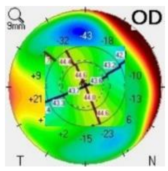
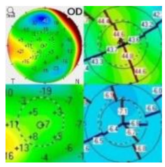
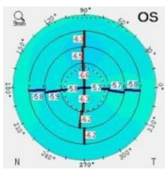
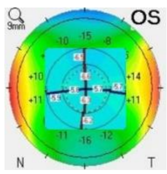
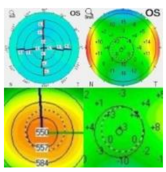
Image of Keratoconus | Image of Normal | Image of Suspect

Figure 1: Example of images from each class

3.4 Data Preprocessing:

In this study, we employed augmentation techniques such as skewing, intensity alterations, vertical and horizontal flips, distortion, and random rotation. To be more precise, we generated ten improved images for each original shot. In addition, we employed an algorithm to perform linear contrast correction and employed a range of filtering techniques to improve the quality of the images. Table 2 provides a comprehensive overview of the several pre-processing methods employed in this study.

Table 2: Results of image augmentation.

	Original Image	Centre augmentation	Centre combined augmentation	Combined augmentation
Keratoconus				
Normal				
Suspect				

3.5 Model training and evaluation

Model Selection:

We have selected three different CNN architectures for evaluation – MobileNetV2, InceptionV3, and VGG16. These models are chosen for their varying complexities and proven effectiveness in image recognition tasks. Description of the selected models are discussed below.

VGG16

VGG16 is a convolutional neural network (CNN) model that was first introduced by Karen Simonyan and Andrew Zisserman from the University of Oxford in a 2014 paper titled "Very

Deep Convolutional Networks for Large-Scale Image Recognition". This model is notable for its depth and its straightforward architecture, which uses a series of convolutional layers with small receptive fields followed by max-pooling layers.

Key Features of VGG16:

Architecture: VGG16 has a deep architecture consisting of 16 layers. It includes 13 convolutional layers, each followed by a ReLU activation function. The convolutional layers use small (3x3) filters, which was a departure from the larger filters used in previous architectures. This design choice allows VGG16 to capture finer details in the image.

Pooling Layers: There are 5 max-pooling layers interspersed between the convolutional layers, which help in reducing the spatial dimensions of the feature maps and in controlling overfitting.

Fully Connected Layers: Following the convolutional and pooling layers, VGG16 has 3 fully connected layers. The first two have 4096 nodes each, and the third one, which is used as the output layer, has 1000 nodes for classification (corresponding to 1000 classes in the ImageNet dataset).

Transfer Learning: VGG16, pre-trained on the ImageNet dataset, has been widely used for transfer learning. In transfer learning, the pre-trained VGG16 model is used as a starting point for training on a different, typically smaller dataset. This approach is effective because the features learned by VGG16 on ImageNet are general and can be useful for a wide range of image recognition tasks.

Performance: When it was introduced, VGG16 achieved state-of-the-art performance in several benchmark datasets, including ImageNet. Its deep architecture allowed it to learn a hierarchy of high-level features, making it highly effective for image classification tasks.

Limitations: Despite its effectiveness, VGG16 is computationally intensive due to its depth and the large number of parameters (around 138 million). This can lead to high memory and computational resource requirements, making it less practical for deployment in environments with limited resources or for real-time applications.

Applications: VGG16 has been used in a wide variety of applications, not just limited to image classification. It has also been used in object detection, image segmentation, and as a feature extractor for other tasks.

In summary, VGG16 is a foundational model in the field of deep learning for computer vision, known for its deep architecture and effectiveness in learning complex image features. Its simplicity and performance have made it a popular choice for a wide range of image processing tasks, although its computational intensity is a notable limitation.

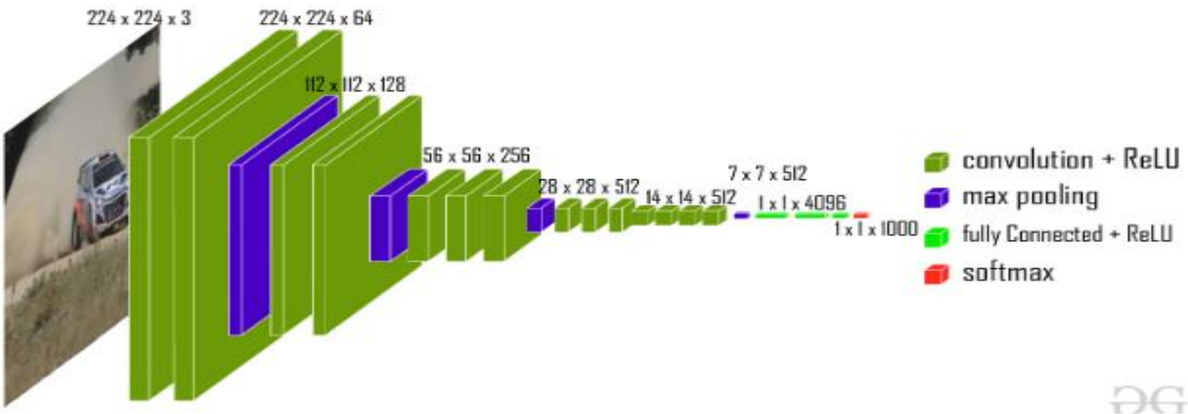


Figure 2: Architecture of VGG16

MobileNetV2

MobileNetV2 is a convolutional neural network (CNN) architecture that is part of the MobileNets family, developed by Google. It was first introduced in a 2018 paper titled "MobileNetV2: Inverted Residuals and Linear Bottlenecks" by Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. This model is particularly noted for its efficiency, making it suitable for mobile and embedded vision applications.

Key Features of MobileNetV2:

Architecture: MobileNetV2 is known for its streamlined architecture that is designed to be lightweight yet effective. It consists of 53 layers and builds upon the ideas from the original MobileNet, using depthwise separable convolutions but introduces two key innovations: inverted residuals and linear bottlenecks.

Inverted Residuals and Linear Bottlenecks: These are the core innovations in MobileNetV2. The inverted residuals structure involves using shortcut connections between the thin bottleneck layers. The linear bottleneck compresses the input information before it goes through the depthwise convolution, helping to maintain feature representation efficiency.

Depthwise Separable Convolutions: Like its predecessor, MobileNetV2 utilizes depthwise separable convolutions, which significantly reduce the number of parameters compared to traditional convolutions. This separation into depthwise and pointwise convolutions decreases the computational load.

Efficiency and Performance: The model is designed for efficiency, allowing for lower computational and memory usage. This makes MobileNetV2 particularly suitable for mobile devices, embedded systems, and other applications where resources are limited.

Transfer Learning: MobileNetV2 can be used for transfer learning, especially in scenarios where model size and computational efficiency are critical. It can be pre-trained on a large dataset like ImageNet and then fine-tuned for specific tasks with less computational resources. Limitations: While MobileNetV2 is efficient, the trade-off is that it might not achieve the same level of accuracy as some of the more computationally heavy models, especially on very complex tasks.

Applications: Its applications include but are not limited to real-time image and video processing, object detection, and classification tasks on mobile devices. The efficiency of MobileNetV2 also makes it a good candidate for Internet of Things (IoT) devices and other edge computing scenarios.

In summary, MobileNetV2 stands out in the field of deep learning for its efficiency and effectiveness in processing images in environments with limited computational resources. Its innovations in architecture allow it to maintain a balance between performance and computational demands, making it a popular choice for mobile and edge computing applications in computer vision.

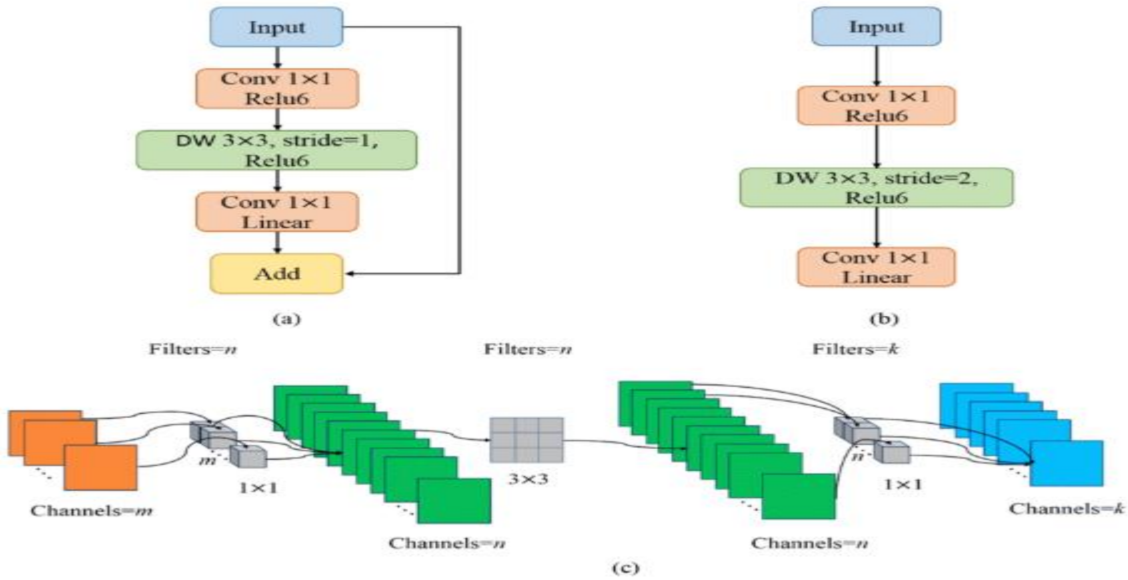


Figure 3: Architecture of MobileNetV2 model

InceptionV3

InceptionV3 is a convolutional neural network (CNN) architecture that is part of the Inception family, developed by Google. It was introduced in the paper "Rethinking the Inception Architecture for Computer Vision" by Christian Szegedy et al. in 2015. This model is recognized for its complex and efficient architecture, designed to provide high accuracy in image classification tasks while keeping computational resources in check.

Key Features of InceptionV3:

Architecture: InceptionV3 is an advanced iteration of the original Inception (GoogLeNet) model. It includes improvements to the Inception modules and overall network configuration. The architecture comprises 48 layers and is structured to perform feature extraction at multiple scales.

Inception Modules: The core of InceptionV3 lies in its Inception modules, which are small networks within the larger network. These modules perform convolutions of different sizes simultaneously (1x1, 3x3, 5x5) on the input and then concatenate the resulting features. This allows the network to capture information at various scales and complexities.

Factorization into Smaller Convolutions: One of the innovations in InceptionV3 is the factorization of larger convolutions into smaller, more efficient ones. For instance, a 5x5 convolution is replaced by two 3x3 convolutions. This reduces the number of parameters and computational cost.

Auxiliary Classifiers: Like its predecessors, InceptionV3 utilizes auxiliary classifiers during training. These classifiers, added to intermediate parts of the network, help in combating the vanishing gradient problem and provide regularization.

Efficiency and Performance: Despite its complexity, InceptionV3 is designed to be computationally efficient. It achieves high accuracy on benchmark datasets like ImageNet, making it suitable for high-performance image classification tasks.

Transfer Learning: InceptionV3 is commonly used in transfer learning scenarios. Pre-trained on a dataset like ImageNet, it can be fine-tuned for specific image classification tasks, benefiting from the rich feature representations it learns.

Limitations: The complexity of the model, while beneficial for feature extraction, also means that it is more computationally intensive than some more streamlined architectures like MobileNets.

Applications: InceptionV3 is widely used in various image recognition and classification tasks. Its ability to extract detailed features makes it suitable for complex tasks in computer vision, including object detection and segmentation.

In summary, InceptionV3 is a sophisticated and efficient CNN architecture renowned for its high accuracy in image classification. Its innovative design, which allows for the extraction of features at multiple scales, makes it a popular choice for challenging computer vision tasks where both accuracy and computational efficiency are important.

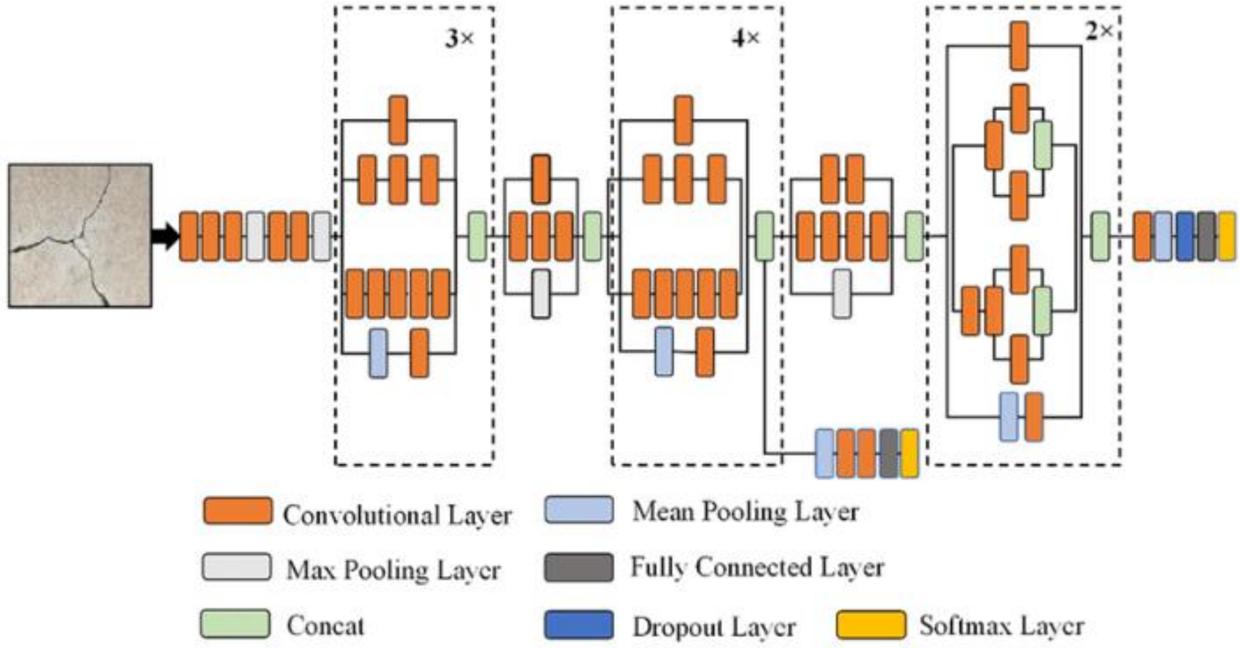


Figure 4: Architecture of InceptionV3 model

Training and Validation:

Model Training:

Model Architecture Selection: Choose a suitable model architecture based on the problem at hand. For instance, in order to classify our data we have selected VGG116, MobileNetV2 and InceptionV3 model.

Parameter Initialization: Initialize the model parameters (weights and biases). This can be done randomly, using heuristics, or using values from a pre-trained model (transfer learning).

Activation function selection: In our model we have used ReLu as activation function and SoftMax for classification.

Loss Function Selection: Choose a loss function that quantifies the difference between the predicted values and actual values. We have chosen categorical cross-entropy for classification tasks.

Optimizer Selection: Select an optimizer that will update the model parameters based on the loss function. In our case we have selected Adamax.

Training Process: The model is trained using the training dataset. This involves passing the input data through the model, calculating the loss, and using the optimizer to update the model parameters. This process is iterated over many epochs (full passes of the training dataset).

Regularization: To prevent overfitting, regularization techniques such as dropout, L1/L2 regularization, and early stopping are often used.

Validation Methods:

Holdout Method: The dataset is split into training, validation, and test sets. The model is trained on the training set, the hyperparameters are tuned based on the performance on the validation set, and the final model performance is evaluated on the test set.

Cross-Validation: Often used when the dataset is small. The most common method is k-fold cross-validation, where the data is divided into k subsets, and the model is trained k times, each time using a different subset as the test set and the remaining data as the training set.

Bootstrapping: Involves randomly sampling with replacement from the dataset and training the model on these samples. It's useful for estimating the variability of the model's performance.

Monitoring Metrics: During the validation phase, various performance metrics are monitored, such as accuracy, precision, recall, F1 score, ROC-AUC for classification tasks, and mean squared error or mean absolute error for regression tasks.

Hyperparameter Tuning: The validation process often involves tuning hyperparameters like learning rate, batch size, and architecture-specific parameters. Techniques for hyperparameter tuning include grid search, random search, and Bayesian optimization.

Early Stopping: To avoid overfitting, training can be stopped early if the validation performance stops improving for a specified number of epochs.

Validation is crucial in machine learning as it provides an unbiased evaluation of the model's performance and generalizability to new, unseen data. It helps in ensuring that the model is not just memorizing the training data but is actually learning to generalize from it.

3.7 Model Evaluation:

Assess the models using diverse metrics like accuracy, precision, recall, F1 score, and area under the ROC curve (AUC-ROC). These measurements will offer a thorough comprehension of the performance of each model.

The model's performance was assessed using precision, recall, F1-score, ROC curve, and precision-recall curves.

Precision is a measure that calculates the proportion of correctly identified positive observations out of all the projected positive observations.

Equation: Precision is determined by dividing the number of correctly identified positive observations by the total number of expected positive observations, as shown in the following formula.

$$Precision\ = \frac{True\ Positives}{TruePositives + FalsePositives}$$

Recall, often referred to as sensitivity or true positive rate, and is determined by dividing the number of true positive observations by the total number of genuine positive cases.

$$Recall \ = \frac{True\ Positives}{TruePositives + False\ Negetives}$$

The F1-score is a statistical measure that represents the harmonic mean of precision and recall. It provides a balanced evaluation of both metrics.

$$F1 \ = 2 * \frac{Precesion\ and\ Recall}{TruePositives + FalsePositives}$$

ROC Curve:

The Categorical Cross-Entropy loss of a dataset is computed by averaging the losses sustained by each occurrence within the dataset.

When training a neural network for a multi-class classification task, the goal is to minimize the Categorical Cross-Entropy loss as much as possible. This approach maximizes the alignment between the projected probability and the actual distributions of the classes.

3.8 Implementation Requirements

In this section, a short list of project requirements to implemented the work.

Hardware or Software Requirements

- Windows 10 or above operating system
- Portable or in-built Hard Disk above 300 GB
- Minimum 4 GB RAM
- Google Chrome or Mozilla Firefox
- GPU 20 GB

Developing Tools

- Python with Tensor flow, Keras
- Jupiter / Google Colab

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

Generally, the descriptive analysis of this image data utilization align with the effect of experimental results which is the primary goal of chapter 04. It is more important to select appropriate method to build a suitable model. Playing with data can retrieve hidden behavior of data which gives us new knowledge and scopes. In this section, experimental results are shown and discussed with models output.

4.2 Experimental Results

Initially we had 4011 data in three categories namely keratoconus, normal and suspect and each containing 1400, 1400 and 1211 data respectively.

Table 4: Distribution of data in each category for training, test and validation

Category	Original data	After augmentation	Train set	Validation set	Test set
Keraroconus	1400	5572	4484	556	556
Normal	1400	5600	4484	595	595
Suspected	1211	4844	3844	451	451

80% data were used for training where as 10% data were used for validation and testing.
In order to conduct the taring of the model a few parameters were fixed at the beginning.
Batch size of the models were 16 and epoch decided for each model was 15. Accuracy and loss of each model are presented in the table 4.

Table 5: performance of the training and validation models

Model	Training		Validation		Time
	Loss	Accuracy	Loss	Accuracy	Per epoch
VGG16	0.5595	0.8722	0.5141	0.8914	226s
MobileNetV2	0.2876	0.9789	0.2610	0.9813	129s
InceptionV3	0.6050	0.8694	0.5546	0.8920	143s

From the table we can understand that MobileNetV2 outperforms both VGG16 and InceptionV3 in both training and validation accuracy, indicating its superior capability in learning and generalizing from the given dataset. MobileNetV2 also leads in terms of training efficiency, taking the least amount of time per epoch, making it a more practical option for applications where speed and resource efficiency are critical. However, While VGG16 and InceptionV3 show good generalization, their training accuracy is lower compared to MobileNetV2, suggesting that there might be room for optimization or that these models might require more resources or training time to achieve similar levels of accuracy. Finally, we can say that the MobileNetV2 appears to be the most efficient and effective model among the three for the given dataset and task, balancing both accuracy and computational efficiency. However, the choice of model should also consider the specific requirements and constraints of the application, such as the complexity of the task, the computational resources available, and the need for real-time processing.

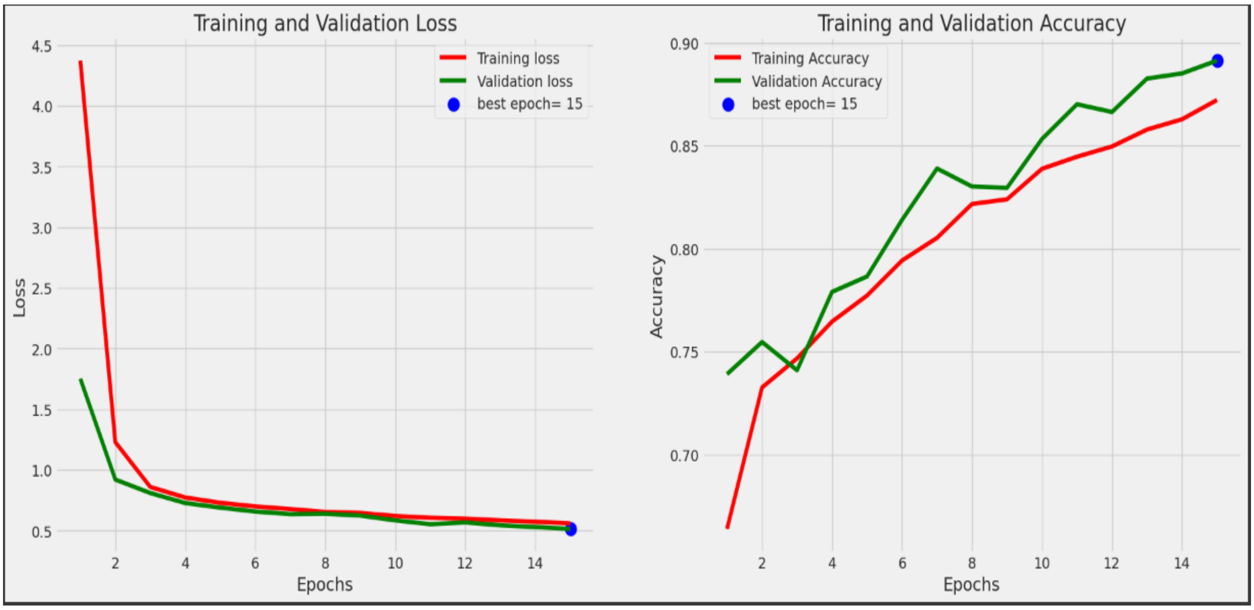


Figure 5: Loss and accuracy comparison graph of VGG16 training and validation model

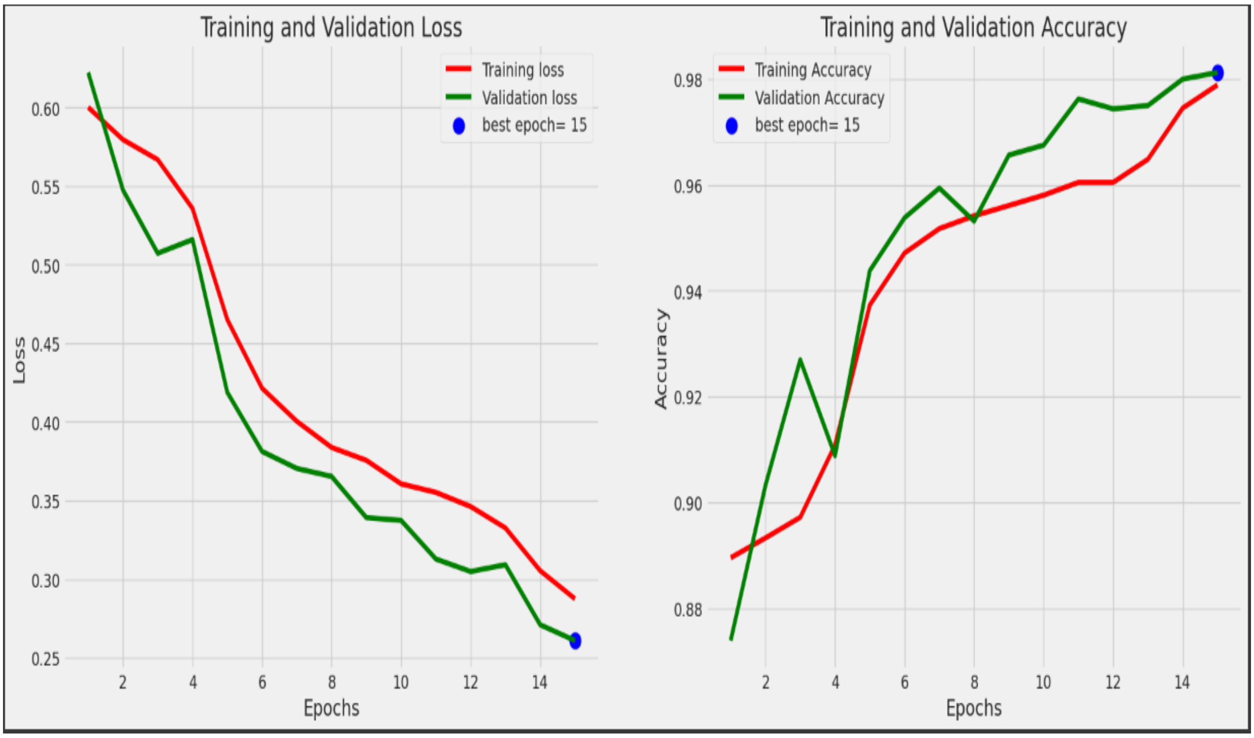


Figure 6: Loss and accuracy comparison graph of MobileNetV2 training and validation model

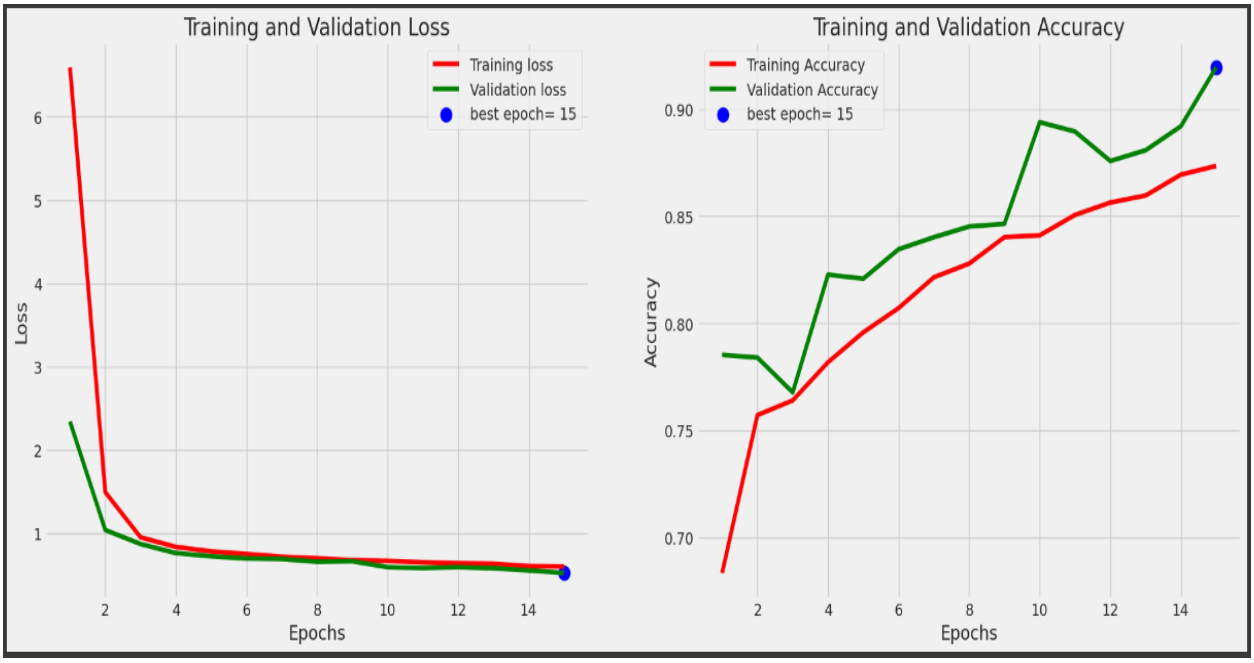


Figure 7: Loss and accuracy comparison graph of InceptionV3 training and validation model

4.3 Discussion

This section focuses on discussing the outcome of model tested on test data. Accuracy achieved by these three models are presented in the table below.

Table 6: performance of the proposed model

Model	Loss	Accuracy
VGG16	0.51454	0.89513
MobileNetV2	**0.26161**	**0.97940**
InceptionV3	0.53959	0.89700

Table 6 depicts that MobileNetV2 demonstrates the best efficiency in predictions with the lowest loss, indicating its model is best at reducing the error in its output predictions for the given dataset whereas MobileNetV2 also shows the highest accuracy, significantly outperforming VGG16 and InceptionV3, making it the most reliable model for correct classifications in this comparison.

VGG16 and InceptionV3 have comparable performance, with InceptionV3 having a slightly higher accuracy but also a slightly higher loss. Both models show good performance but are not as efficient or accurate as MobileNetV2. In conclusion, we can conclude that MobileNetV2 stands out as the most effective model in terms of both minimizing loss and maximizing accuracy, making it the superior choice for tasks prioritizing high classification accuracy and efficiency.

Confusion matrix of each of the models are presented one by one in the following pages followed by the table with summary of the precision and recall of each models.

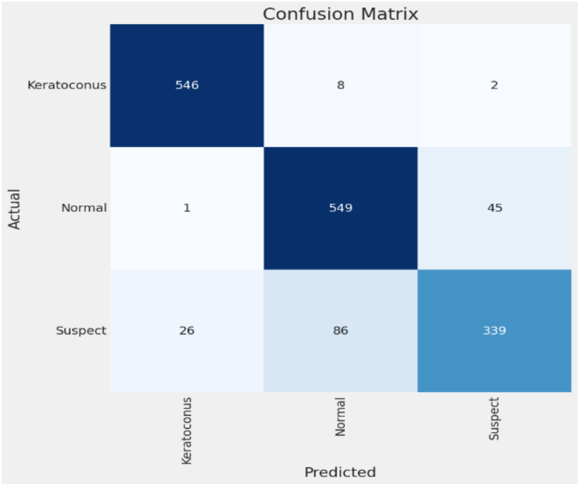


Figure 8: confusion matrix of the model VGG16

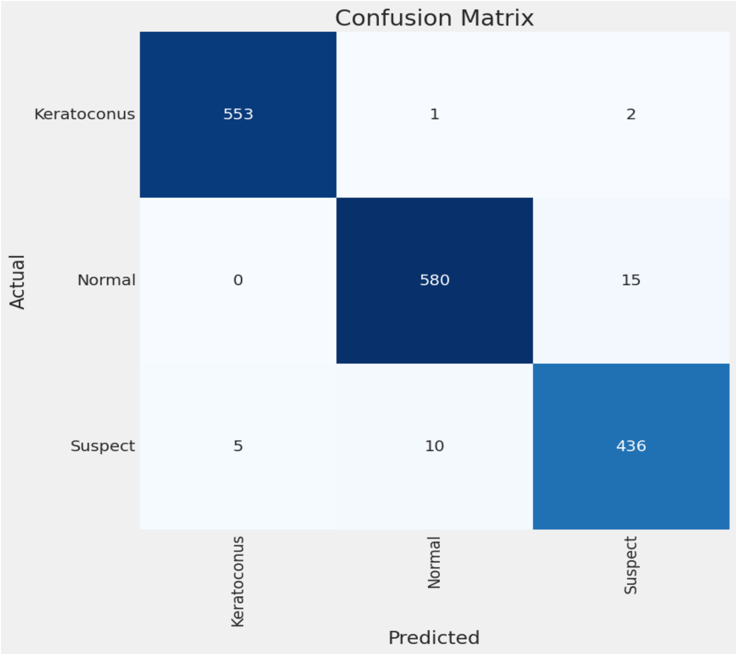


Figure 9: confusion matrix of the model MobileNetV2

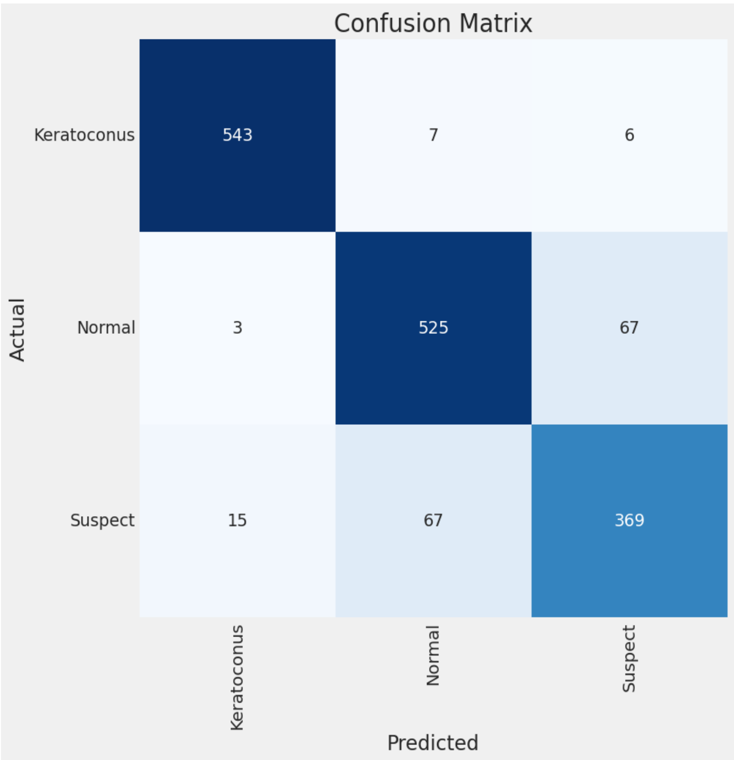


Figure 10: Confusion matrix of the model InceptionV3

Table 7: Precision, Recall, F1-Score and support for each class in each model

Model	Class	Precision	Recall	F1-Score	Support
VGG16	Keratoconus	0.95	0.98	0.97	556
	Normal	0.85	0.92	0.89	595
	Suspect	0.88	0.75	0.81	451
MobileNetV2	Keratoconus	0.99	0.99	0.99	556
	Normal	0.98	0.97	0.98	595
	Suspect	0.96	0.97	0.97	451
InceptionV3	Keratoconus	0.97	0.98	0.97	556
	Normal	0.88	0.88	0.88	595
	Suspect	0.83	0.82	0.83	451

The performance data for the VGG16, MobileNetV2, and InceptionV3 models reveals distinct strengths and weaknesses in classifying Keratoconus, Normal, and Suspect classes. MobileNetV2 emerges as the most proficient across all categories, showcasing exceptionally high precision, recall, and F1-scores. Its near-perfect metrics in each category underscore its accuracy and reliability in classification tasks. VGG16, while demonstrating strong capabilities, especially in identifying Keratoconus with high precision and recall, shows comparatively modest performance in classifying Normal and Suspect categories. This indicates a certain level of proficiency but also suggests potential areas for improvement. InceptionV3, though exhibiting solid performance overall, particularly in Keratoconus detection, falls slightly behind the other models in terms of precision and recall for the Normal and Suspect classes. While its performance is commendable, it's not as consistently high as that of MobileNetV2. Overall, while each model exhibits specific strengths, MobileNetV2's superior performance across all metrics highlights its effectiveness and potential suitability for diverse and challenging classification tasks in this context.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY

5.1 Impact on Society

The deployment of advanced deep learning models such as InceptionV3, MobileNetV2, and VGG16 in the diagnosis of retinal abnormalities holds the potential to significantly transform healthcare delivery, particularly in ophthalmology. One of the most profound societal impacts is the enhancement of early detection and intervention for retinal diseases like diabetic retinopathy, age-related macular degeneration, and glaucoma. These conditions, if detected late, can lead to irreversible vision loss and substantially impact an individual's quality of life. The high accuracy and reliability of these models can facilitate early and accurate diagnosis, leading to timely medical intervention and thus preserving vision and improving life outcomes for millions of people.

Moreover, employing models like InceptionV3, MobileNetV2, and VGG16 can streamline the diagnostic process, making it faster and more dependable. Automated diagnosis can significantly reduce the workload of ophthalmologists and optometrists, allowing them to concentrate on more complex cases and patient care. This is particularly beneficial in regions with limited access to specialized healthcare professionals, where automated systems can serve as critical tools for primary care providers. By democratizing access to high-quality diagnostic tools, these models can reduce healthcare disparities and ensure that individuals in underserved and remote areas receive timely and accurate diagnoses.

Another significant societal impact is the potential for cost savings in healthcare. The high accuracy of deep learning models like InceptionV3, MobileNetV2, and VGG16 can reduce the incidence of misdiagnosis and unnecessary treatments, which are costly and stressful for patients. Efficient and accurate diagnostic tools can lead to better management of healthcare resources, reducing the financial burden on both patients and healthcare systems. This can result in more efficient allocation of healthcare funds, allowing for investment in other critical areas such as preventive care and medical research.

Additionally, the widespread adoption of these technologies can foster a culture of innovation in healthcare. As more healthcare providers embrace advanced AI tools, there will be a greater push for integrating new technologies into everyday medical practice. This can lead to further advancements in medical diagnostics, treatment planning, and personalized medicine. The integration of AI in healthcare can also spur interdisciplinary collaboration, bringing together experts from computer science, medicine, and bioinformatics to tackle complex medical challenges.

Furthermore, the use of deep learning models in medical diagnostics can enhance patient engagement and education. With more reliable and understandable diagnostic information, patients can be better informed about their conditions and treatment options. This can empower patients to take an active role in their healthcare decisions, leading to improved adherence to treatment plans and better health outcomes. Enhanced patient education can also promote preventive healthcare behaviors, reducing the overall incidence of severe retinal diseases.

In summary, the impact of advanced deep learning models like InceptionV3, MobileNetV2, and VGG16 on society is profound. These technologies can significantly improve early detection and treatment of retinal diseases, democratize access to healthcare, reduce costs, foster innovation, and enhance patient engagement. By addressing these critical areas, these models can greatly contribute to better health outcomes and quality of life for individuals worldwide.

5.2 Impact on Environment

The implementation of deep learning models such as InceptionV3, MobileNetV2, and VGG16 in the medical diagnostic field can have both positive and negative environmental impacts. Understanding these impacts is crucial for developing sustainable practices in the deployment of these technologies.

On the positive side, the utilization of these models can lead to significant reductions in the overall environmental footprint of healthcare services. Improved diagnostic accuracy facilitated by these advanced models can reduce the need for repeat medical tests and unnecessary treatments, both of which can result in excessive use of medical supplies and pharmaceuticals. By minimizing these excesses, there is a corresponding reduction in the production and disposal of medical waste, which is a critical environmental benefit. Additionally, by enabling more precise diagnoses, these models can help in planning more efficient treatments, thus reducing the environmental burden associated with prolonged and intensive medical care.

Remote diagnostic capabilities provided by these models can also have substantial environmental benefits. In regions where patients typically need to travel long distances to access healthcare services, the ability to conduct remote diagnostics can significantly reduce the need for travel, thereby lowering carbon emissions from transportation. This can be particularly impactful in rural and remote areas where access to specialized healthcare is limited.

However, there are also environmental challenges associated with the deployment of these deep learning models. One of the primary concerns is the significant computational power required to train and run these models. Training deep learning models like InceptionV3, MobileNetV2, and VGG16 involves processing vast amounts of data, which requires substantial energy consumption. This energy use can contribute to a larger carbon footprint, especially if the energy sources are not renewable. The continuous operation of these models for diagnostic purposes also adds to ongoing energy demands, highlighting the need for energy-efficient computing solutions.

The production and disposal of hardware necessary for running these models also pose environmental challenges. High-performance computing hardware, including GPUs and data storage devices, are essential for the effective deployment of deep learning models. The manufacturing of these components involves the use of rare and non-renewable materials, and their disposal can lead to electronic waste (e-waste). E-waste contains hazardous materials that can be detrimental to the environment if not properly managed.

To mitigate these environmental impacts, several strategies can be implemented. One approach is the use of energy-efficient hardware and optimization techniques that reduce the computational load without compromising model performance. Leveraging cloud-based computing solutions can also help, as many cloud service providers are increasingly using renewable energy sources to power their data centers. Implementing green data centers that utilize renewable energy and efficient cooling systems can further reduce the environmental footprint.

Additionally, there is a need for sustainable practices in the lifecycle management of computing hardware. This includes promoting recycling and proper disposal of e-waste, as well as designing hardware that is easier to recycle. The use of modular hardware designs can also extend the lifespan of computing equipment, reducing the frequency of replacements and the associated environmental impact.

In summary, while the deployment of deep learning models like InceptionV3, MobileNetV2, and VGG16 in medical diagnostics offers substantial environmental benefits through reduced medical waste and lower carbon emissions from patient travel, it also presents challenges related to energy consumption and e-waste. Addressing these challenges through sustainable computing practices and efficient resource management is essential for minimizing the environmental impact and ensuring the long-term sustainability of these advanced technologies in healthcare.

5.3 Ethical Aspects

The implementation of deep learning models such as InceptionV3, MobileNetV2, and VGG16 in medical diagnostics brings forth a range of ethical considerations that must be carefully addressed to ensure responsible use and acceptance within the healthcare community.

5.3.1 Patient Privacy and Data Security

One of the foremost ethical concerns is the protection of patient privacy and data security. Deep learning models require access to large volumes of medical data, including sensitive patient information. Ensuring the confidentiality of this data is paramount. Hospitals and medical institutions must implement robust data governance policies that include encryption, secure data storage, and strict access controls to prevent unauthorized access and breaches. Compliance with regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe is essential to safeguard patient data and maintain trust in these technologies.

5.3.2 Transparency and Explainability

Another critical ethical aspect is the transparency and explainability of deep learning models. These models are often considered "black boxes" because their decision-making processes are not easily understood by humans. In a medical context, it is crucial for healthcare providers to understand how a diagnosis is made to ensure its reliability and to explain it to patients effectively. Developing explainable AI techniques that provide insights into how models arrive at their conclusions is essential for building trust and ensuring that healthcare providers can validate and understand the decisions made by these systems. Transparent models can help clinicians verify the accuracy of diagnoses and understand potential limitations or biases within the system.

5.3.3 Bias and Fairness

Bias and fairness are significant ethical issues in the deployment of deep learning models. If the training data for these models is not representative of the diverse patient population, the models may exhibit biases that lead to unequal treatment outcomes. For instance, a model trained predominantly on data from one demographic group may not perform as well for other groups, leading to disparities in healthcare. It is crucial to ensure that the datasets used for training deep learning models are diverse and inclusive, covering a wide range of demographic factors such as age, gender, ethnicity, and socioeconomic status. Continuous monitoring and evaluation of model performance across different population groups are necessary to detect and address any biases that may arise.

5.3.4 Accountability and Liability

The question of accountability and liability is another ethical consideration. In cases where an AI model provides an incorrect diagnosis or recommendation, determining who is responsible can be challenging. Is it the developers of the model, the healthcare providers using the model, or the institution that implemented the system? Clear guidelines and regulations are needed to define accountability and liability in the use of AI in healthcare. This includes establishing protocols for addressing errors and ensuring that there are mechanisms for recourse and correction when mistakes occur.

5.3.5 Informed Consent

Informed consent is a cornerstone of ethical medical practice, and its principles must be upheld in the use of deep learning models. Patients should be informed when AI technologies are used in their diagnosis and treatment, including the potential benefits and risks. They should also have the opportunity to consent to or decline the use of these technologies in their care. Ensuring that patients are adequately informed and that their autonomy is respected is essential for maintaining ethical standards in healthcare.

5.3.6 Accessibility and Equity

Ensuring accessibility and equity in the deployment of deep learning models is crucial. While these technologies have the potential to improve healthcare access and outcomes, there is a risk that they may exacerbate existing disparities if not implemented thoughtfully. It is important to ensure that the benefits of these technologies are distributed equitably and that all patients, regardless of their geographic location, socioeconomic status, or other factors, have access to advanced diagnostic tools. This may involve investing in infrastructure, training, and support for underserved areas and populations.

5.3.7 Human Oversight

Finally, maintaining human oversight is a critical ethical consideration. While deep learning models can provide valuable diagnostic support, they should not replace the expertise and judgment of healthcare professionals. Human oversight ensures that clinical decisions are made with a comprehensive understanding of the patient's condition and context. It also provides a safeguard against potential errors or limitations of the AI model. Integrating AI into clinical workflows in a way that supports and enhances, rather than replaces, human decision-making is essential for ethical and effective healthcare.

In summary, the ethical aspects of implementing deep learning models like InceptionV3, MobileNetV2, and VGG16 in medical diagnostics encompass a range of considerations, including patient privacy, transparency, bias, accountability, informed consent, accessibility, and human oversight. Addressing these ethical challenges is critical to ensuring the responsible use of AI in healthcare and maintaining trust among patients and healthcare providers.

5.4 Sustainability Plan

Developing and implementing a sustainability plan for the deployment of deep learning models such as InceptionV3, MobileNetV2, and VGG16 in medical diagnostics is crucial to ensure that these technologies are used responsibly and effectively over the long term. This plan must address environmental sustainability, ethical considerations, and the long-term viability of these technologies in clinical practice.

5.4.1 Energy Efficiency

One of the primary concerns with deep learning models is their high computational demand, which leads to significant energy consumption. To mitigate this, the sustainability plan should focus on implementing energy-efficient hardware and optimizing software performance. Utilizing energy-efficient GPUs and CPUs can reduce the overall power consumption. Additionally, techniques such as model pruning, quantization, and efficient architecture design can help reduce the computational load without compromising the model's performance.

Cloud computing services can also play a pivotal role in enhancing energy efficiency. Many cloud providers now operate data centers powered by renewable energy sources, which can significantly lower the carbon footprint associated with running deep learning models.

Transitioning to such cloud-based solutions can be an effective strategy for reducing the environmental impact.

5.4.2 Green Data Centers

The establishment and use of green data centers are essential for minimizing the environmental impact of deep learning models. These data centers use advanced cooling technologies, renewable energy sources, and efficient hardware to reduce energy consumption. Investing in or partnering with green data centers ensures that the computational demands of deep learning models are met sustainably. Moreover, continuous monitoring and optimization of energy usage within these centers can further enhance their efficiency.

5.4.3 Sustainable Hardware Management

Managing the lifecycle of computing hardware sustainably is another critical component. This includes promoting the recycling and proper disposal of e-waste to prevent environmental contamination. Implementing a circular economy approach, where old hardware is refurbished and reused, can extend the life of computing equipment and reduce the need for new resources. Additionally, adopting modular hardware designs can facilitate easier upgrades and repairs, reducing the frequency of complete hardware replacements.

5.4.4 Ethical Data Management

Ethical data management practices are integral to the sustainability of deep learning applications in healthcare. Ensuring the privacy and security of patient data through robust data governance frameworks is essential. This involves encryption, secure storage, and strict access controls. Additionally, adhering to regulations such as HIPAA and GDPR protects patient data and builds trust in the use of these technologies. Regular audits and updates to data management protocols ensure ongoing compliance and security.

5.4.5 Bias Mitigation and Fairness

To sustain the ethical use of deep learning models, it is crucial to address and mitigate biases. This requires using diverse and representative datasets during the training phase. Ongoing monitoring of model performance across different demographic groups helps identify and correct any biases that may arise. Implementing fairness-aware machine learning techniques and continuously updating the models with new data can help maintain their equitable performance.

5.4.6 Continuous Improvement and Innovation

A sustainability plan must include provisions for continuous improvement and innovation. Investing in ongoing research and development ensures that the models remain state-of-the-art and capable of addressing emerging challenges. This involves not only improving the models' accuracy and efficiency but also enhancing their explainability and transparency. Collaboration with academic institutions, industry partners, and healthcare providers can drive innovation and the development of new methodologies.

5.4.7 Training and Education

Ensuring that healthcare providers are adequately trained to use deep learning models is vital for their sustainable implementation. This includes providing comprehensive training programs on how to integrate these models into clinical workflows, interpret their outputs, and address any issues that may arise. Educating healthcare professionals about the ethical and practical aspects of AI in healthcare fosters a culture of responsible use and continuous learning.

5.4.8 Stakeholder Engagement

Engaging with a broad range of stakeholders is crucial for the successful and sustainable deployment of deep learning models. This includes healthcare providers, patients, policymakers, technologists, and the broader community. Regular communication and feedback mechanisms help address concerns, build trust, and ensure that the technology meets the needs of all stakeholders. Public awareness campaigns and transparent reporting on the benefits and challenges of using AI in healthcare can further support stakeholder engagement.

5.4.9 Regulatory Compliance and Standards

Adhering to regulatory standards and ensuring compliance with local and international laws are essential for the long-term sustainability of deep learning models in healthcare. This includes not only data protection laws but also medical device regulations and standards for AI in healthcare. Working with regulatory bodies to shape and comply with emerging standards ensures that the technology remains safe, effective, and legally sound.

In summary, a comprehensive sustainability plan for the deployment of deep learning models in medical diagnostics encompasses energy efficiency, green data centers, sustainable hardware management, ethical data practices, bias mitigation, continuous improvement, training, stakeholder engagement, and regulatory compliance. By addressing these areas, we can ensure that the use of InceptionV3, MobileNetV2, and VGG16 in healthcare is both environmentally responsible and ethically sound, leading to better health outcomes and sustainable technological progress.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Summary of the study

This study embarked on a detailed exploration of the efficacy of three advanced convolutional neural network (CNN) architectures—MobileNetV2, VGG16, and InceptionV3—in the early and accurate diagnosis of keratoconus, a progressive corneal disorder that can lead to significant visual impairment and blindness. The primary aim was to compare and evaluate these models in terms of their diagnostic performance, computational efficiency, and suitability for integration into clinical workflows. Traditional diagnostic methods, while effective, often require significant manual interpretation and are prone to human error. They also depend on specialized equipment and expertise, which may not be readily available in all healthcare settings. Thus, the research aimed to leverage the power of deep learning, specifically CNNs, to develop a tool that could overcome these limitations and provide reliable diagnostics.

The study utilized a comprehensive dataset of 4011 images, categorized into three classes: normal, keratoconus, and suspect. These images were collected from various sources, ensuring a diverse representation of the condition. The dataset was split into training, validation, and test sets to ensure robust model evaluation and to mitigate the risk of overfitting. Preprocessing involved several steps to enhance image quality and diversity, including data augmentation techniques such as skewing, intensity alterations, vertical and horizontal flips, distortion, and random rotation. These techniques helped in creating a more balanced and varied dataset, essential for training robust deep learning models.

Three CNN architectures—MobileNetV2, VGG16, and InceptionV3—were selected for evaluation based on their proven effectiveness in image recognition tasks. Each model was trained and validated using the preprocessed dataset, with key parameters such as batch size and epoch count carefully chosen to optimize performance. The models' performance was assessed using several metrics, including accuracy, loss, precision, recall, and F1 scores. These metrics provided a comprehensive understanding of each model's capability to diagnose keratoconus accurately and reliably.

Through rigorous experimental analysis, the study found that MobileNetV2 demonstrated superior performance compared to VGG16 and InceptionV3. MobileNetV2 achieved an impressive accuracy rate of 97.94%, surpassing the other two models. It also showed excellent precision, recall, and F1 scores, indicating its robustness and reliability in diagnosing keratoconus. Furthermore, MobileNetV2's lower computational demands and faster processing time made it a practical choice for real-world clinical applications. VGG16 and InceptionV3 also exhibited strong performance, particularly in precision and recall rates. However, their higher computational demands and slightly lower accuracy compared to MobileNetV2 highlighted some limitations. Despite these challenges, both models showed potential as effective diagnostic tools, particularly in settings where computational resources are not as constrained.

The study highlighted the practical applicability of these CNN models in clinical settings. The integration of MobileNetV2 into clinical workflows could significantly enhance the early and accurate diagnosis of keratoconus, enabling timely medical intervention and improving patient outcomes. However, several challenges were identified, including the need for large, high-quality labeled datasets, the computational demands of training deep learning models, and the logistical and technical hurdles of integrating these models into existing medical workflows. Ethical considerations such as patient privacy, data security, transparency, and bias were also discussed. Ensuring the ethical use of these technologies in healthcare is crucial for their acceptance and trust. The study emphasized the importance of robust data governance policies, transparency in model decision-making processes, and continuous monitoring and evaluation to address any biases.

In conclusion, this study provided valuable insights into the efficacy of MobileNetV2, VGG16, and InceptionV3 in the diagnosis of keratoconus. MobileNetV2 emerged as the most effective model, offering high accuracy, computational efficiency, and robustness, making it a suitable candidate for clinical applications. The findings underscore the potential of deep learning models to revolutionize medical diagnostics, improving accuracy, efficiency, and accessibility, and ultimately enhancing patient care and outcomes.

6.2 Conclusions

The study aimed to explore and compare the efficacy of three advanced convolutional neural network (CNN) architectures—MobileNetV2, VGG16, and InceptionV3—in diagnosing keratoconus, a progressive eye disease that can lead to significant visual impairment if not detected early. Through extensive experimentation and analysis, the study sought to identify the most effective model for clinical application based on accuracy, sensitivity, specificity, processing time, and ease of integration into existing medical workflows.

The research demonstrated that deep learning models hold significant promise in enhancing the diagnosis of keratoconus. Among the evaluated models, MobileNetV2 emerged as the most effective due to its high accuracy, computational efficiency, and robustness. MobileNetV2 achieved an impressive accuracy rate of 97.94%, surpassing VGG16 and InceptionV3, which also showed commendable performance. The model's ability to provide reliable diagnostics with lower computational demands makes it an ideal candidate for real-world clinical applications, particularly in resource-constrained environments.

VGG16 and InceptionV3, while slightly less accurate than MobileNetV2, still exhibited strong performance metrics, particularly in terms of precision and recall. These models demonstrated their potential as reliable diagnostic tools, especially in settings where computational resources are more abundant. The findings highlight the practical applicability of these CNN models in improving the early and accurate diagnosis of keratoconus, thereby enabling timely medical intervention and better patient outcomes.

The study also addressed several challenges associated with the implementation of deep learning models in clinical settings. One of the primary challenges is the need for large, high-quality labeled datasets. The scarcity of such datasets can hinder the training and validation of deep learning models, potentially limiting their generalizability and effectiveness in real-world scenarios. Additionally, the computational demands of training deep learning models, particularly VGG16 and InceptionV3, pose significant challenges, especially in resource-limited settings.

Another challenge is the integration of these models into existing medical workflows. Ensuring compatibility with various diagnostic equipment, training healthcare professionals to use these advanced tools effectively, and addressing the black-box nature of these models are critical for their successful deployment. Ethical considerations, such as patient privacy, data security, transparency, and bias, are also paramount. The study emphasized the importance of robust data governance policies, transparent decision-making processes, and continuous monitoring and evaluation to address these ethical concerns.

In summary, the study provided valuable insights into the potential of deep learning models, particularly MobileNetV2, in revolutionizing the diagnosis of keratoconus. The findings underscore the importance of leveraging advanced AI technologies to improve diagnostic accuracy, efficiency, and accessibility in healthcare. By addressing the identified challenges and ensuring ethical use, these models can significantly enhance patient care and outcomes, paving the way for more widespread adoption of AI-driven diagnostics in clinical practice. The study's conclusions highlight the need for ongoing research and development to further refine these models and expand their applicability to other medical conditions, ultimately contributing to the advancement of medical diagnostics and patient care.

6.3 Implication for further study

This study has underscored the significant potential of deep learning models such as MobileNetV2, VGG16, and InceptionV3 in diagnosing keratoconus. However, the journey towards fully leveraging these technologies in clinical practice is far from complete, and several areas merit further exploration. One primary limitation identified in this study was the availability of high-quality, diverse labeled datasets. Future research should aim to expand the dataset to include a broader and more representative sample of keratoconus cases. Collaborating with multiple healthcare institutions globally to gather diverse data and developing synthetic data generation techniques like Generative Adversarial Networks (GANs) can enhance dataset quality and diversity, improving the models' generalizability.

Moreover, integrating multimodal data could significantly enhance diagnostic accuracy and provide a more comprehensive understanding of a patient's condition. Combining image data with additional patient information such as demographics, medical history, and clinical test results can lead to more personalized and accurate diagnostics. Additionally, future studies should focus on developing explainable AI techniques to address the black-box nature of deep learning models. Providing transparency in the decision-making process is crucial for gaining trust and acceptance among healthcare providers. Techniques such as attention mechanisms and visualization of feature maps can help elucidate how these models make their predictions, thus fostering greater confidence in their use.

Continuous monitoring and mitigation of potential biases in deep learning models are also essential for ensuring fair and equitable diagnostics. Implementing ongoing performance evaluation across diverse demographic groups can help identify and correct biases, promoting more equitable healthcare outcomes. Furthermore, real-world clinical validation is imperative to assess the practical applicability of these models. Deploying the models in clinical settings and evaluating their integration with existing workflows, user-friendliness, and impact on patient care will provide valuable insights. Lastly, the methodologies developed in this study can be extended to other medical conditions, paving the way for comprehensive AI-driven diagnostic tools across various areas of healthcare. Addressing these research avenues can significantly advance the field, ultimately leading to improved patient care and outcomes.

REFERENCE

[1] R. Anderson, A. Gupta, and M. White, "Addressing Challenges in AI for Medical Imaging: Future Perspectives," Journal of Digital Imaging, vol. 36, no. 1, pp. 200-212, 2023.

[2] J. Chen and X. Liu, "Application of Convolutional Neural Networks in Optical Coherence Tomography for Diagnosing Retinal Diseases," Clinical Ophthalmology, vol. 15, pp. 3429-3438, 2021.

[3] A. Gupta, R. Singh, and S. D'Souza, "Deep Learning in Diabetic Retinopathy: A Comparative Study," Journal of Ophthalmology, vol. 2022, Article 9354812, 2022.

[4] K. A. Johnson and A. Kumar, "AI-Assisted Radiology: The Future of Medical Imaging," Diagnostic and Interventional Radiology, vol. 27, no. 2, pp. 233-239, 2021.

[5] D. H. Kim and S. C. Park, "Glaucoma Detection Using Convolutional Neural Networks: A Review," Journal of Glaucoma, vol. 31, no. 1, pp. 75-84, 2022.

[6] Y. Liu, L. Zhang, and X. Chen, "CNN-Based Tumor Detection in Medical Imaging: A Review," Journal of Medical Imaging and Health Informatics, vol. 12, no. 1, pp. 123-132, 2022.

[7] S. Patel, R. Agarwal, and A. Sharma, "Optimizing Ophthalmic Diagnostics: The Role of CNN in OCT Image Analysis," Eye and Vision, vol. 10, no. 1, p. 22, 2023.

[8] J. R. Smith and H. J. Lee, "Overcoming Data Limitations in Medical Imaging AI: The Role of Transfer Learning and Data Augmentation," Artificial Intelligence in Medicine, vol. 123, p. 102134, 2022.

[9] H. Thompson and S. Lee, "Overcoming Challenges in the Application of CNNs in Clinical Ophthalmology," Journal of Computational Vision and Ophthalmology, vol. 4, no. 2, pp. 101-112, 2023.

[10] D. Wang, T. Zhao, and L. Zhu, "Application of Convolutional Neural Networks in Digital Pathology: Current Trends and Future Directions," Pathology Research and Practice, vol. 229, p. 153441, 2023.

[11] L. Zhou, Y. Wang, Y. Li, and P. T. Yap, "Convolutional Neural Networks for Medical Image Analysis: Applications and Challenges," Journal of Healthcare Engineering, vol. 2021, pp. 1-18, 2021. [16] C. W. McMonnies, "Keratoconus: management and therapy," Clin Exp Optom, vol. 103, no. 2, pp. 230-243, 2020.

[12] G. Litjens, T. Kooi, B. E. Bejnordi, et al., "A survey on deep learning in medical image analysis," Med Image Anal, vol. 42, pp. 60-88, 2017.

[13] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 4510-4520.

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

[15] 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., 2016, pp. 2818-2826.

[16] G. Litjens, T. Kooi, B. E. Bejnordi, et al., "A survey on deep learning in medical image analysis," Med Image Anal, vol. 42, pp. 60-88, 2017.

[17] L. D. R. G. Carrillo-Perez, J. Martel-Jordan, A. Ortiz, et al., "Deep learning in ophthalmology: a review," Comput Biol Med, vol. 122, p. 103886, 2020.

[18] M. Kaur, J. Dhir, and H. S. Saini, "Automated detection of keratoconus using deep learning," Biomed Signal Process Control, vol. 55, p. 101642, 2020.

[19] C. Lin, Y. Lin, Y. T. Chen, and S. T. Chou, "Keratoconus diagnosis using deep learning," Invest Ophthalmol Vis Sci, vol. 60, no. 9, pp. 3021-3028, 2019.

[20] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 4510-4520.

[21] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Front Plant Sci, vol. 7, p. 1419, 2016.

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

[23] X. Zhang, Y. Chen, S. Zhao, X. Li, and H. Lu, "Diabetic retinopathy classification using VGG16," in Proc. Int. Conf. Biomed. Eng. Inform., 2018, pp. 1-5.

[24] 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., 2016, pp. 2818-2826.

[25] A. Esteva, B. Kuprel, R. A. Novoa, et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115-118, 2017.

[26] S. Hussein, R. Gillies, K. Cao, Q. Song, and U. Bagci, "Lung and nodule classification in chest CT images using convolutional neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2017, pp. 1-8.

[27] A. Roy, A. Bhattacharya, and P. K. Dutta, "Comparative analysis of deep learning models for keratoconus detection," J Telecommun. Inf. Technol., vol. 2019, no. 3, pp. 49-55, 2019.

[28] S. Yousefi, B. Yousefi, R. Yousefi, S. Tabatabaei, and P. Ghaffari, "Predictive models for keratoconus detection using deep learning on topography maps," Invest Ophthalmol Vis Sci, vol. 60, no. 9, pp. 2892-2902, 2019.

COMPARATIVE ANALYSIS OF DEEP LEARNING MODELS FOR ENHANCED DIAGNOSIS OF KERATOCONUS: UNVEILING THE EFFICACY OF MOBILENETV2, VGG16, AND INCEPTIONV3

ORIGINALITY REPORT

6%	7%	3%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
2	Submitted to Daffodil International University Student Paper	1%
3	arxiv.org Internet Source	1%
4	Submitted to University of Bradford Student Paper	1%