School of Computer Science and Engineering

Faculty of Engineering

The University of New South Wales

Myopia Detection, Incidence and Progression Prediction in Adolescents with Machine Learning Methods

by

Zhao WANG

Thesis submitted as a requirement for the degree of

Master of Information Technology

Submitted: July 2021 Student ID: z5219645

Supervisor: Prof. Arcot Sowmya Topic ID: COMP9993

Acknowledgments

Postgraduate study is expected to conclude upon completion of the Research Project. This thesis was written under Professor Sowmya's close supervision. I want to express my sincere gratitude to her. I would also want to express my gratitude to Professor Padmaja, Dr. Thomas, and Dr. Divya at Brien Holden Vision Institute for their assistance and feedback on the optometry portion of my thesis. This thesis would not have been feasible without their kindness and willingness to share their findings.

I am extremely thankful for my friends: Hasindu, Ying, Pingyang, Yongzhen, Ciel, Jack, Tak, Max, Oscar, Matias, Tim, Fred and Johnny who have been there for me throughout the year.

Finally, I would want to express my gratitude to my family for their unwavering support, which has enabled my graduate to really live up to the initial goal and grow holistically. Although I have many regrets, I am still ready to go on. Learn how to get along with yourself, learn how to get along with the world, I believe this is a subject that everyone must confront, and I believe I have discovered my own solution at this point. Thank you to everyone who has supported me, guided me, and assisted me.

Table of Contents

Acknowledgments I

Table of Contents II

Abstract IV

List of Figures V

List of Tables VII

List of Abbreviations VIII

Chapter 1. Introduction 1

Chapter 2. Background 3

2.1 Myopia 3

2.2 Dataset 5

2.2.1 Exploratory Data Analysis 6

2.3 Data Mining Methods 9

2.3.1 Ensemble Learning 10

2.3.2 Random Forest 14

2.3.3 Gradient Boosting Decision Tree 14

2.3.4 Support Vector Machine 15

2.4 Deep Learning Methods 15

2.4.1 Convolutional Neural Networks 15

2.4.2 CNN for Image Classification 16

2.5 Related Literatures 19

Chapter 3. Data Mining Models for Myopia Detection, Incidence and Progression Prediction 22

3.1 Methods 22

3.1.1 Data Mining Models 22

3.1.2 Model Evaluation Criteria 23

3.2 Results 26

3.2.1 Results without Axial Length to Corneal Radius Ratio 26

3.2.2 Results with Axial Length to Corneal Radius Ratio 31

3.3 Summary & Discussion 36

3.3.1 Summary 36

3.3.2 Discussion 36

Chapter 4. Deep Learning Models for Myopia Detection, Incidence and Progression Prediction 38

4.1 Methods 38

4.1.1 Data Augmentation 38

4.1.2 Model Design 42

4.1.3 Convolutional Neural Network Optimization 45

4.1.4 Model Pipeline 47

4.1.5 Model Evaluation Criteria 48

4.1.6 Model Visualization 49

4.2 Results 50

4.2.1 Results of Image Data 50

4.2.2 Results of Multi-modal Data Training 55

4.3 Summary & Discussion 57

4.3.1 Summary 57

4.3.2 Discussion 58

Chapter 5. Conclusion & Future Works 59

5.1 Conclusion 59

5.1.1 Summary 59

5.1.2 Conclusion 60

5.2 Future works 61

References 62

Abstract

Myopia is a growing global public health concern. Myopia will result in vision loss and is a significant cause of blindness if left uncorrected. It is critical to conduct studies on the risk of acquiring myopia or hyperopia in adolescents in order to intervene as soon as feasible.

The purpose of this project is to build a system for identifying myopia in teens, forecasting its onset and progression, and providing medical advises using machine learning.

This article demonstrates that the LGBM model beats the Logistic Regression model in detecting myopia but falls short in predicting myopia incidence and development without requiring time-consuming manual features selection. In detection, incidence, and progression prediction tasks, the feature axial length to corneal radius ratio can improve LGBM AUC values by 7%, 10%, and 14%, respectively. This feature's significance is consistent with clinical research findings.

Without the need for laborious feature extraction, deep learning can produce higher AUC results simply by inputting retinal image data. Models of different structures and sizes are trained and optimized. As demonstrated through model visualisation, the neural network learns information that is unique from optometry research. Finally, this work modified the model to carry out multi-modal data training and discussed the results.

Additional refinements to enhance models are also apparent, and further study is suggested.

List of Figures

Figure 2-1 Myopia vs Normal Eye[7] 3

Figure 2-2 The estimated global prevalence of myopia and high myopia per decade from 2000 to 2050 based on current trends[2] 4

Figure 2-3 Posterior segment changes associated with myopia[6] 5

Figure 2-4 Annotated retinal images[14] 6

Figure 2-5 View of first 38 features of the dataset 7

Figure 2-6 Axis Length distribution with Myope and Non-myope at baseline 7

Figure 2-7 Myope distribution at baseline month 8

Figure 2-8 Myope incidence distribution at 12 months after baseline 8

Figure 2-9 Myope distribution with gender 9

Figure 2-10 Bagging vs Boosting[16] 11

Figure 2-11 Bias vs Variance[17] 12

Figure 2-12 Convolutional kernel 16

Figure 2-13 ImageNet Top-5 Classification Error Over Time[21] 17

Figure 2-14 The development of neural network structures for classification 18

Figure 2-15 ROC curve for the predictive model to detection of myopia with LR model[6] 20

Figure 2-16 ROC curve for predictive model for incidence of myopia with LR model[6] 20

Figure 2-17 ROC curve for the predictive model for the progression of myopia with LR model[6] 20

Figure 3-1 Hyper-parameter search with Bayesian Optimization Algorithm 23

Figure 3-2 An ROC Curve example 24

Figure 3-3 ROC curves of each detection model on datasets 27

Figure 3-4 ROC curves of each incidence prediction model on datasets 28

Figure 3-5 ROC curves of each progression prediction model on datasets.....29

Figure 3-6 ROC curves of each detection model on datasets.....32

Figure 3-7 ROC curves of each incidence prediction model on datasets.....33

Figure 3-8 ROC curves of each progression prediction model on datasets.....34

Figure 4-1 Retinal Images Geometric Transformation.....39

Figure 4-2 Retinal Images Colour Transformation.....40

Figure 4-3 Retinal Images cut and mix operation......41

Figure 4-4 ResNet architectures(above) and ResNet 50 architecture (below)[36].....42

Figure 4-5 Model Size Vs ImageNet accuracy[37].....43

Figure 4-6 Stem and Final Layers in EfficientNet[38].....44

Figure 4-7 Five modules for architecture construction[38].....44

Figure 4-8 Sub-blocks for architecture construction[38].....45

Figure 4-9 Architecture for EfficientNet-B4[38].....45

Figure 4-10 Hyper-parameter search with TPE Algorithm.....47

Figure 4-11 Pipeline Flow for Image Data Training.....48

Figure 4-12 Pipeline Flow for Multi-modal Data Training.....48

Figure 4-13 An incidence prediction image of model visualization.....49

Figure 4-14 ROC curves of each model on image data of three tasks.....51

Figure 4-15 GRAD-CAM visualization on Resnet models of three tasks with different retinal images.....54

Figure 4-16 ROC results of EfficientNet-B4 model on different datasets of three tasks 56

List of Tables

Table 2-1 Comparison of different data mining models[15] 10

Table 2-2 Common CNN architectures for classification 18

Table 3-1 Data Mining Models for datasets and tasks in this study 22

Table 3-2 Confusion Matrix 23

Table 3-3 AUC results of each model on datasets on three tasks 26

Table 3-4 Pair-wised Independent Samples t-test on Detection Results 30

Table 3-5 Pair-wised Independent Samples t-test on Incidence Prediction Results 30

Table 3-6 Pair-wised Independent Samples t-test on Progression Prediction Results 30

Table 3-7 AUC results of each model on datasets on three tasks 31

Table 3-8 Pair-wised Independent Samples t-test on Detection Results 35

Table 3-9 Pair-wised Independent Samples t-test on Incidence Prediction Results 35

Table 3-10 Pair-wised Independent Samples t-test on Progression Prediction Results 35

Table 4-1 AUC results of each model on image data of three tasks 50

Table 4-2 AUC results of EfficientNet-B4 model on different datasets of three tasks 55

List of Abbreviations

AL Axial length

AL/CR Axial length to corneal radius ratio

AUC Area Under Curve

BHVI The Brien Holden Vision Institute

CNN Convolutional Neural Network

CV Computer Vision

D Dioptres

FP False Positive

FN False Negative

LGBM Light Gradient Boosting Machine

ML Machine Learning

mm millimetre

NN Neural Network

RE Refractive Error

RF Random Forest

ROC Receiver Operating Characteristic

SGD Stochastic Gradient Descent

SVM Support Vector Machine

TP True Positive

TN True Negative

Chapter 1. Introduction

Myopia is a public health issue that is becoming more prevalent throughout the globe. Light is refracted through the lens and focussed in front of the retina, resulting in a refractive error[1]. According to a research conducted by the American Academy of Ophthalmology, by 2050, 4.7 billion people (about half of the world's population) would develop myopia, with around 10% of the population developing severe myopia[2]. Uncorrected myopia causes vision loss and is a leading cause of blindness. Because excessive myopia may lead to visual pathologies including retinal detachment, glaucoma, cataracts, and macular degeneration[3]. In addition, untreated myopia costs the global economy tens of billions of dollars[4]. As a result, predicting the likelihood of acquiring myopia or extreme myopia in teenagers and intervening as soon as feasible is an important study subject.

A large quantity of data has been gathered to study the incidence and development of myopia in teenagers. With the application of statistical learning, there provides an opportunity for data analysis. Deep learning models with automated feature learning have been the primary force pushing machine learning methods after the removal of the computational performance restriction. Image segmentation, classification, object recognition, and retrieval tasks have all yielded acceptable results using CNNs[5].

The work in this thesis is based on research by Divya[6] from BHVI. Divya carried out semi-automatic feature extraction of fundus images, and manually screened the extracted features. Finally, Logistic Regression models were established to detect myopia and predict myopia incidence and progression of young children. BHVI provides a collection of fundus images and associated semi-automatically extracted features and clinical data.

As a result, this research is intended to develop a system for detecting myopia, predicting its occurrence, and development in young children using machine learning methods in order to offer medical advice.

This study indicates that while the LGBM model outperforms the Logistic Regression model in terms of detecting myopia, it falls short in terms of forecasting myopia incidence and development without requiring time-consuming manual feature selection. The feature axial length to corneal radius ratio can increase LGBM AUC values by 7%, 10%, and 14%,

respectively, in detection, incidence, and progression prediction tasks. The significance of this characteristic is consistent with clinical research findings.

Without the requirement for time-consuming feature extraction, deep learning may provide superior AUC values just by inputting retinal image data. We train and optimize models with a variety of various topologies and sizes. As shown via model visualization, the neural network acquires knowledge not relevant to optometry research. Finally, this work changed the model and analyzed the outcomes in order to perform multi-modal data training. Additional model refinements are also apparent, and further research is recommended.

The background material in Chapter 2 includes data mining methodologies, deep learning ideas employed in this thesis, and related research. Chapter 3 discusses data mining models for detecting myopia, predicting its occurrence, and predicting its progression, while Chapter 4 discusses deep learning models for detecting myopia, predicting its incidence, and predicting its progression. Finally, Chapter 5 summarises the research findings and makes recommendations for future research.

Chapter 2. Background

We begin this chapter by discussing myopia's background and the data set used in this study. After that, it discusses data mining and deep learning methods. Finally, this chapter summarizes published works on myopia research.

2.1 Myopia

Myopia is a refractive error, in which light is refracted through the lens and focused in front of the retina[1]. The tear film, cornea, and lens of the eye bend light in such a way that it focuses on the retina. The retina is the film that receives the image produced by these light rays. It transmits the signal to the brain through the optic nerve. Myopia develops when the eye is either excessively long or has an abnormally steep cornea. As a consequence, light rays concentrate forwards of the retina rather than directly on it. In this scenario, you will see close things clearly but will perceive distant ones as fuzzy.

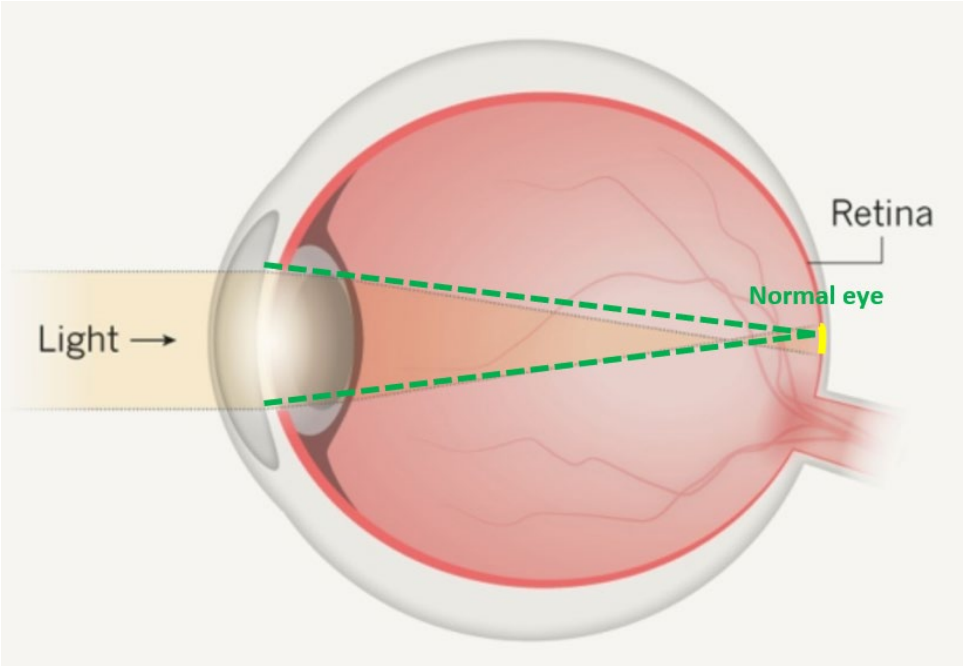


Figure 2-1 Myopia vs Normal Eye[7]

Myopia is the most frequent cause of visual impairment in individuals under the age of 40. Its incidence has been increasing at an alarming pace in recent years.

Globally, research indicates that about 25% of the world's population was short-sighted in 2000, but by 2050, it is predicted that nearly half of the world's population would be myopic[2].

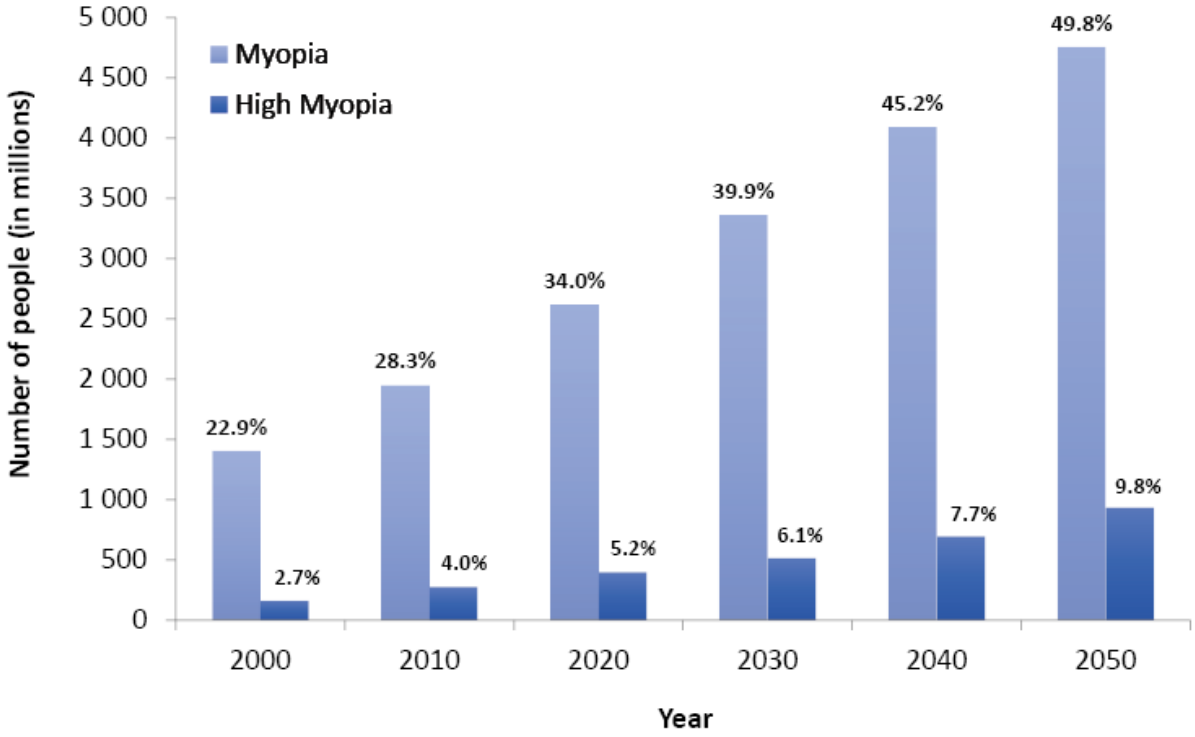


Figure 2-2 The estimated global prevalence of myopia and high myopia per decade from 2000 to 2050 based on current trends[2]

Myopia, particularly severe myopia, has been linked to a variety of ocular complications. Vitreoretinal diseases, particularly peripheral pathologic alterations in the retina, are well-known complications of hyperopia. A cross-sectional research discovered that up to 61.7 percent of severely myopic eyes had peripheral retinal changes[8]. High myopia has also been linked to bilateral rhegmatogenous retinal detachment, a disease associated with significant vision impairment[9]. Nuclear cataract was shown to be more prevalent in highly myopic eyes with increasing axial length. High myopes also likely to develop cataracts with a greater nuclear density than normal controls[10]. Both nuclear and posterior subcapsular cataracts are strongly related with myopia[11]. Myopia was linked to idiopathic focal subretinal neovascularization[12].

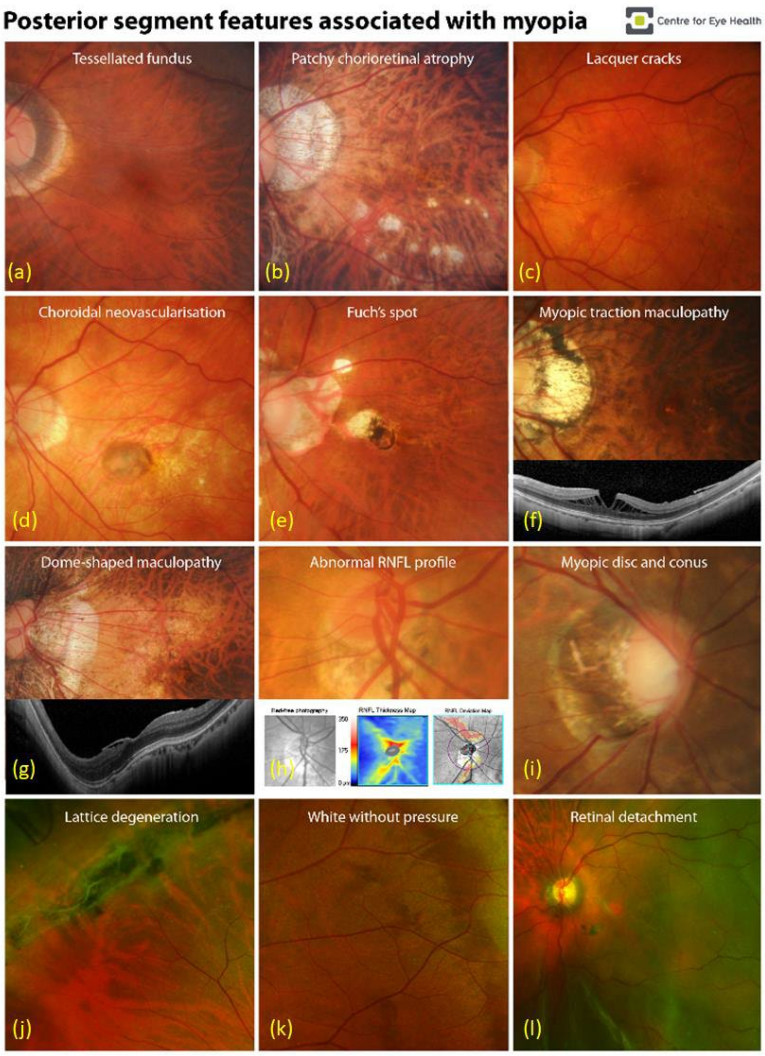


Figure 2-3 Posterior segment changes associated with myopia[6]

2.2 Dataset

The Brien Holden Vision Institute provided a collection of 2875 fundus photographs and accompanying labelling (BHVI). These are retinal fundus photographs taken with a 45-degree field of view using optical coherence tomography equipment. Additional demographic information was provided, including the patient's age at the time of the appointment, parental myopia, gender, and axial length.

The data for analysis came from a cluster-based, randomised prospective school-based population study conducted in Shanghai, China from October 2016 to November 2018[13]. The Shanghai General Hospital Ethics Committee, Shanghai Eye Diseases Prevention and Treatment Centre (SEDPTC), Shanghai, China (ClinicalTrials.gov Identifier: NCT02980445),

and the UNSW Ethics Committee (project no: HC 190575) approved this study. It was conducted in accordance with the Declaration of Helsinki's principles for human experimentation.

Only data from children aged 12 to 24 months and children aged 6 to 9 years were included in the current analysis. Data from the preceding 12 months were used as a baseline. The sample consisted of 2875 children enrolled in 24 primary schools located throughout Shanghai, China's eight districts. Figures below show the annotated retinal images.

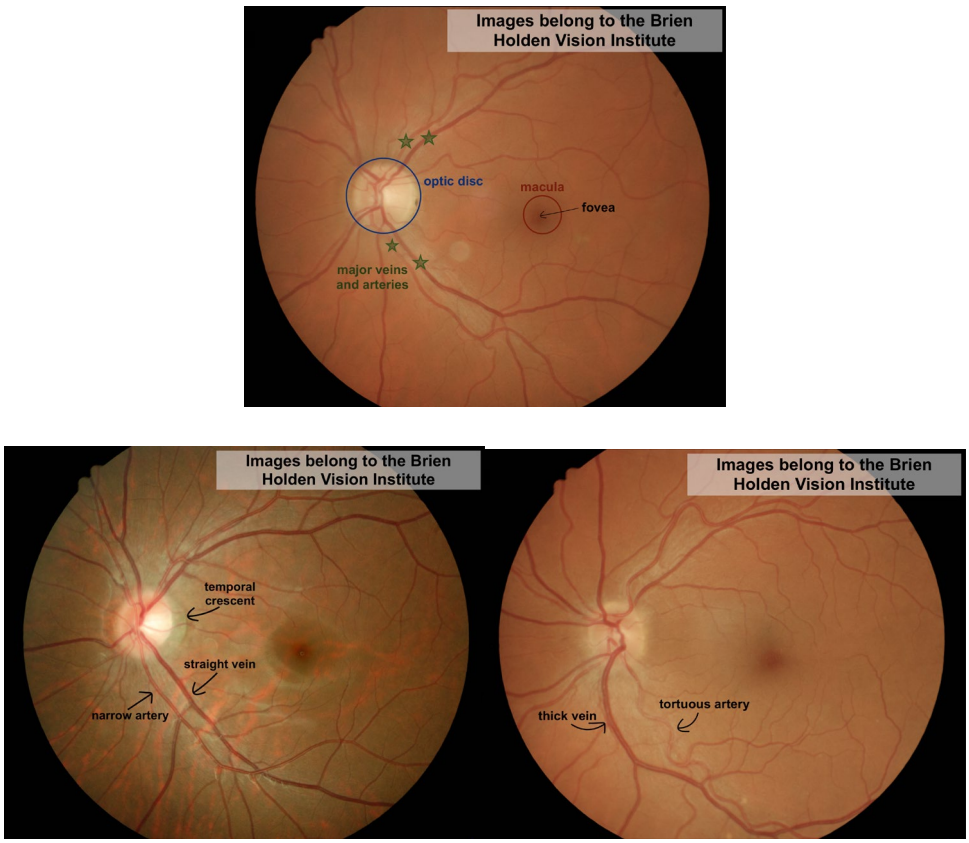


Figure 2-4 Annotated retinal images[14]

2.2.1 Exploratory Data Analysis

Brien Holden Vision Institute (BHVI) supplied a dataset of semi-automated characteristics extracted from 2875 fundus pictures of children and related refractive error in 12 and 24 months. Additional demographic data was also provided, including age at the time of the appointment, parental myopia and gender, as well as machine data. First 38 features were listed in Figure 2-5.

```
RangeIndex: 2875 entries, 0 to 2874
Data columns (total 240 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0    Filename                     2875 non-null   object
 1    Eye                          2875 non-null   object
 2   Parental_Myopia               2753 non-null   float64
 3   Gender                        2875 non-null   object
 4   Age_2017                      2875 non-null   float64
 5   SE_12m                        2875 non-null   float64
 6   SE_24m                        2875 non-null   float64
 7   SE_Prog_12_to_24m             2875 non-null   float64
 8   Myope_12M                     2875 non-null   object
 9   Myope_24M                     2875 non-null   object
 10  RE_Prog_12_to_24M             2875 non-null   object
 11  AL_12m                        2875 non-null   float64
 12  AL_24m                        2875 non-null   float64
 13  AL_Prog_12_to_24m             2875 non-null   float64
 14  AL_by_CR_12m                  2875 non-null   float64
 15  AL_by_CR_24m                  2875 non-null   float64
 16  AL_by_CR_Prog_12_to_24m       2875 non-null   float64
 17  Major axis mm                 2875 non-null   float64
 18  Minor axis mm                 2875 non-null   float64
 19  FDD mm                        2875 non-null   float64
 20  Ovality Index                 2875 non-null   float64
 21   Major Axis                   2875 non-null   float64
 22   Minor Axis                   2875 non-null   float64
 23   Disc Tilt                    2875 non-null   float64
 24   Disc Rotation                2875 non-null   float64
 25   Foveal Dist                  2875 non-null   float64
 26   STRAA                        2875 non-null   float64
 27   ITRAA                        2875 non-null   float64
 28   STRVA                        2875 non-null   float64
 29   ITRVA                        2875 non-null   float64
 30   Fit RMS                      2875 non-null   float64
 31   Disc X                       2875 non-null   float64
 32   Disc Y                       2875 non-null   float64
 33   Fovea X                      2875 non-null   float64
 34   Fovea Y                      2875 non-null   float64
 35   Rel Major Axis               2875 non-null   float64
 36   Rel Minor Axis               2875 non-null   float64
 37   Rel Fov Dist                 2875 non-null   float64
 38   Rel Fit RMS                  2875 non-null   float64
```

Figure 2-5 View of first 38 features of the dataset

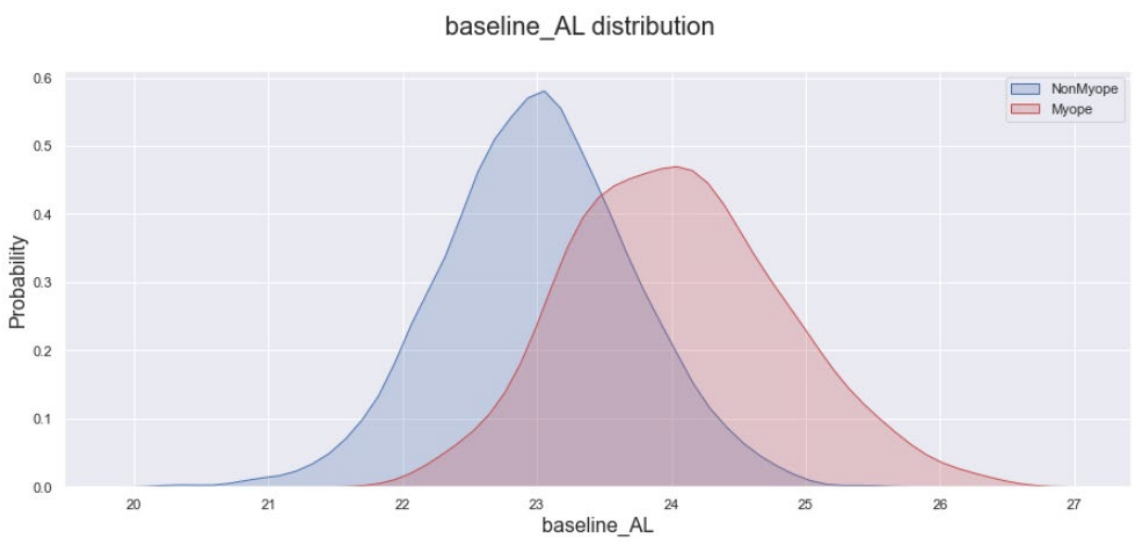


Figure 2-6 Axis Length distribution with Myope and Non-myope at baseline

Distribution analysis of all features was conducted in this study, and the distribution of all features was Gaussian distribution. Figure 2-6 shows the Axis Length Distribution with Myope and Non-Myope in Month 12.

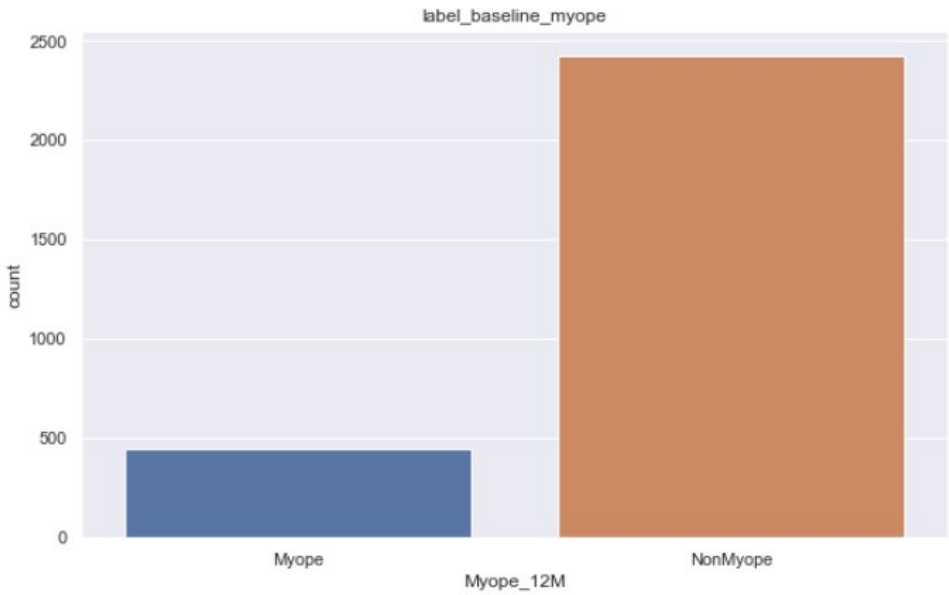


Figure 2-7 Myope distribution at baseline month

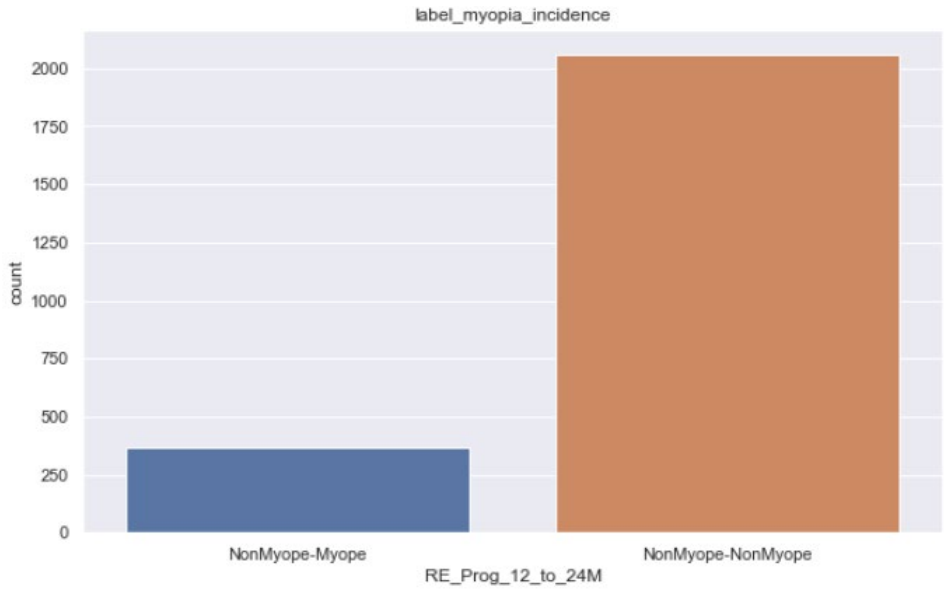


Figure 2-8 Myope incidence distribution at 12 months after baseline

The histograms in Figures 2-7 and 2-8 illustrate the distribution of Myope children at baseline and the incidence of Myope 12 months after baseline.

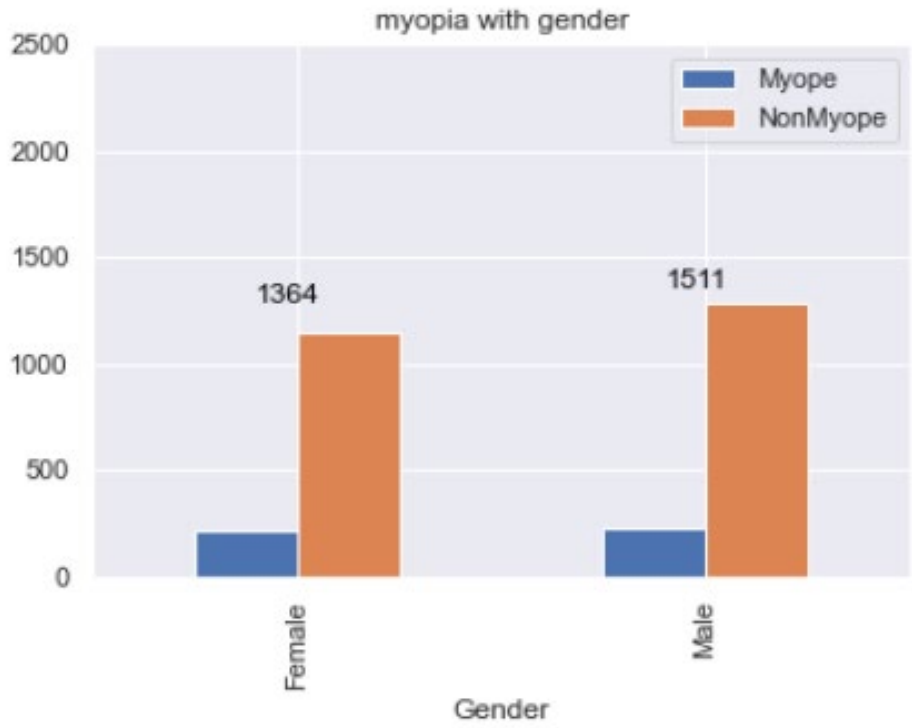


Figure 2-9 Myope distribution with gender

The data is severely imbalanced, as seen in the figures and tables above, with the bulk of samples falling within a low myopic and emmetropic range. The system is severely hampered by this class disparity.

2.3 Data Mining Methods

Data mining is the process of extracting patterns and models from existing data, i.e., prioritising the most critical data for future machine learning and artificial intelligence applications. The primary objective is to establish connections between variables in the data. The primary driver of its growth has been the growth of big data. Traditional data analysis techniques are no longer capable of dealing with such a large number of seemingly unconnected pieces of data. As a result, data mining method is required to discover relationships between diverse data and factors in order to improve data.

Data mining is basically the cornerstone of machine learning, with the primary goal of extracting supersets of information from a number of data sources and then combining the information to uncover previously unknown patterns and interrelationships. This indicates that

data mining is not a technique for establishing a hypothesis, but for developing a range of possibilities.

Table 2-1 Comparison of different data mining models[15]

Method	Applicability	Model Feature	Model Type	Learning Strategy	Loss Function	Learning Algorithm
Perceptron	Binary classification	Hyperplane separation	Discriminative	Minimize distance from mispoint to hyperplane	Distance from mispoint to hyperplane	Stochastic gradient descent
k-Nearest Neighbor	Multi-class classification / regression	Feature space	Discriminative	---	---	---
Naive Bayes	Multi-class classification	Joint probability distribution	Generative	Maximum posterior probability estimation	Log-likelihood loss	Probability calculation formula
Decision Tree	Multi-class classification / regression	Classification tree & Regression tree	Discriminative	Regularized maximum likelihood estimation	Log likelihood loss	Generating and pruning
Logistic Regression	Multi-class classification	Conditional probability distribution	Discriminative	Maximum likelihood estimation	Logistic loss	Gradient descent
Support Vector Machine	Binary classification	Hyperplane separation	Discriminative	Minimize regularized hinge loss	Hinge loss	Sequence minimum optimization
Boosting Method	Binary classification	Weak classifiers linear combination	Discriminative	Minimize additive model exponential loss	---	Forward stepwise addition

2.3.1 Ensemble Learning

There are three popular frameworks for ensemble learning: Bagging, Boosting, and Stacking[16].

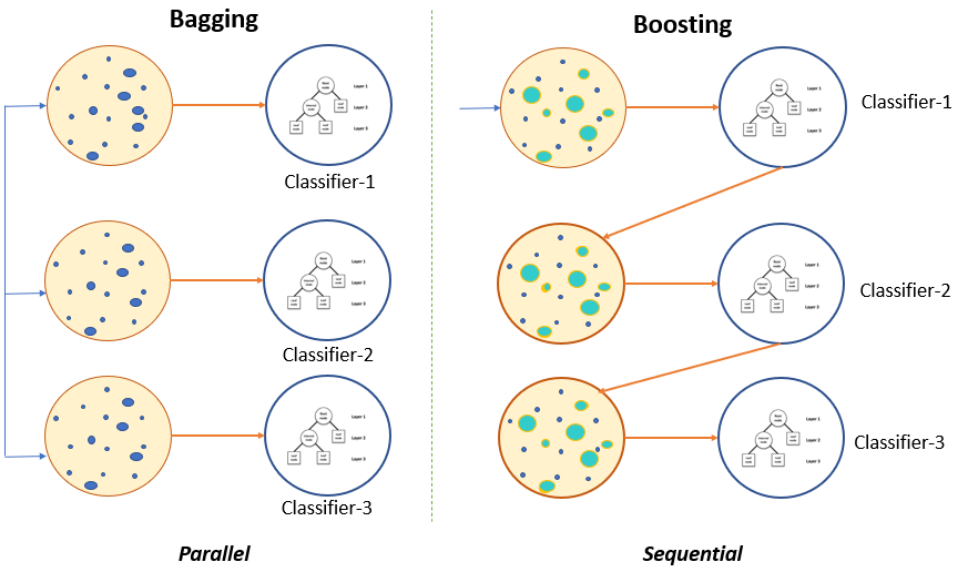


Figure 2-10 Bagging vs Boosting[16]

Bagging: The training set is extracted using the data from the original sample set. Each round, N training samples will be randomly selected from the original set using the Bootstrapping method (some samples may be drawn multiple times and some samples may not be drawn at all). To create k training sets, a total of k rounds were extracted. (each of the k training sets is self-contained). Each time a model is obtained, a single training set is employed, and k training sets are used to obtain k models. Voting was used to determine the categorisation results for the k models generated in the preceding phase. The ultimate outcome for regression issues is the mean value of the aforementioned model. (Each model is significant in its own right)

Boosting: The addition model combines the base model in a linear fashion. Each round of training raises the weight assigned to those basic models with a low mistake rate and decreases the weight assigned to those with a high error rate. By adjusting the weights or probability distributions of training data in each round, the weights of samples misclassified by the weak classifier in the previous round and the weights of samples misclassified in the previous round can be reduced, allowing the classifier to have a greater effect on the misclassified data.

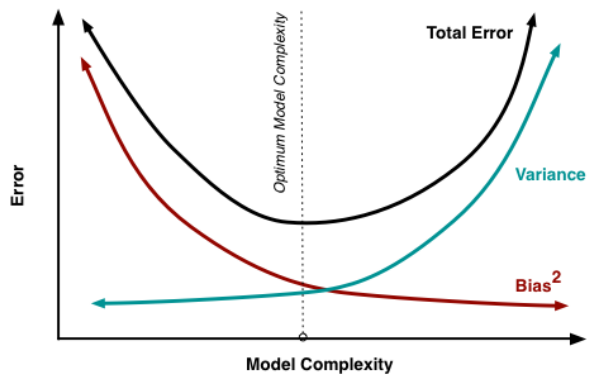


Figure 2-11 Bias vs Variance[17]

Because Bagging and Boosting basis models are both linear, so have:

The overall expectation:

$$\begin{aligned} E(F) &= E(\sum_{i}^{m} r_i f_i) \\ &= \sum_{i}^{m} r_i E(f_i) \end{aligned}$$

The overall variance:

$$\begin{aligned} Var(F) &= Var(\sum_{i}^{m} r_i f_i) \\ &= \sum_{i}^{m} Var(r_i f_i) + \sum_{i \neq j}^{m} Cov(r_i f_i, r_j f_j) \\ &= \sum_{i}^{m} r_i^2 Var(f_i) + \sum_{i \neq j}^{m} \rho r_i r_j \sqrt{Var(f_i)} \sqrt{Var(f_j)} \\ &= m r^2 \sigma^2 + m(m-1) \rho r^2 \sigma^2 \\ &= m r^2 \sigma^2 (1-\rho) + m^2 r^2 \sigma^2 \rho \end{aligned}$$

The model accuracy:

$$Error = bias^2 + var + \xi$$

For Bagging, each base model's weight is equal to 1/m and is assumed to be roughly equal, allowing us to obtain:

$$
\begin{aligned}
E(F) &= \sum_{i}^{m} r_i E(f_i) \\
&= m\frac{1}{m}\mu \\
&= \mu \\
Var(F) &= mr^2\sigma^2(1-\rho) + m^2r^2\sigma^2\rho \\
&= m\frac{1}{m^2}\sigma^2(1-\rho) + m^2\frac{1}{m^2}\sigma^2\rho \\
&= \frac{\sigma^2(1-\rho)}{m} + \sigma^2\rho
\end{aligned}
$$

The expectation of the global model is equal to the expectation of the base model, indicating that the global model's deviation is comparable to the base model's deviation. The variance of the entire model is less than or equal to the variance of the base model. As the number of base models increases, the variance of the overall model decreases, boosting the ability to prevent overfitting and enhancing the model's accuracy.

For Boosting: Due to the fact that the basis models share the same training set, there is a high degree of correlation between them, and the correlation coefficient between the models is nearly equal to one. According to Boosting, the following formula is used to calculate the reduction:

$$
\begin{aligned}
E(F) &= \sum_{i}^{m} r_i E(f_i) \\
Var(F) &= mr^2\sigma^2(1-\rho) + m^2r^2\sigma^2\rho \\
&= m\frac{1}{m^2}\sigma^2(1-1) + m^2\frac{1}{m^2}\sigma^2 1 \\
&= \sigma^2
\end{aligned}
$$

The total model's variance is equal to the base model's variance. If the base model is not a weak model, its variance will be very high, resulting in a large variance for the entire model, which will be unable to avoid overfitting. As a result, the Boosting framework's base model must be a weak model.

Additionally, since the boosting framework employs a greedy approach, the expectation of the overall model is accumulated by the expectation of the basis model, as the number of basis models grows, the anticipated value of the overall model increases, and the overall model's accuracy improves.

2.3.2 Random Forest

Random Forest (RF)[18]: a forest that has been generated randomly. The RF algorithm is made up of many decision trees, each of which is unrelated to the others. After the forest is established, and fresh samples are added, each decision tree will be evaluated independently, and the classification results will be determined using the voting method.

Random Forest is a more sophisticated version of Bagging. Random feature selection is further incorporated into the decision tree training process based on the Bagging, which is created using a decision tree-based learner.

2.3.3 Gradient Boosting Decision Tree

Gradient Boosting Decision Tree[19] is a multi-tree iterative Decision Tree method. Three ideas comprise the GBDT: Regression Decision Trees, Gradient Boosting (GB), and Shrinkage. Because the purpose of GBDT is to aggregate the results of all trees, the trees used in GBDT are regression trees rather than classification trees. Each tree in GBDT updates the target value using residuals from previous trees, so that the sum of the values in each tree equals the anticipated value of GBDT.

The model's anticipated value is as follows:

$$F_k(x) = \sum_{i=1}^{k} f_i(x)$$

Adopt a greedy approach: train the basic model sequentially. rephrase the model's general structure as iterative:

$$F_k(x) = F_{\{k-1\}}(x) + f_{\{k\}}(x)$$

Thus, each iteration should answer just one training issue for the basic model.

The residual represents the inverse gradient of the least mean square loss function applied to the anticipated value:

$$-\frac{\partial(\frac{1}{2}(y - F_k(x))^2)}{\partial F_k(x)} = y - F_k(x)$$

2.3.4 Support Vector Machine

A Support Vector Machine (SVM)[15] is a classification model that employs binary classification. Its fundamental model is a linear classifier specified in the biggest feature space, which differentiates it from perceptron. Additionally, the support vector machine incorporates the kernel trick, transforming it into a mostly nonlinear classifier. The support vector machine's learning approach is to maximize the interval, which may be described as a convex quadratic programming problem that is equal to reducing the regularized hinge loss function. The support vector machine learning method is the optimal approach for solving convex quadratic programming.

When the training data is separable linearly, a linear classifier is learnt by maximizing the hard interval, i.e., a linear separable support vector machine. When the training data is roughly separable linearly, a linear classifier, called the linear support vector machine, is also learnt by maximizing the soft interval. When the training data are linear and non-separable, a nonlinear support vector machine is trained using the kernel trick and soft interval maximization. The kernel method and trick involve implicitly learning a linear support vector machine in a high-dimensional feature space by using the kernel function. The maximal hard interval, defined as the greatest distance between the support point and the hyperplane, is converted into a constrained optimization problem, which is also a convex quadratic programming issue. Convert it to a Lagrange problem, and then solve $\boldsymbol{w}$ and $\boldsymbol{b}$ using the dual function under the KKT condition.

When the input space is either Euclidean space or a discrete set and the feature space is Hilbert space, the kernel function is the inner product of the feature vectors produced by mapping the input space to the feature space. Nonlinear support vector machines and linear support vector machines may be implicitly learnt in high-dimensional feature Spaces by using kernel functions.

2.4 Deep Learning Methods

2.4.1 Convolutional Neural Networks

Convolutional neural networks (CNN), which may seem like an odd fusion of computer science, biology, and mathematics, have become the most important advancement in computer vision. Neural networks made their debut in 2012, when Alex[20] used them to win the ImageNet

Challenge (approximately equal to the Olympic equivalent of computer vision of the year). He astounded the globe by reducing the categorisation error rate from 26% to 15%.

A more thorough explanation of how CNN works is as follows: you start with an image and run it through a succession of convolutional layers, non-linear layers, pooling (downsampling) layers, and fully connected layers.

The following figure illustrates the convolutional layer's functioning. To sweep the whole image, a convolutional kernel is used:

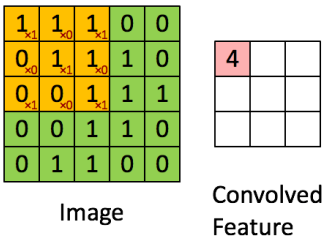


Figure 2-12 Convolutional kernel

This method may be viewed as filtering each tiny region of the picture with a filter (convolution kernel) in order to get the eigenvalues of these small areas. Multiple convolution cores are often used in particular applications. Each convolution kernel may be thought of as representing an image pattern. If an image block's value convolved with this convolution kernel is big, this image block is regarded to be extremely near to this convolution kernel.

Pooling is simple downsampling, which reduces the dimensionality of the data significantly. The rationale for this is because even after convolution, the picture is still very big (due to the convolution kernel's short size), and so subsampling is used to decrease the dimension of the data.

The last stage is the complete connection layer, which receives the data processed by the convolution and pooling layers. A typical CNN is not just a three-tier structure as mentioned above, but a multi-tier structure.

2.4.2 CNN for Image Classification

Classification of images is a critical job in image processing. In traditional machine learning, the typical method for identifying and classifying pictures includes feature extraction, feature screening, and ultimately, input of the feature vector to the appropriate classifier for feature

classification. Alex Krizhevsky achieved a breakthrough in 2012 by introducing the AlexNet network topology[20]. The three modules of image feature extraction, screening, and classification were combined using a deep learning algorithm, and a deep convolutional neural network structure with five convolutional layers and three full connection layers was designed to mine and extract image information in various directions layer by layer. For instance, shallow convolution often yields generic features such as picture edges, while deep convolution yields distribution characteristics unique to certain data sets. AlexNet won the 2012 ILSVRC annual championship with an error rate of 15.4 percent, a record low.

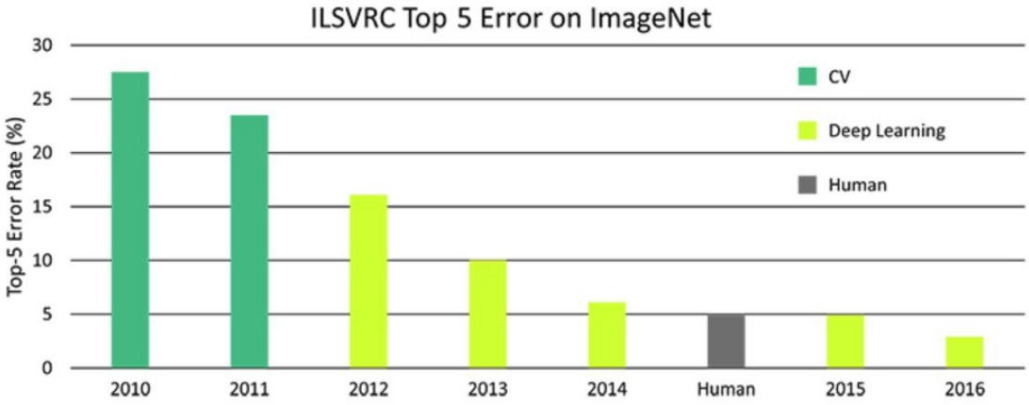


Figure 2-13 ImageNet Top-5 Classification Error Over Time[21]

GoogLeNet[22] and VGGNet[23] were the ILSVRC champion and runner-up networks in 2014. VGGNet is one of them, with two variants of layer 16 and layer 19, totalling 550M parameters. To simplify the construction of convolutional neural networks, 33 convolution kernels and 22 maximum pooling kernels are utilised. VGGNet is an excellent illustration of how you can boost a network's performance simply by adding layers and depth to the network on top of the prior network design. VGGNet is a simple but surprisingly effective benchmark model that is still used for a wide variety of jobs today.

The Inception Module is at the core of GoogLeNet, since it uses a parallel approach. Four components comprise a conventional inception structure. 1×1 convolution, 3×3 convolution, 5×5 convolution, 3×3 maximum pooling, and finally the results of the four component operations are combined on the channel. This is the fundamental concept behind the Inception Module. A more accurate picture representation may be produced by extracting information from the image's various scales and then merging it. Since then, deep learning algorithms have reached an accuracy level comparable to that of humans. In comparison to VGGNet, the

GoogleNet model architecture is deeper, more compact, and computationally efficient when combined with the well-designed Inception framework.

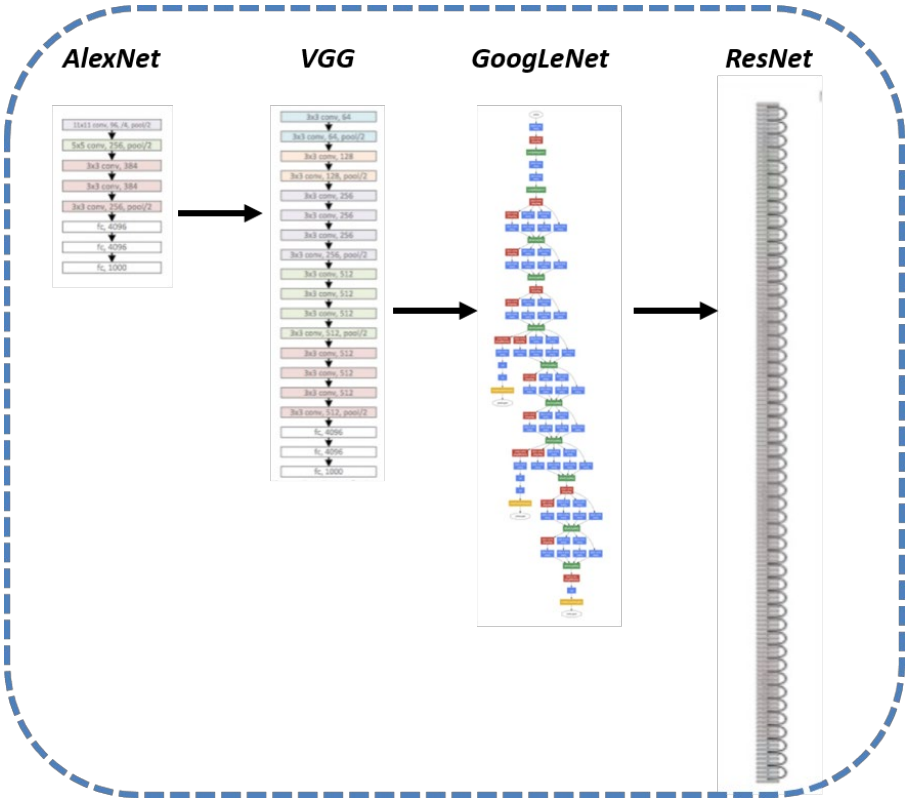


Figure 2-14 The development of neural network structures for classification

ResNet[24] is the next evolutionary step in this fast developing series. Historically, ResNet variations have been used as a reference model architecture for a variety of applications and as a foundation for more sophisticated designs. Apart from winning the ILSVRC 2015 Classification Challenge, ResNet is notable for its depth in comparison to other network designs. ResNet effectively addresses the issue of gradient dissipation in deep neural networks and enables the training of networks with thousands of layers.

Table 2-2 Common CNN architectures for classification

Model	Size(M)	Image-Net Top-1/top-5 error(%)	Layers	Model description
AlexNet	238	41.00/18.00	8	5 conv + 3 fc
VGG-16	540	28.07/9.33	16	13 conv + 3 fc
VGG-19	560	27.30/9.00	19	16 conv + 3 fc
GoogleNet	40	29.81/10.04	22	21 conv + 1 fc
ResNet-50	100	22.85/6.71	50	49 conv + 1 fc
ResNet-152	235	21.43/3.57	152	151 conv + 1 fc

2.5 Related Literatures

At the moment, no study has been conducted on the topic of using deep learning to predict the incidence and progression of myopia. Several publications emphasise the significance of myopia risk and risk factors[25]. Using feature analysis methods, Sara McCullough et al. established a Latent growth mixture model for the prediction of myopia in European children and young adults[26]. However, these studies highlight the significance of the features but do not provide a prediction model for specific patients. By examining refraction data, Haotian Lin et al. established a machine learning model that effectively predicted the onset of myopia in Chinese kids[27]. This technique, however, may suffer from over-fitting.

Numerous effective instances of fundus image-based diagnosis and interpretation have been shown, including the identification of glaucoma[28], cataracts[29], diabetic retinopathy[30], and macular degeneration[31]. Varadarajan et al. conducted the first and only study to date using deep learning and computer vision to investigate myopia.[32] Using the ResNet deep learning network, they extract the refractive error information from retinal fundus images. Additionally, Shanthi et al. enhance the Alexnet4 architecture for diabetic retinopathy picture categorisation. They established a successful precedent for the use of deep learning in ophthalmology, where its high accuracy and sensitivity aid in diagnosis.

The work in this thesis is based on Divya's research from BHVI[6]. The purpose of her study is to ascertain the relationship between myopia and presenting optic disc and retinal features, as well as to develop models incorporating optic disc and retinal features for discriminating myopes from non-myopes, predicting the incidence of myopia, and predicting a myopic shift in refractive error. Divya extracted semi-automatic features from fundus pictures and manually inspected the derived features with statistical analysis methods. Finally, Multiple logistic regression analysis was performed to identify myopia and forecast its occurrence and development in young children.

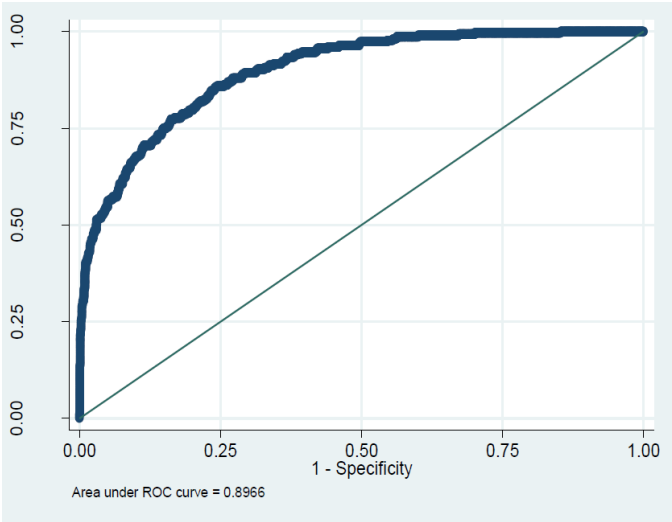


Figure 2-15 ROC curve for the predictive model to detection of myopia with LR model[6]

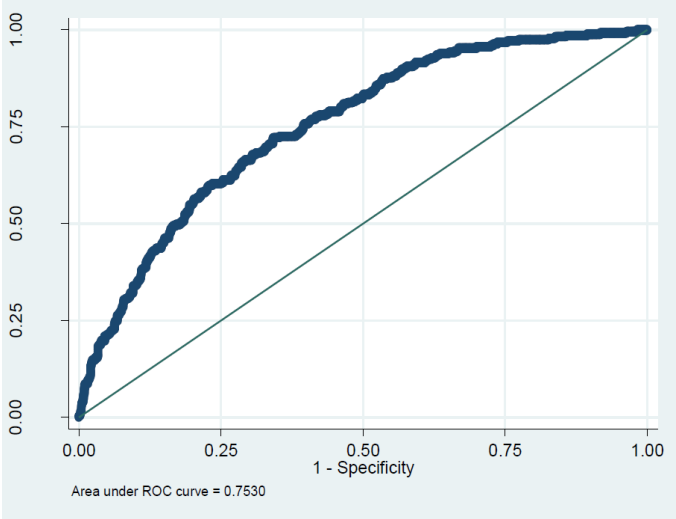


Figure 2-16 ROC curve for predictive model for incidence of myopia with LR model[6]

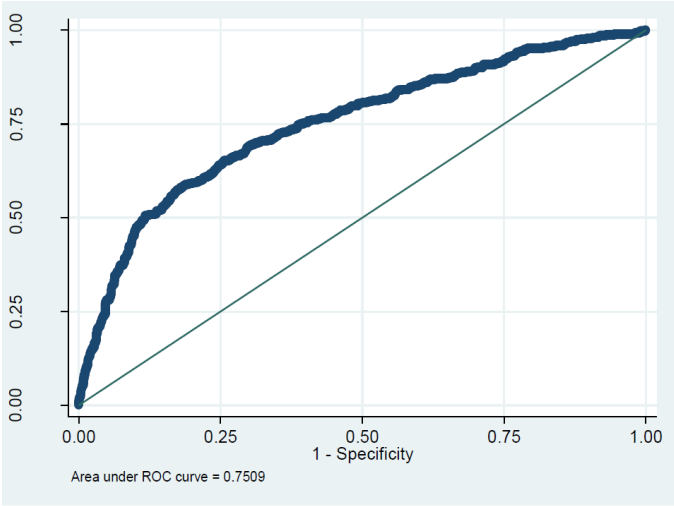


Figure 2-17 ROC curve for the predictive model for the progression of myopia with LR model[6]

The ROC curves for the predictive model for the detection, incidence, and progression of myopia using the LR model are shown in Figures 2-15, 2-16, and 2-17, respectively. In Divya's research, the AUC for detecting adolescent myopia was 0.89, the AUC for predicting myopia incidence was 0.75, and the AUC for predicting myopia progression was 0.75. However, Divya's work has a lot of room for improvement. To begin, manually extracting features takes a significant amount of time and is also quite tedious. Then, because the result is not validated via cross-validation, it is over-fitting. However, given that the model is LR, over-fitting may not be a significant issue.

As a result, this research uses machine learning to make use of image data that have not before been utilised to predict myopia. Additionally, via multi-modal data analysis, there is a possibility of developing a model with a high degree of resilience and accuracy for predicting the start and progression of myopia.

Chapter 3. Data Mining Models for Myopia Detection, Incidence and Progression Prediction

In this chapter, we examine the extracted features from retinal images and ophthalmic data using data mining methods. Because the axial length to corneal radius ratio is a critical feature in myopia prediction, this chapter does data mining to analyse data that have and do not contain this feature.

This chapter will introduce the methods of data mining, and show the results of myopia detection, incidence and progression prediction models with and without axial length to corneal radius ratio feature respectively. Finally, the results are summarized and discussed.

3.1 Methods

3.1.1 Data Mining Models

With the rise in popularity of machine learning in engineering and scientific research, different models and approaches have found their optimal application sectors, including image identification, natural language processing, and other disciplines. LGBM, Random Forest, and SVM are the most often employed models in the field of classification models.

Table 3-1 Data Mining Models for datasets and tasks in this study

Models \Training data set	Extracted features	Extracted features + ophthalmic data(intersection)
LGBM	Myope Detection	Myope Detection
	Incidence Prediction	Incidence Prediction
	Progression Prediction	Progression Prediction
RF	Myope Detection	Myope Detection
	Incidence Prediction	Incidence Prediction
	Progression Prediction	Progression Prediction
SVM(SVC)	Myope Detection	Myope Detection
	Incidence Prediction	Incidence Prediction
	Progression Prediction	Progression Prediction

The data mining model method used in this study, as well as the corresponding data sets and classification tasks, are listed in Table 3-1. Due to the lack of ophthalmic data captured by the optical coherence tomography system's machine, only 2169 participants' data were obtained following the intersection.

Two stages can be distinguished in the optimisation and tweaking of model parameters. The first are the so-called model parameters, which include the tree's depth and breadth, and the second are the hyper-parameters, which include the characteristic dimension. To optimise the parameter space, lattice search, random search, and heuristic search can be utilised. In this study Bayesian Optimization Algorithm will be implemented for parameter search. Figure 3-1 shows the parameter search with Bayesian optimization algorithm.

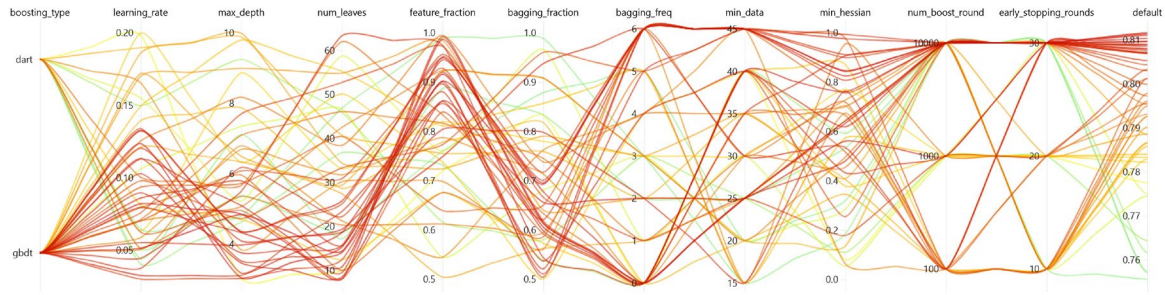


Figure 3-1 Hyper-parameter search with Bayesian Optimization Algorithm

3.1.2 Model Evaluation Criteria

3.1.2.1 ROC Curves and AUC

This project chooses the cross-validation method for off-line evaluation. The model evaluation method for the ensemble learning classification problem is identical to that for the common classification problem. The confusion matrix is frequently used to construct the dichotomy model.

Table 3-2 Confusion Matrix

Predicted Class \ Actual Class	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Among them, the TP (True Positive) is an example that is both true and predicted to be true. FN (False Negative): a positive example is classified incorrectly as a negative example. FP (False Positive): This is a true negative example that has been classified incorrectly as a positive example. The TN (True Negative) is, in fact, a negative example that is correctly classified as such.

The Receiver Operating Characteristic Curve (ROC) and the Area Under the Curve (AUC) are both popular and practical methods. The ROC curve has a horizontal coordinate of FPR (false positive rate) and a vertical coordinate of TPR (true positive rate). True positive rate (TPR) is TP / (TP + FN) and false positive rate (FPR) is FP / (FP + TN). A group of FPR and TPR point pairs for different classification thresholds can be given in the space formed by it, and the point pairs connecting all classification thresholds form the ROC curve, which is typically used to reflect the model's classification accuracy on the data set. The further to the right the ROC curve is, the higher the false positive rate, indicating that the predicted positive examples contain a greater number of actual negative samples. The greater the area under the receiver operating characteristic curve, the greater the true rate, implying that there are more actual positive samples among the predicted positive examples. The smoother the ROC curve, the greater the degree of generalisation; the closer it is to the upper left, the greater the classification accuracy. In other words, the larger the AUC under the ROC, the better.

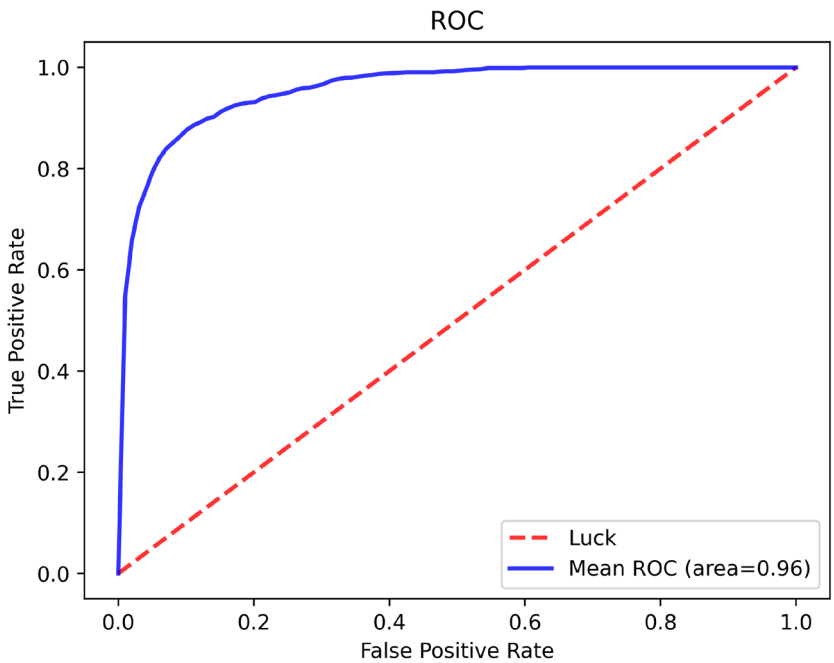


Figure 3-2 An ROC Curve example

3.1.2.2 t-test

The Independent Samples t-test analyses the means of two independent groups to see if statistical evidence exists that the related population means differ substantially. A parametric test is the Independent Samples t-test. The AUC values of multiple models within the same data set will be compared using the Independent Samples T-test to determine the results' statistical significance.

3.2 Results

The growth in vitreous chamber depth in children differed considerably between myopic and emmetropic eyes. Prior to the start of myopia, these disparities were more noticeable. The sensitivity and specificity of predicting the start of myopia can be increased marginally by using the axial length/corneal radius of curvature (AL/CR ratio) as possible predictive markers. In this study, the results will be presented with and without this feature[33].

3.2.1 Results without Axial Length to Corneal Radius Ratio

3.2.1.1 AUC Results

Table 3-3 below shows the AUC values without regard for the feature axial length to corneal radius ratio. As shown in the table, the LGBM model produces the best results when combined with the extracted features and ophthalmic data. The results for Detection, Incidence, and Progression are 0.89, 0.71, and 0.70, respectively. The LR in the table represents the result of the Divya[6] thesis' retest using the same method. The results of LGBM model in Detection task are slightly better, while the results of Incidence and Progression task are slightly worse.

Table 3-3 AUC results of each model on datasets on three tasks

AUC	Detection		Incidence		Progression	
Training Methods / Data Set	Extracted features [37]	Extracted features + ophthalmic data(intersection) [226]	Extracted features [37]	Extracted features + ophthalmic data(intersection) [226]	Extracted features [37]	Extracted features + ophthalmic data(intersection) [226]
LGBM	0.88	0.89	0.71	0.71	0.70	0.70
RF	0.86	0.86	0.69	0.69	0.69	0.68
SVM	0.88	0.85	0.71	0.69	0.70	0.65
LR	N/A	0.88	N/A	0.74	N/A	0.71

3.2.1.2 ROC Curves

Detection:

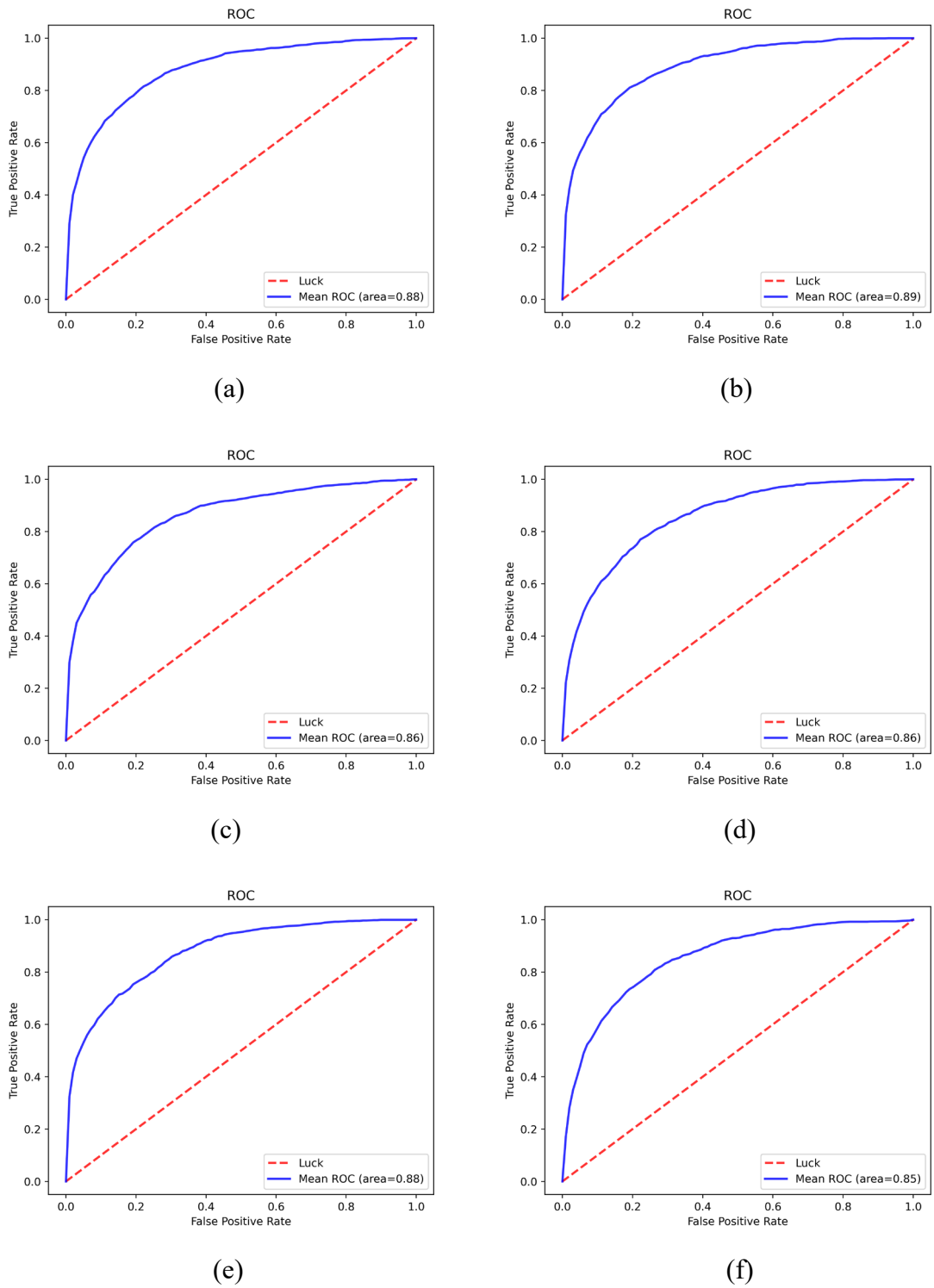


Figure 3-3 ROC curves of each detection model on datasets

(a: LGBM model on extracted features dataset, b: LGBM model on extracted features + ophthalmic dataset, c: RF model on extracted features dataset, d: RF model on extracted features + ophthalmic dataset, e: SVM model on extracted features dataset, f: SVM model on extracted features + ophthalmic dataset)

Incidence:

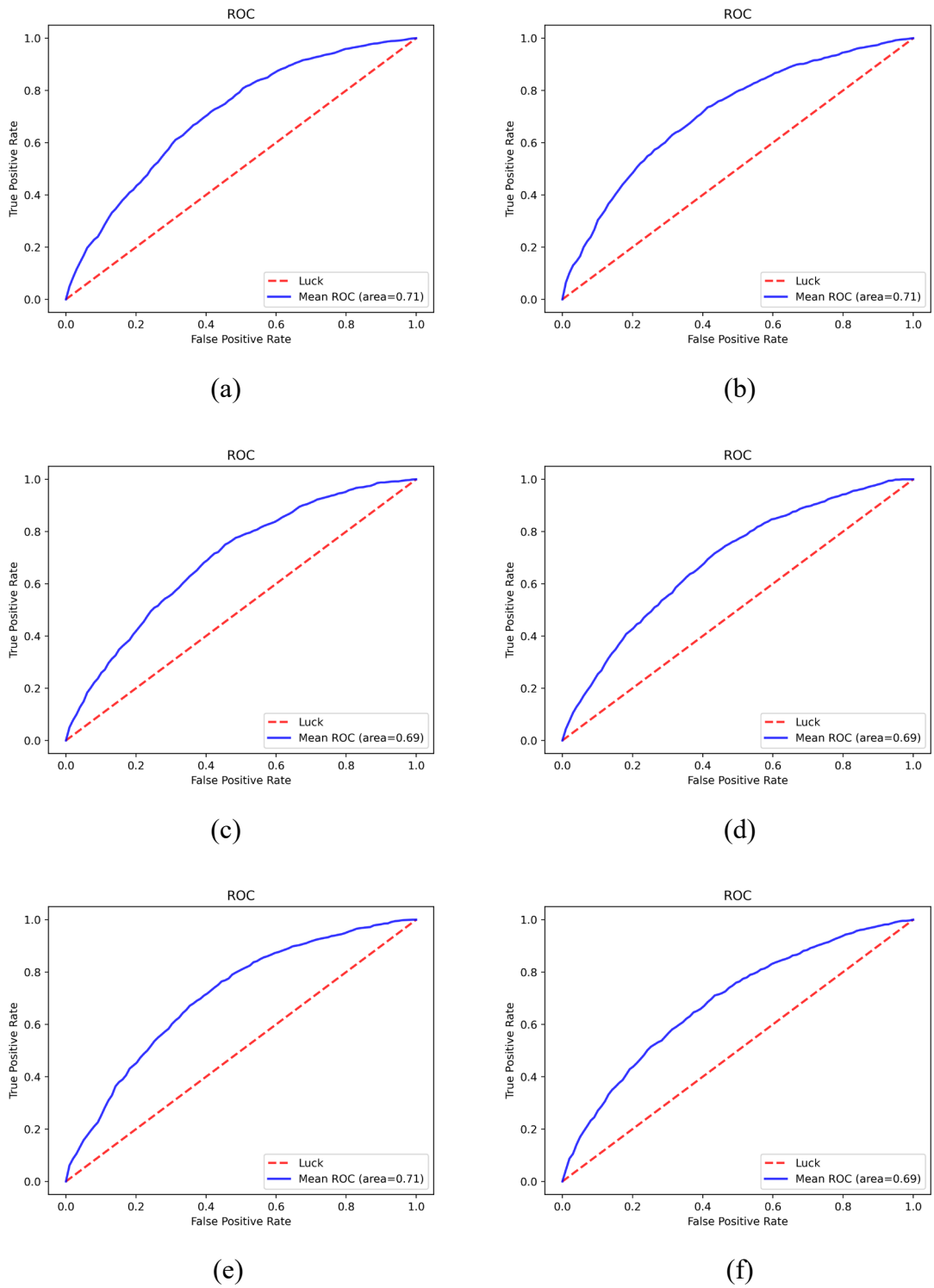


Figure 3-4 ROC curves of each incidence prediction model on datasets

(a: LGBM model on extracted features dataset, b: LGBM model on extracted features + ophthalmic dataset, c: RF model on extracted features dataset, d: RF model on extracted features + ophthalmic dataset, e: SVM model on extracted features dataset, f: SVM model on extracted features + ophthalmic dataset)

Progression:

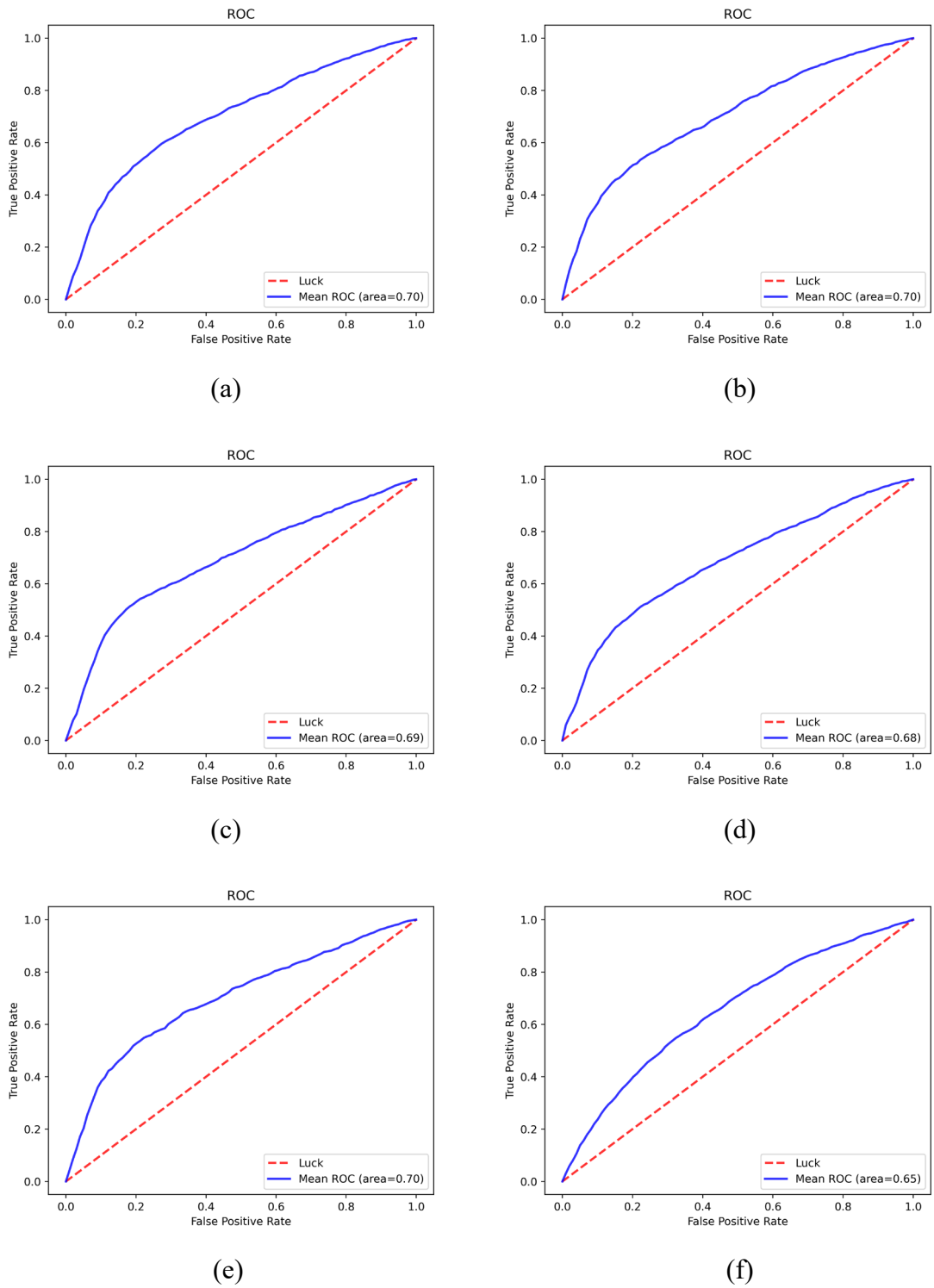


Figure 3-5 ROC curves of each progression prediction model on datasets

(a: LGBM model on extracted features dataset, b: LGBM model on extracted features + ophthalmic dataset, c: RF model on extracted features dataset, d: RF model on extracted features + ophthalmic dataset, e: SVM model on extracted features dataset, f: SVM model on extracted features + ophthalmic dataset)

Figure 3-3,3-4 and 3-5 respectively show ROC curves of Detection, Incidence and Progression tasks without regard for the feature axial length to corneal radius ratio. a: LGBM model on extracted features dataset, b: LGBM model on extracted features + ophthalmic dataset, c: RF model on extracted features dataset, d: RF model on extracted features + ophthalmic dataset, e: SVM model on extracted features dataset, f: SVM model on extracted features + ophthalmic dataset.

3.2.1.3 t-test on AUC Results

The results of the pair-wise independent samples t-test for detection, incidence, and progression on AUC results are shown in Table 3-4,3-5, and 3-6, respectively. The results indicate that the LGBM model's output is statistically significantly different from the output of the other models in the three tasks.

Table 3-4 Pair-wised Independent Samples t-test on Detection Results

Group One	Group Two	p value	$p < 0.05$
LGBM	LR	0.0498	True
LGBM	RF	0.001	True
LGBM	SVM	0.001	True
LR	RF	0.0063	True
LR	SVM	0.0011	True
RF	SVM	0.9	False

Table 3-5 Pair-wised Independent Samples t-test on Incidence Prediction Results

Group One	Group Two	p value	$p < 0.05$
LGBM	LR	0.0025	True
LGBM	RF	0.0142	True
LGBM	SVM	0.0078	True
LR	RF	0.001	True
LR	SVM	0.001	True
RF	SVM	0.9	False

Table 3-6 Pair-wised Independent Samples t-test on Progression Prediction Results

Group One	Group Two	p value	$p < 0.05$
LGBM	LR	0.434	True
LGBM	RF	0.058	True
LGBM	SVM	0.001	True
LR	RF	0.001	True
LR	SVM	0.001	True
RF	SVM	0.0033	True

3.2.2 Results with Axial Length to Corneal Radius Ratio

3.2.2.1 AUC Results

Table 3-7 below shows the AUC values with regard for the feature axial length to corneal radius ratio. As shown in the table, the LGBM model produces the best results. The results for Detection, Incidence, and Progression are 0.96, 0.81, and 0.74, respectively. In comparison to features without this feature, the AUC of the LGBM model increases by 7%, 10%, and 14% in the detection, incidence, and progression tasks, respectively.

Table 3-7 AUC results of each model on datasets on three tasks

AUC	Detection		Incidence		Progression	
Training Methods / Data Set	Extracted features [37]	Extracted features + ophthalmic data(intersection) [226]	Extracted features [37]	Extracted features + ophthalmic data(intersection) [226]	Extracted features [37]	Extracted features + ophthalmic data(intersection) [226]
LGBM	0.96	0.96	0.81	0.81	0.74	0.74
RF	0.94	0.92	0.78	0.77	0.72	0.71
SVM	0.92	0.88	0.75	0.70	0.70	0.66

3.2.2.2 ROC Curves

Detection:

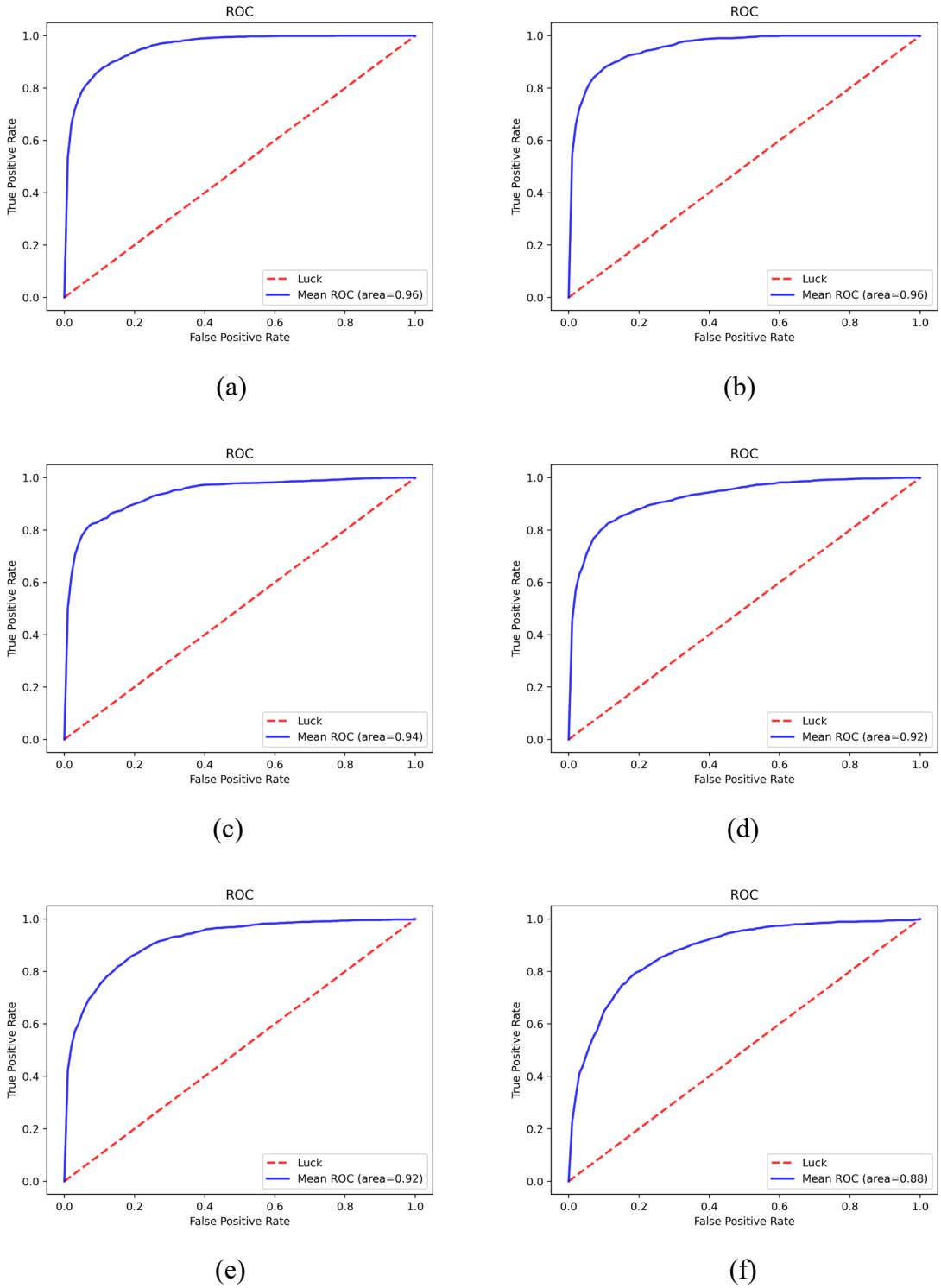


Figure 3-6 ROC curves of each detection model on datasets

(a: LGBM model on extracted features dataset, b: LGBM model on extracted features + ophthalmic dataset, c: RF model on extracted features dataset, d: RF model on extracted features + ophthalmic dataset, e: SVM model on extracted features dataset, f: SVM model on extracted features + ophthalmic dataset)

Incidence:

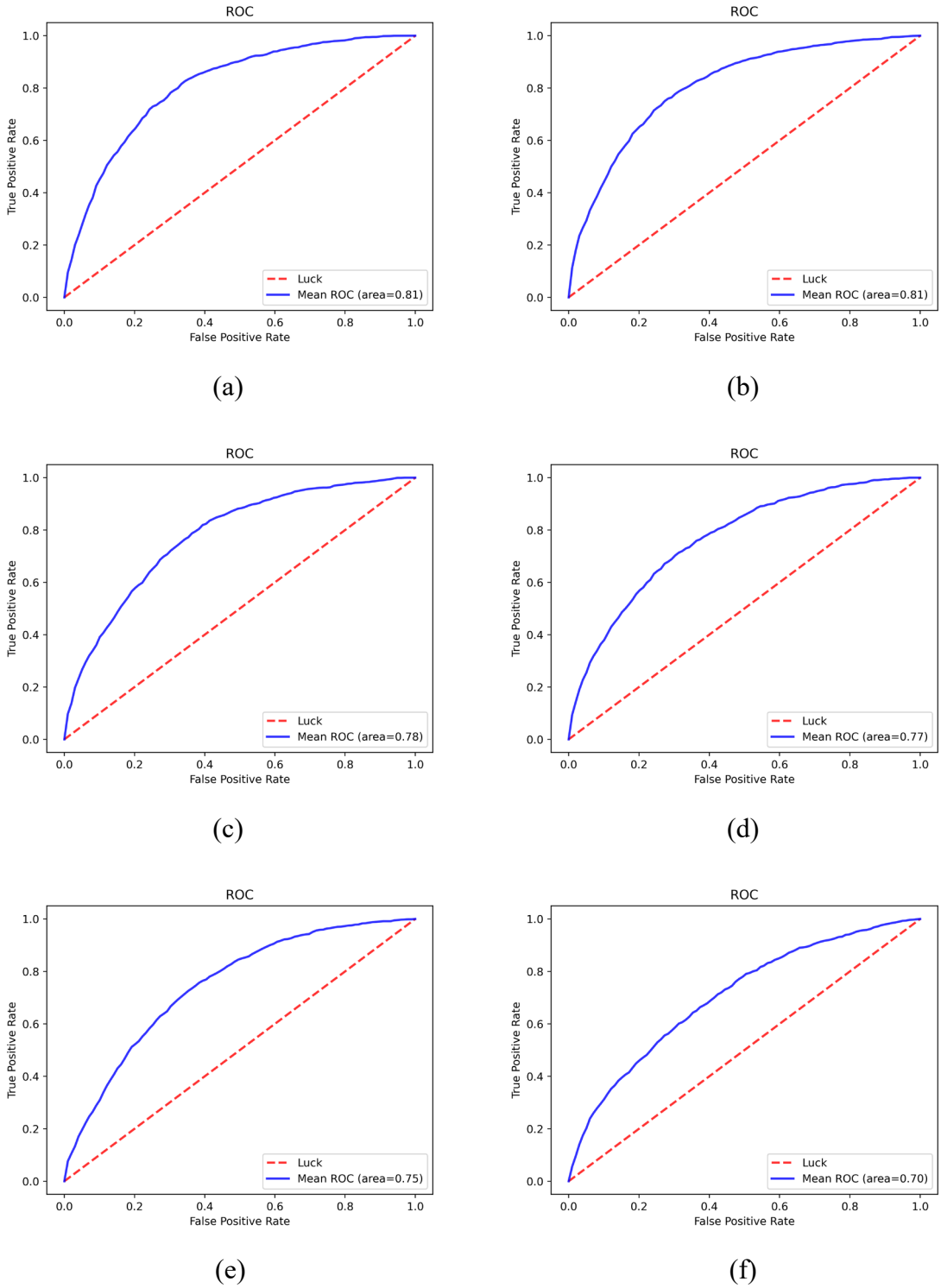


Figure 3-7 ROC curves of each incidence prediction model on datasets

(a: LGBM model on extracted features dataset, b: LGBM model on extracted features + ophthalmic dataset, c: RF model on extracted features dataset, d: RF model on extracted features + ophthalmic dataset, e: SVM model on extracted features dataset, f: SVM model on extracted features + ophthalmic dataset)

Progression:

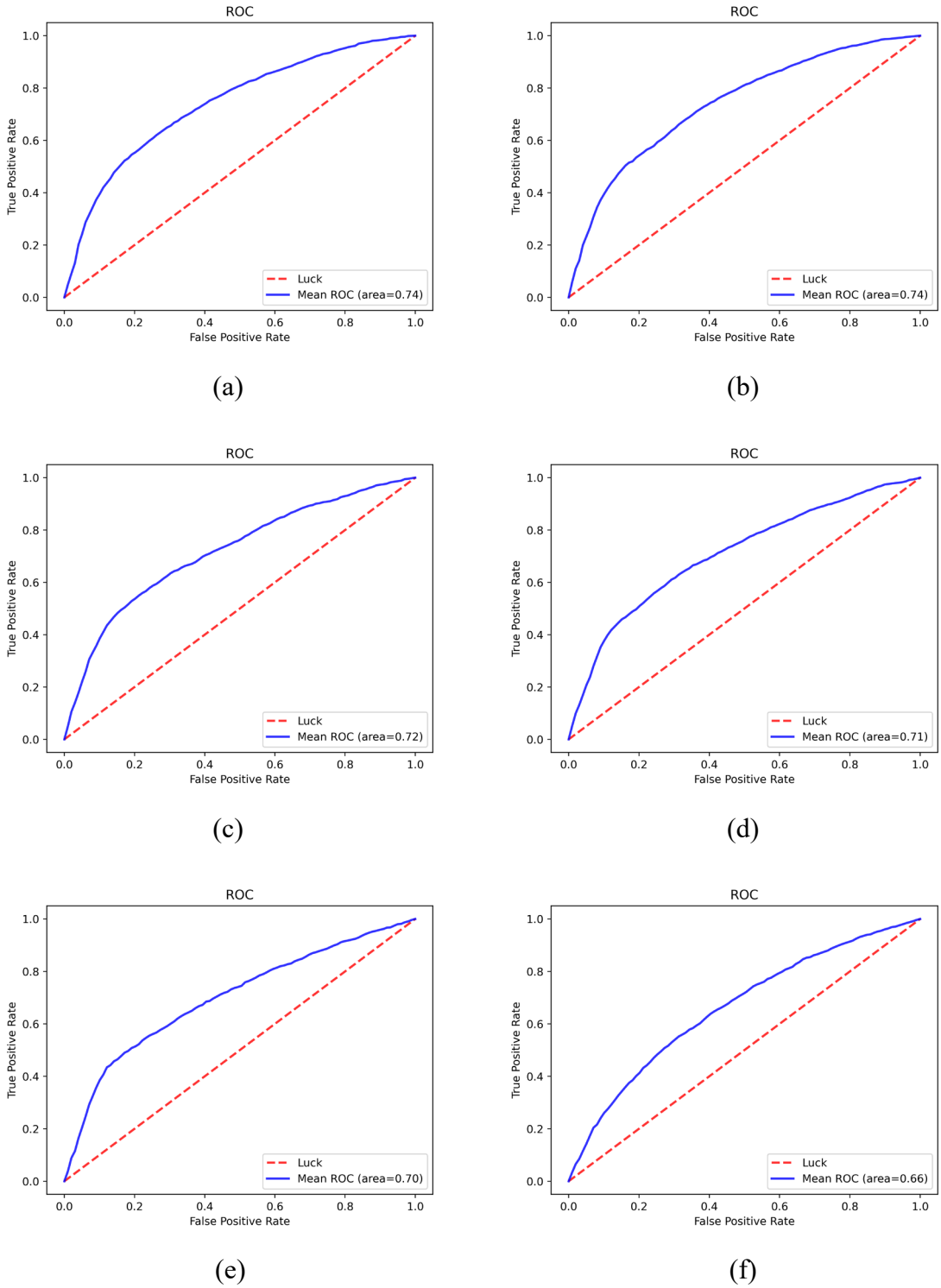


Figure 3-8 ROC curves of each progression prediction model on datasets

(a: LGBM model on extracted features dataset, b: LGBM model on extracted features + ophthalmic dataset, c: RF model on extracted features dataset, d: RF model on extracted features + ophthalmic dataset, e: SVM model on extracted features dataset, f: SVM model on extracted features + ophthalmic dataset)

Figure3-6, 3-7 and 3-8 respectively show ROC curves of Detection, Incidence and Progression tasks with regard for the feature axial length to corneal radius ratio. a: LGBM model on extracted features dataset, b: LGBM model on extracted features + ophthalmic dataset, c: RF model on extracted features dataset, d: RF model on extracted features + ophthalmic dataset, e: SVM model on extracted features dataset, f: SVM model on extracted features + ophthalmic dataset.

3.2.2.3 t-test on AUC Results

The results of the pair-wise independent samples t-test for detection, incidence, and progression on AUC results are shown in Table 3-4,3-5, and 3-6, respectively. The results indicate that all model's output is statistically significantly different from the output of the other models in the three tasks.

Table 3-8 Pair-wised Independent Samples t-test on Detection Results

Group One	Group Two	p value	$p < 0.05$
LGBM	RF	0.001	True
LGBM	SVM	0.001	True
RF	SVM	0.001	True

Table 3-9 Pair-wised Independent Samples t-test on Incidence Prediction Results

Group One	Group Two	p value	$p < 0.05$
LGBM	RF	0.001	True
LGBM	SVM	0.001	True
RF	SVM	0.001	True

Table 3-10 Pair-wised Independent Samples t-test on Progression Prediction Results

Group One	Group Two	p value	$p < 0.05$
LGBM	RF	0.001	True
LGBM	SVM	0.001	True
RF	SVM	0.001	True

3.3 Summary & Discussion

3.3.1 Summary

In this chapter, we examine the extracted features from retinal images and ophthalmic data using a data mining. Because the axial length to corneal radius ratio is a critical feature in ophthalmic data, this chapter does data mining to analyse data that have and do not contain this feature.

Without regard for the axial length to corneal radius ratio, the following data are available: When paired with the extracted features and ophthalmic data, the LGBM model delivers the best results. Detection, Incidence, and Progression all had values of 0.89, 0.71, and 0.70, respectively. The LR value in the table is the result of a retest of the Divya[6] thesis using the same approach. As can be observed, the LGBM model performs marginally better in the Detection task while slightly worse in the Incidence and Progression tasks than Logistic Regression models. The results reveal that the output of the LGBM model is statistically substantially different than the output of the other models in all three tasks.

The following summarises the findings of data analysis using the feature axial length to corneal radius ratio: The LGBM model yields the most accurate results. Detection, Incidence, and Progression all had values of 0.96, 0.81, and 0.74, respectively. In comparison to results that do not include this feature, the LGBM model's AUC increases by 7%, 10%, and 14% in the detection, incidence, and progression tasks, respectively. The pair-wise independent samples t-test findings for detection, incidence, and progression on the AUC values suggest that the output of all models is statistically substantially different from the output of the other models in all three tasks.

3.3.2 Discussion

The following conclusions can be drawn from the results obtained from the experiment:

1. The results of the LGBM model are significant different from the other methods. Obviously, due to different loss functions of different models and different learning algorithms, the training model converges to different vertices for loss minimization under the same sample space. As can be seen, the outcomes of each model are distinct.
2. The data mining model performs marginally better in the myopia detection task than the Logistic Regression model, but slightly worse in the myopia incidence and

progression prediction tasks. Notably, the data mining model can eliminate the need for tedious manual feature selection.

3. In comparison to results that do not include the feature axial length to corneal radius ratio, the LGBM model's AUC increases by 7%, 10%, and 14% in the detection, incidence, and progression prediction tasks, respectively. This is consistent with the results of clinical research.

Chapter 4. Deep Learning Models for Myopia Detection, Incidence and Progression Prediction

In this chapter, two neural network models with different structures were used to train on the retinal image dataset then optimized. The activation images of the Resnet model on three tasks were provided for clinical research. The EfficientNet model were modified and trained for the multi-modal data of image and clinical data. We also train the model with and without feature axial length to corneal radius ratio in clinical dataset.

This chapter will introduce the dataset and methods of deep learning. Then we will show the results of Resnet and EfficientNet-B4 models on the retinal image dataset of myopia detection, incidence and progression prediction. We will also provide the results of EfficientNet-B4 model on different datasets of myopia detection, incidence and progression prediction tasks. Finally, the results are summarized and discussed.

4.1 Methods

4.1.1 Data Augmentation

In general, data augmentation methods[34], such as spatial deformation, random conversion, rotation at a certain angle, reversal, and cropping, may all be utilized to address the issue of insufficient training data and can be used to restricted biomedical picture training sets. By including a number of data augmentation methods into the training data set, the deep neural network may get additional training data for parameter training and learning, as well as more robust picture characteristics. As a result, data augmentation is critical while training a deep neural network, particularly when the training data is inadequate.

4.1.1.1 Geometric Transformation

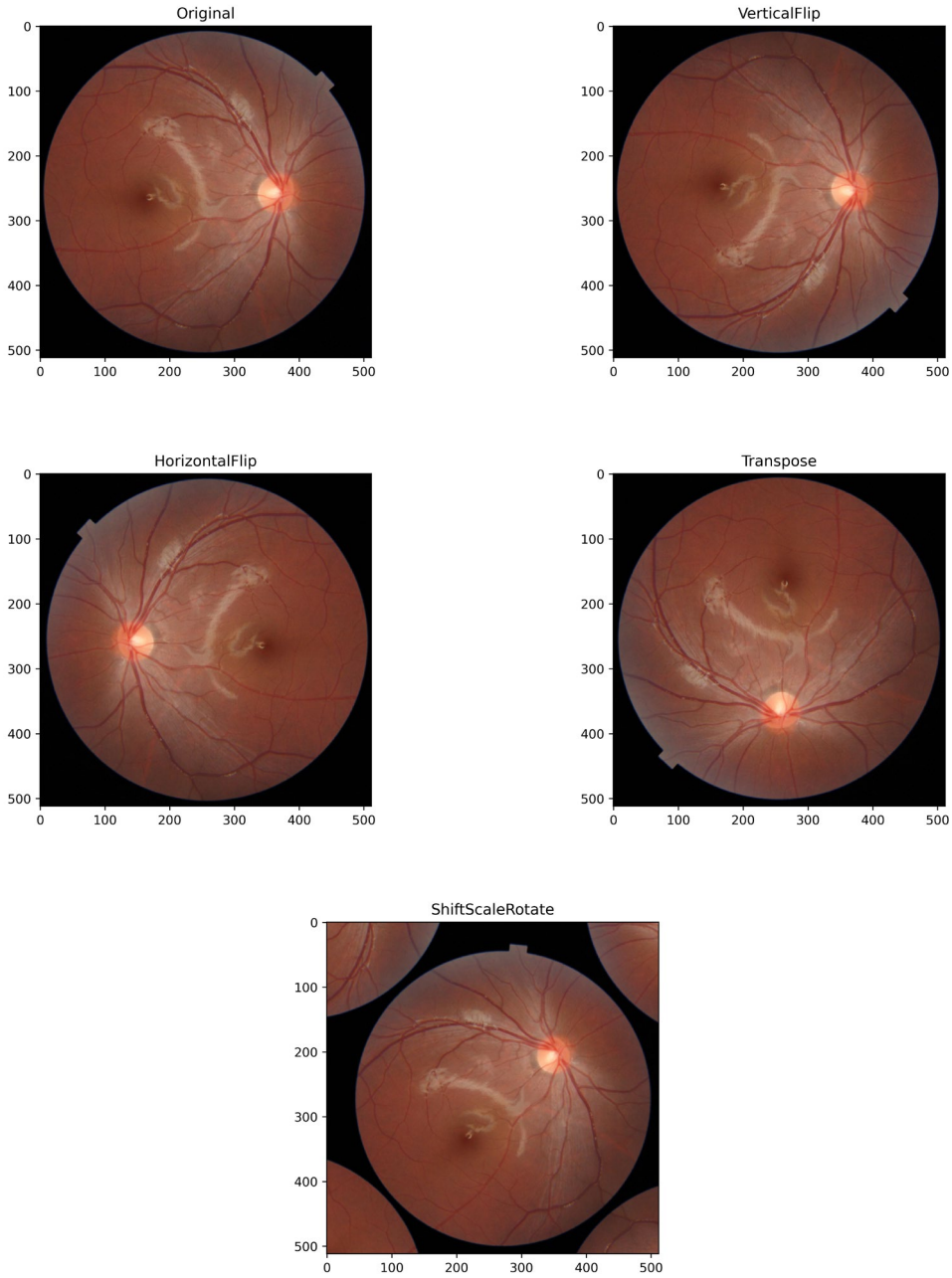


Figure 4-1 Retinal Images Geometric Transformation

[These images belong to BHVI]

Vertical flip, Horizontal flip, Transpose and Shift scale rotation are implemented in this study, shown above in Figure 4-1.

4.1.1.2 Colour Transformation

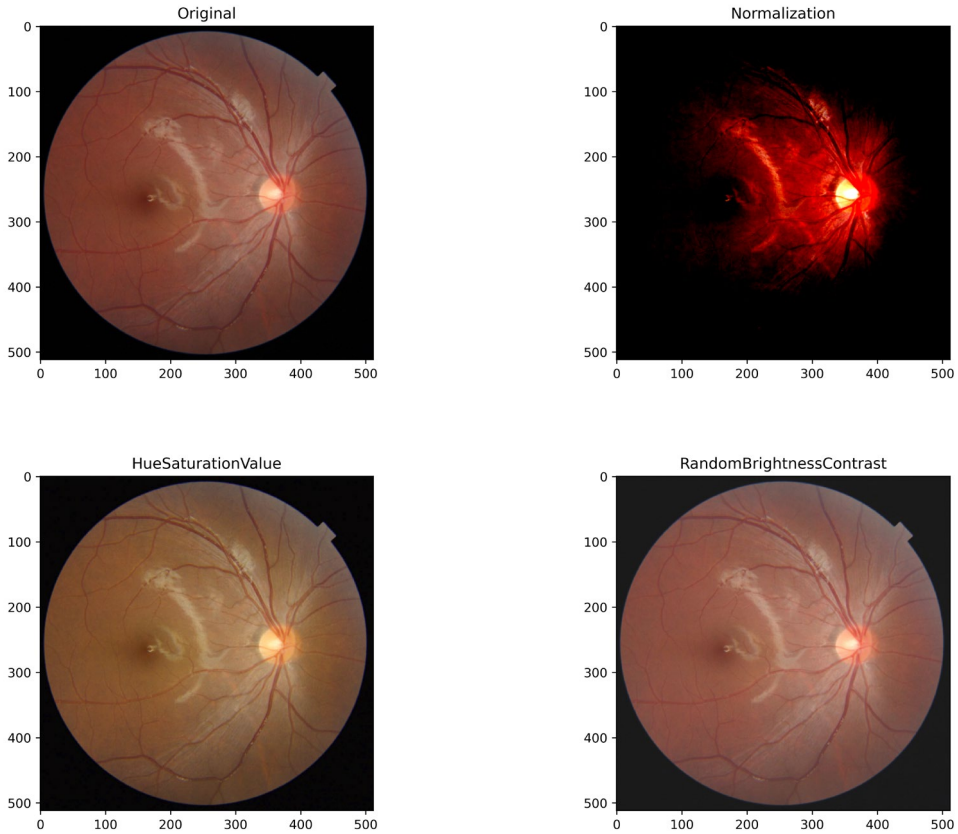


Figure 4-2 Retinal Images Colour Transformation

[These images belong to BHVI]

Because the Image-Net pre-training model is used for transfer learning, the image normalization parameters are identical to those used in Image-Net. Hue-Saturation Value change and Random brightness contrast change were implemented in this study as shown above in Figure 4-2.

4.1.1.3 Cut and Mix

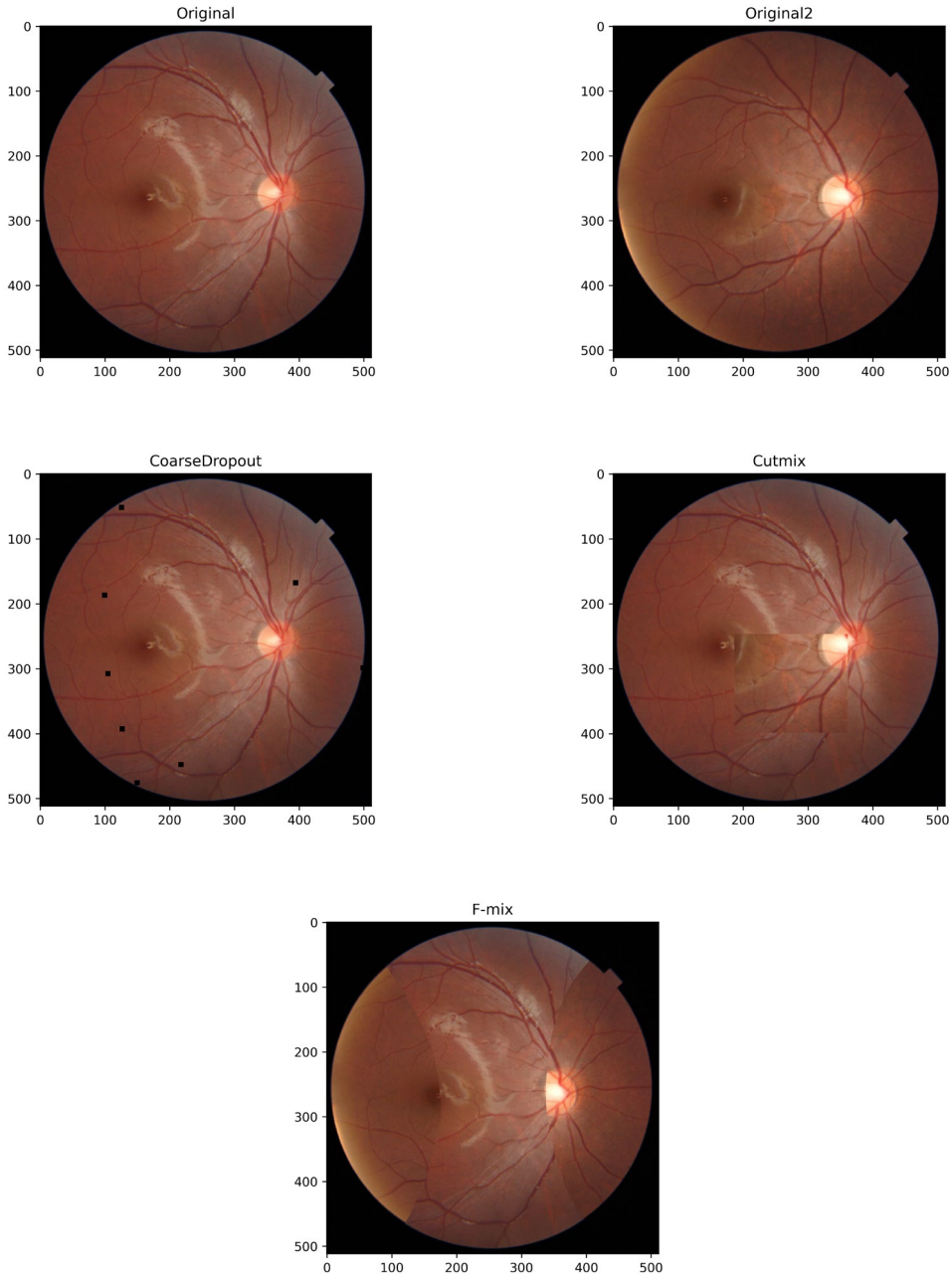


Figure 4-3 Retinal Images cut and mix operation.

[These images belong to BHVI]

Coarse dropout, Cut-mix and F-mix were used in this study, presented in the Figure 4-3 above.

4.1.2 Model Design

4.1.2.1 Model Architecture of Resnet

We have observed an increase in the number of layers in the design and improved performance over the last several CNNs. However, as network depth increases, accuracy becomes saturated and subsequently quickly decreases[24]. Microsoft Research researchers solved this issue with ResNet by using skip connections (shortcut connections, residuals) to construct deeper models[35]. These layers repeat activations from the preceding layer until the subsequent layer learns its weights. ResNet models often begin with a double-layer skip followed by ReLU activation and batch normalization. Each ResNet model consists of five convolutional layers, one fully connected layer, and a softmax layer. Figure 4-4 illustrates the design of the ResNet models.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

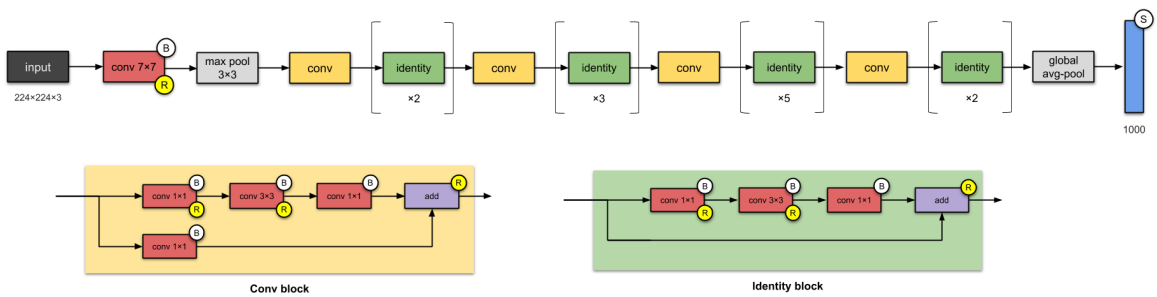


Figure 4-4 ResNet architectures(above) and ResNet 50 architecture (below)[36]

4.1.2.2 Model Architecture of EfficientNet-B4

Convolutional Neural Networks (ConvNets) are usually created with a set resource budget and then scaled up for increased accuracy using more resources, such as increasing network depth, network breadth, and input picture resolution. However, manually adjusting the depth, breadth, and resolution is very tough. When the quantity of calculation available is restricted, it is tough to decide which one to expand and which one to decrease. In other words, such a combination space is unfathomably vast. On the basis of the above, EfficientNet[37] offers a novel model scaling technique that scales the network based on its depth, breadth, and resolution using a simple and efficient compound coefficient. It will not randomly scale the network's dimensions, as the conventional approach does. A neural structure search method may be used to determine the optimum collection of parameters (compound coefficients). As shown in the Figure below, EfficientNet is considerably quicker and more accurate than other networks.

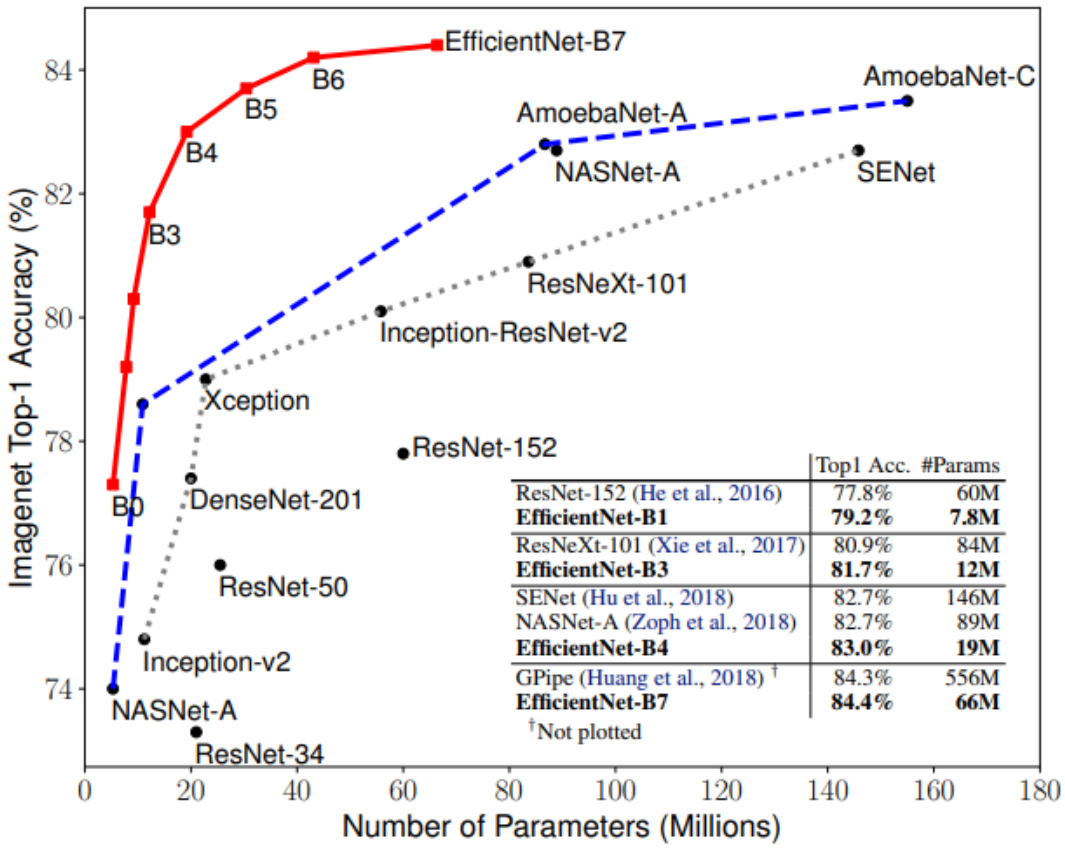


Figure 4-5 Model Size Vs ImageNet accuracy[37]

Any network's stem is the starting point for any experimentation with the design, which is common to all eight EfficientNet models and the final layers.

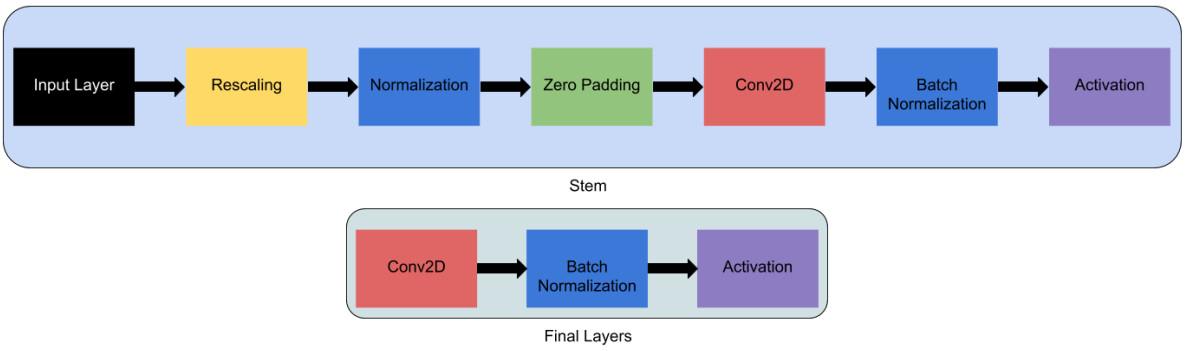


Figure 4-6 Stem and Final Layers in EfficientNet[38]

Following that, each of them has seven blocks. Additionally, these blocks include a variable number of sub-blocks, the number of which increases as we go from EfficientNetB0 to EfficientNetB7.

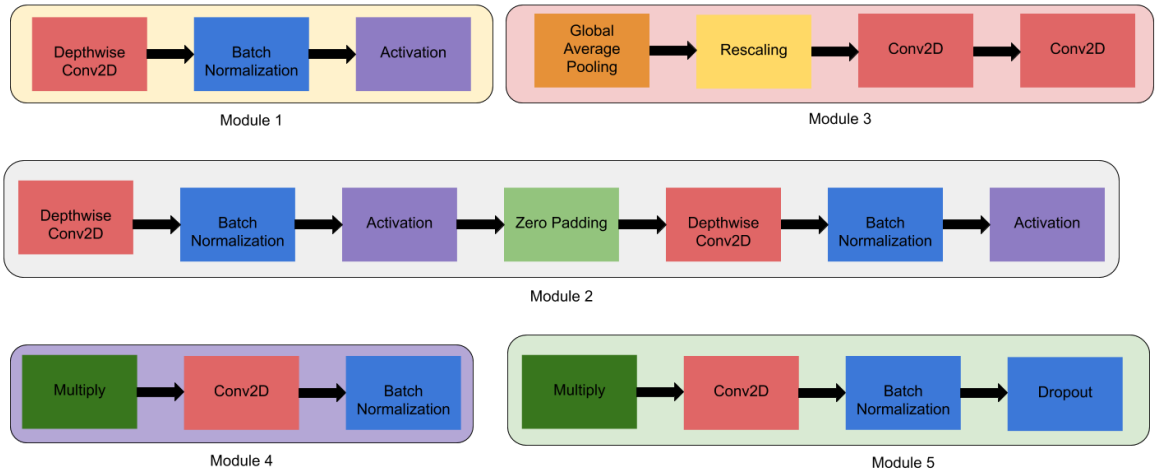


Figure 4-7 Five modules for architecture construction[38]

Module 1 serves as a jumping-off point for the subsequent sub-blocks. Module 2 serves as a jumping-off point for the first sub-block of each of the seven major blocks save the first. Module 3 is linked to all the sub-blocks through a skip connection. Module 4 is where the skip connection from the first sub-blocks is combined. Module 5 connects each sub-block to its preceding sub-block through a skip link and then combines them.

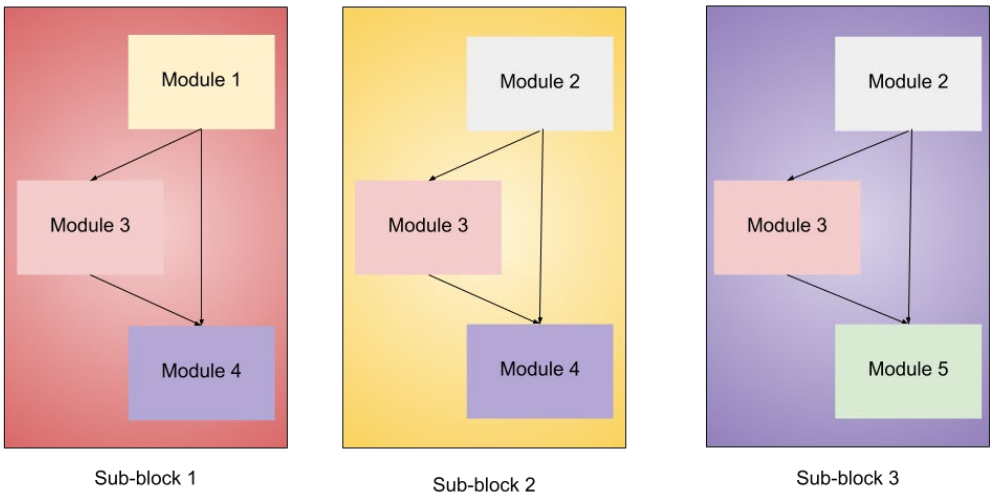


Figure 4-8 Sub-blocks for architecture construction[38]

Sub-block 1 is the sole sub-block in the first block. Sub-block 2 is the initial sub-block in each of the subsequent blocks. Sub-block 3 is used for any sub-block in a block other than the first. Thus, the architecture of EfficientNet-B4 is shown below.

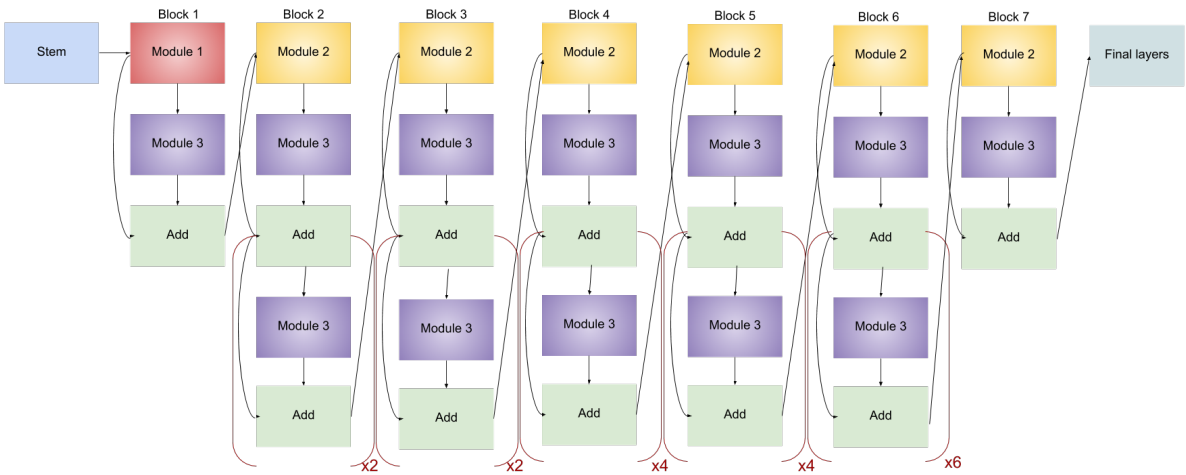


Figure 4-9 Architecture for EfficientNet-B4[38]

4.1.3 Convolutional Neural Network Optimization

The neural network may be thought of as a fitting function with parameterized inputs, and the loss function quantifies the difference between the predicted and actual values. In neural networks, the weight and bias of neurones are referred to as elementary parameters, whereas additional parameters are referred to as hyper-parameters.

The loss function is a surface in the parameter space, and the neural network's training procedure aims to discover the loss function's lowest value. Typically, the loss function is considered to be a lower convex or approximation lower convex function, and the objective of optimisation is to determine the extremum point of the lower convex function.

4.1.3.1 Optimizer

There are many techniques for determining the least value of a loss function, the most basic of which is gradient descent. The Adagrad[39], RMSProp[40], Adam(Adaptive Moment Estimation)[41] and Adamax[41] algorithms are used in this experiment.

Adagrad assigns a greater learning rate to often changing parameters and a lower learning rate to rarely changing ones. The RMSProp method converts Adagrad's sum of gradient squares to an exponentially weighted moving average, which improves performance in non-convex environments. Adam's momentum is included immediately into the estimate of the first instant of gradient. Second, in contrast to the absence of a correction factor, which results in a significant bias in the estimate of the second moment at the start of training, Adam incorporates bias correction, correction of the momentum component initialised from the origin, and non-central second moment estimation. Adamax is an infinity norm-based version of the Adam algorithm.

4.1.3.2 Scheduler

In this study, ReduceLROnPlateau, CosineAnnealingLR and CosineAnnealingWarmRestarts will be implemented. ReduceLROnPlateau reduces learning rate when a metric has stopped improving. CosineAnnealingLR set the learning rate of each parameter group using a cosine annealing. CosineAnnealingWarmRestarts set the learning rate of each parameter group using a cosine annealing schedule, and restarts after certain epochs.

4.1.3.3 Loss Function Selection

Cross Entropy Loss with label smoothing[42], Focal Loss[43] and Bi-Tempered Loss[44] will be utilized in this project. Cross Entropy is a loss function that is frequently used to express the difference between two probability distributions. The neural network will direct its learning in the direction of the greatest difference between the correct and incorrect labels. When there is insufficient training data to represent all the sample features, network overfitting occurs. Label smoothing can be used to address the aforementioned issues. It is a regularization strategy that, primarily by adding noise via soft one-hot, reduces the weight assigned to the categories of real

sample labels in the loss function calculation, thereby suppressing overfitting. Focal loss employs an adjustment item in cross entropy to concentrate on difficult examples and reduce the weight of a large number of easy negatives, thereby resolving the label imbalance problem.

Bi-Tempered Loss is a technique for minimizing the effect of noise on training results.

4.1.3.4 Hyper-parameters Search

The task of optimising a loss function over a graph-structured configuration space is called hyper-parameter optimisation. In this project TPE (Tree-structured Parzen Estimator) is used[45].

TPE is a sequential model-based optimisation (SMBO) approach. The SMBO technique first creates a model from previous index data in order to estimate the hyper-parameter's performance, and then selects the new hyper-parameter based on the model. TPE approach for modelling P (x|y) and P (y), where x denotes hyper-parameters and y denotes related evaluation index. P (x|y) by altering the parameters generation process to model, and instead of utilising the configuration prior distribution, employing non-parametric densities.

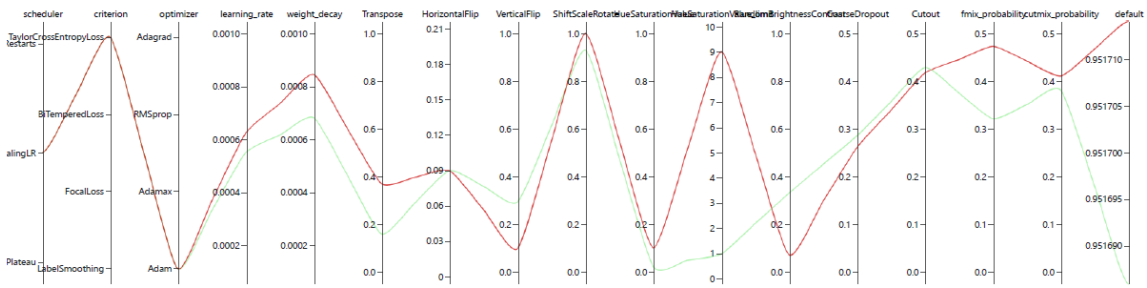


Figure 4-10 Hyper-parameter search with TPE Algorithm

4.1.4 Model Pipeline

The training process for the models, as shown in Figure 4-11, involves passing image data through the Data Augmentation part and then importing the model architecture, training the model with 5-fold cross-validation, and finally exporting the models and results. Hyper-parameters of the models include, Scheduler (Learning rate), Loss Function, Optimizer, and Training Procedure. The hyper-parameters of the models are searched by the TEP algorithm to obtain the optimal models.

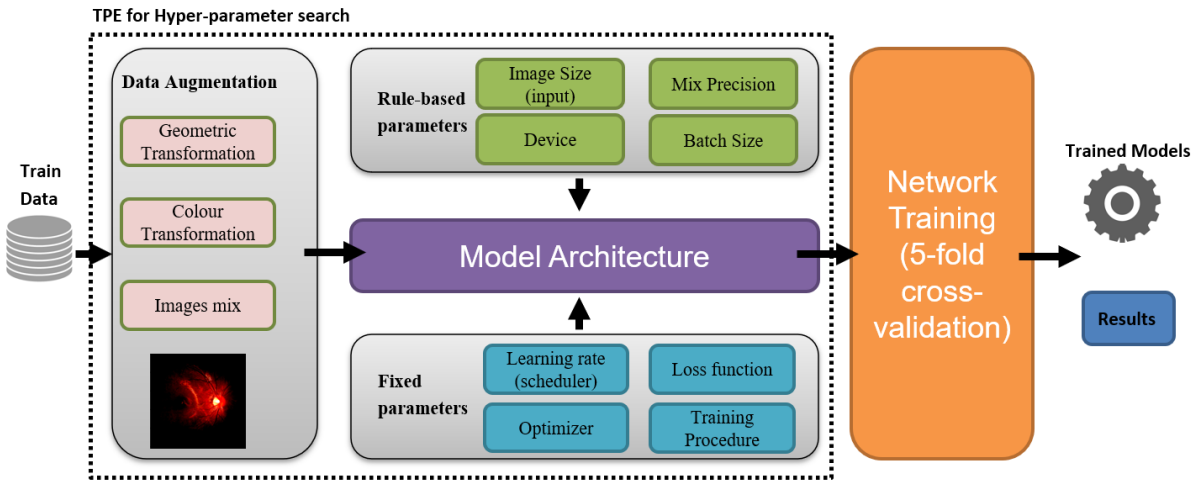


Figure 4-11 Pipeline Flow for Image Data Training

For the training of multi-modal data, as shown in Fig. 4-12, the optimized model in the previous step was selected for the training of image data. The FC layer of the neural network model was exported and combined with clinical data to be input into a data mining model. The whole model is trained with 5-fold cross-validation, and then the trained models and the results are exported.

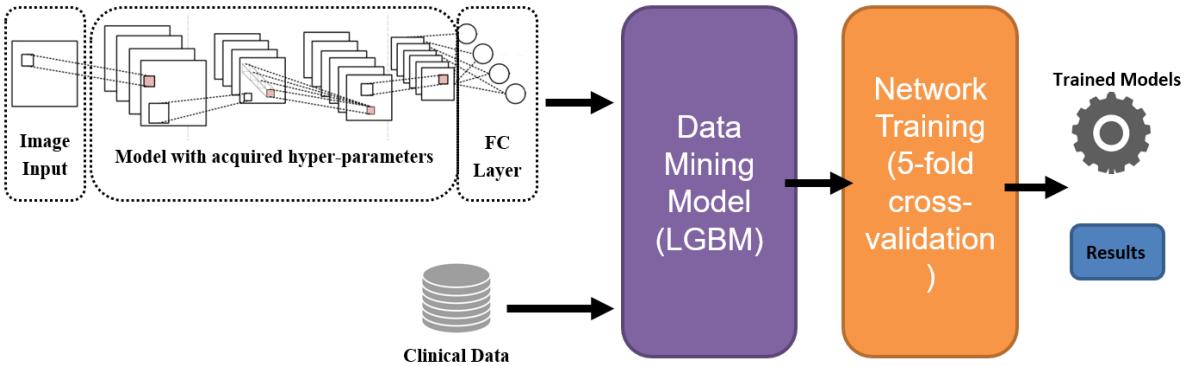


Figure 4-12 Pipeline Flow for Multi-modal Data Training

4.1.5 Model Evaluation Criteria

Model Evaluation Criteria will be the same as described in chapter 3. ROC curves will be plot for image data model as well as Multi-modal data model. AUC will be calculated and listed.

4.1.6 Model Visualization

A critical component of our deep learning strategy was communicating technical, machine learning ideas to medical professionals in an accessible, easy-to-understand way. Given the paucity of research on this subject, it was critical to identify regions of the fundus picture from which classifications might be formed. To visualise the class-specific gradients flowing into the final convolutional layer, we utilised a gradient-weighted class activation mapping (Grad-CAM). This information is used to create a coarse localisation map that emphasises the image's most significant areas[46]. By seeing the trained model's filters, we may get a better understanding of how the CNN learns the intricate spatial and temporal pixel relationships inherent in the picture[46].

This enables us to make associations with the anatomical characteristics associated with the detection, incidence, and progression of myopia. Additionally, we may examine predictions that are inconsistent and get insight into where and why the model may be missing.

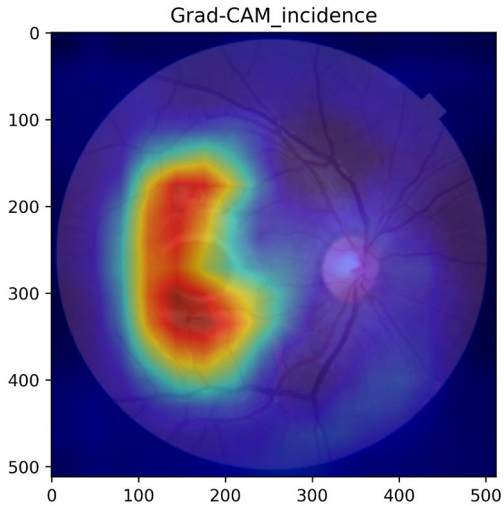


Figure 4-13 An incidence prediction image of model visualization

Heat map color scale: blue (low) - yellow - red (high)

[This image belongs to BHVI]

4.2 Results

4.2.1 Results of Image Data

4.2.1.1 AUC Results

Table 4-1 below shows the AUC values of Resnet and EfficientNet-B4 models on myopia detection, incidence and progression prediction tasks. As shown in the table, the results of Resnet model in detection task is slightly better, while in incidence prediction task is slightly worse than EfficientNet-B4 model. Both models get the same result in myopia progression prediction task. With deep learning the AUC results can be up to 0.95(Resnet) and 0.94(EfficientNet-B4) in detection, 0.80(Resnet) and 0.82(EfficientNet-B4) in incidence prediction and 0.75(Resnet & EfficientNet-B4) in progression prediction tasks.

Table 4-1 AUC results of each model on image data of three tasks

AUC	Detection	Incidence	Progression
Resnet	0.95	0.80	0.75
EfficientNet-B4	0.94	0.82	0.75

4.2.1.2 ROC Curves

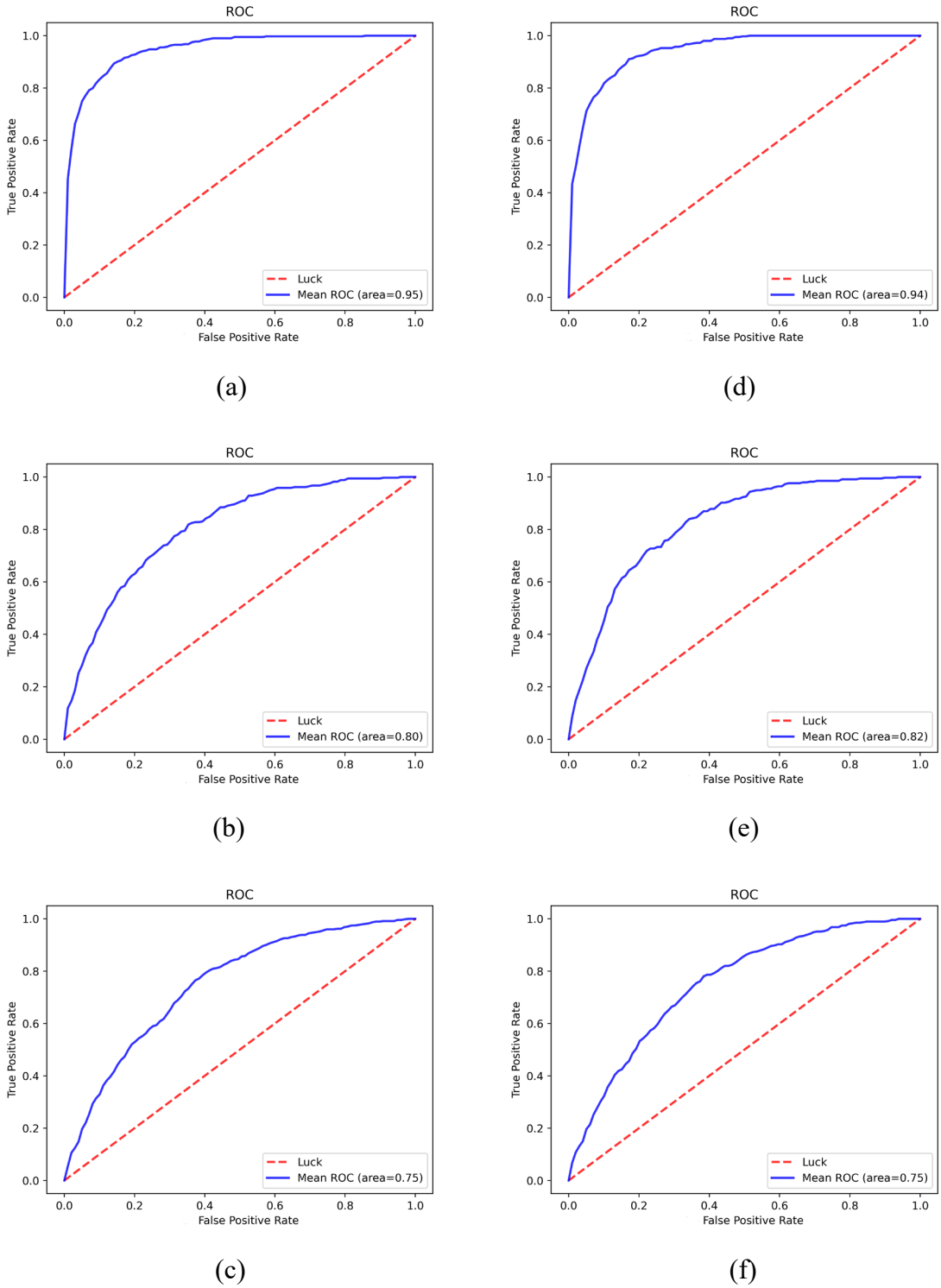


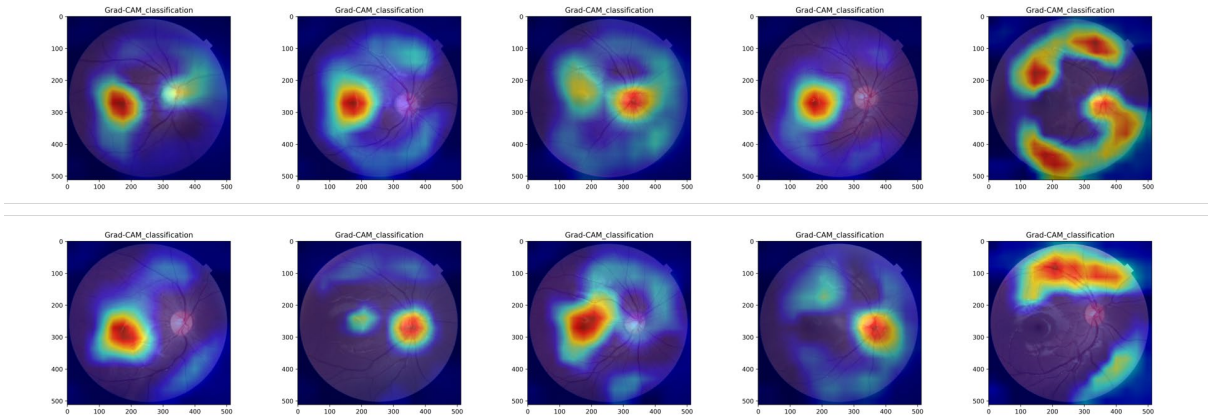
Figure 4-14 ROC curves of each model on image data of three tasks

(a, b, c: Resnet models ROC on detection, incidence and progression prediction tasks respectively. d, e, f: EfficientNet-B4 models ROC on detection, incidence and progression prediction tasks respectively)

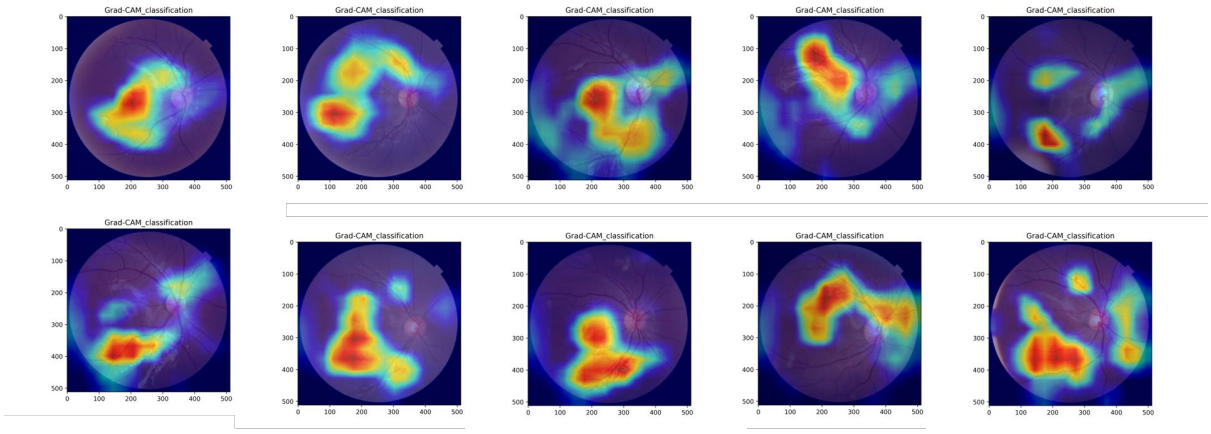
Figure 4-14 shows ROC curves of Resnet and EfficientNet-B4 models on image dataset of myopia detection, incidence and progression prediction tasks. a, b, c: Resnet models ROC on detection, incidence and progression prediction tasks respectively. d, e, f: EfficientNet-B4 models ROC on detection, incidence and progression prediction tasks respectively.

4.2.1.3 Image Visualization

Resnet Detection Model:

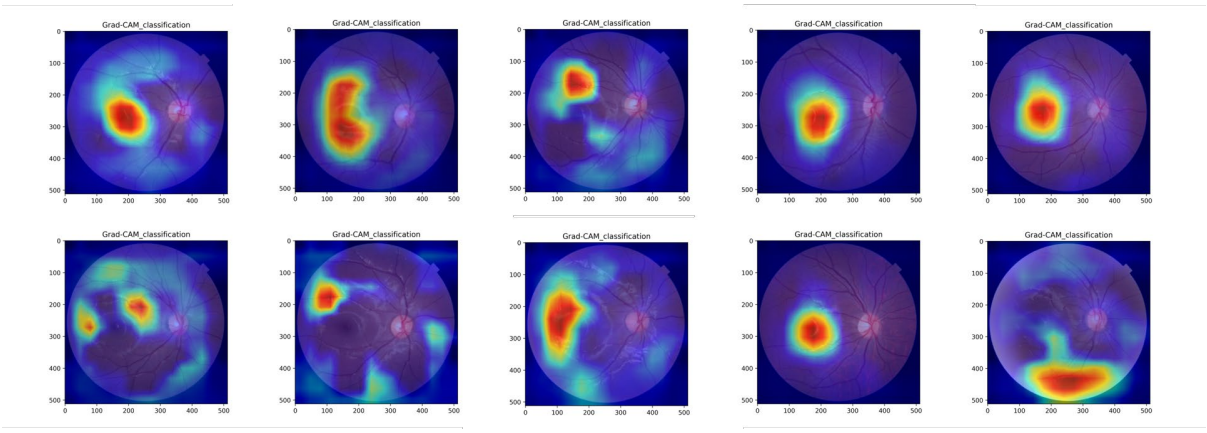


(a) Activation Heatmap of Retinal Images from Myope Children

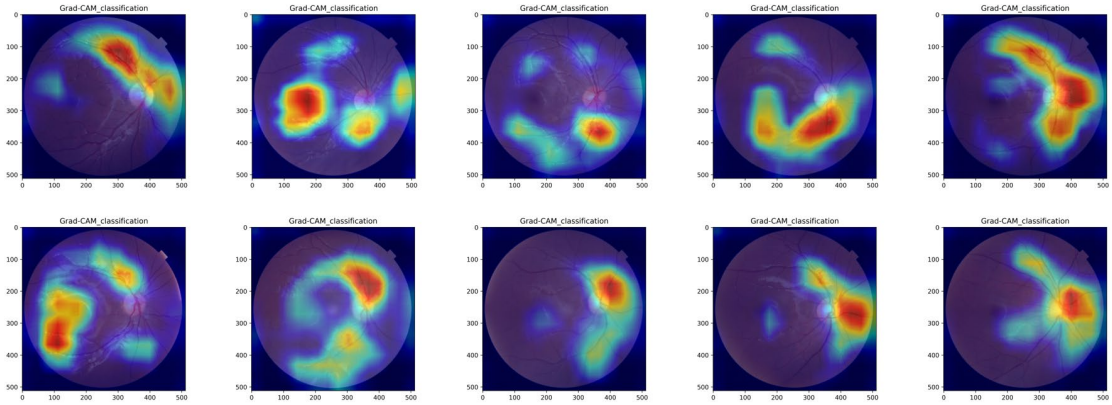


(b) Activation Heatmap of Retinal Images from Non-myope Children

Resnet Incidence Model:

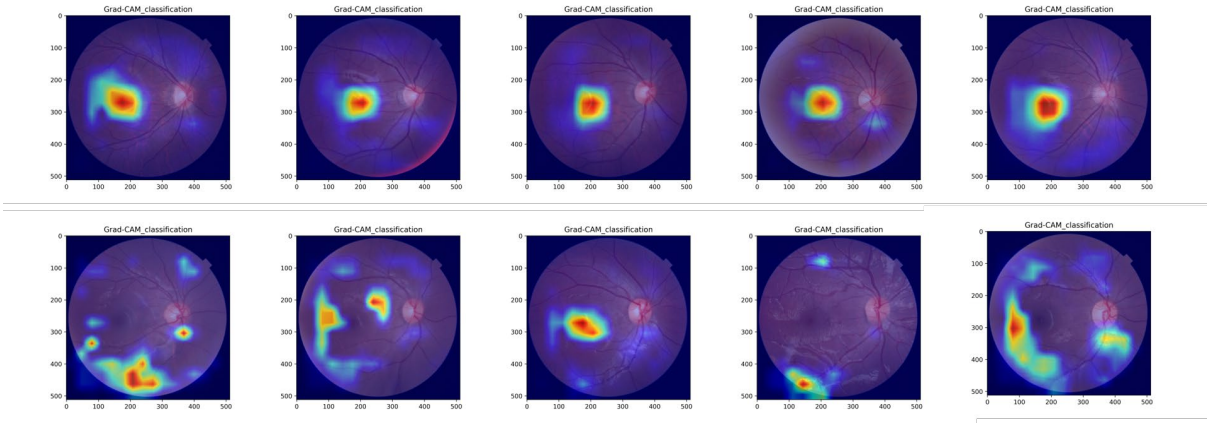


(c) Activation Heatmap of Retinal Images from Children with Incidence of Myopia

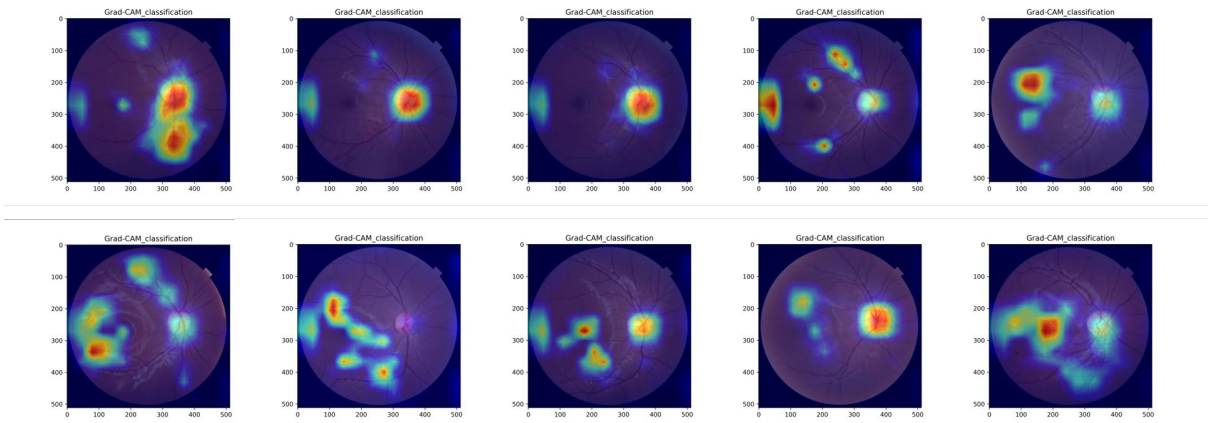


(d) Activation Heatmap of Retinal Images from Children with Non-incidence of Myopia

Resnet Progression Model:



(e) Activation Heatmap of Retinal Images from Children with Progression of Myopia



(e) Activation Heatmap of Retinal Images from Children with Non-progression of Myopia

Figure 4-15 GRAD-CAM visualization on Resnet models of three tasks with different retinal images

Heatmap Colour scale: blue (low) - yellow - red (high)

[These images belong to BHVI]

The GRAD-CAM visualization on Resnet models of three tasks with different retinal images is shown in figure 4-15 above. The heatmaps display the activated images associated with the patient labels. For the detection task, patients with myopia had activated regions concentrated around the macula or on the optic disc and its peripheral vessels. Non-myopic images showed activation mainly in the macula and its surrounding areas. For the myope incidence prediction task, for patients with incidence of myopia, the activation area was only shown around the macula. In contrast, the activated regions of patients with non-incidence of myopia mainly concentrated on optic disc and its peripheral blood vessels. For the myope progression prediction task, the activation area was concentrated in the small area of the macula in patients with myopic progression, while patients with non-progression were concentrated in the small area of the optic disc or the small area of the macula. The results are at odds with medical research and may provide new insights into myopia diagnosis.

4.2.2 Results of Multi-modal Data Training

4.2.2.1 AUC Results

Table 4-2 below shows the AUC values of EfficientNet-B4 models on multi-modal datasets of myopia detection, incidence and progression prediction tasks. As shown in the table, all the datasets have the same value on myopia detection. Image with clinical data plus al_by_cr is 1% higher than Image with clinical data (0.82) while 1% lower with 0.75.

Table 4-2 AUC results of EfficientNet-B4 model on different datasets of three tasks

AUC	Detection	Incidence	Progression
Image with clinical data	0.94	0.81	0.75
Image with clinical data plus al_by_cr	0.94	0.82	0.74
Image only	0.94	0.82	0.75

4.2.2.2 ROC Curves

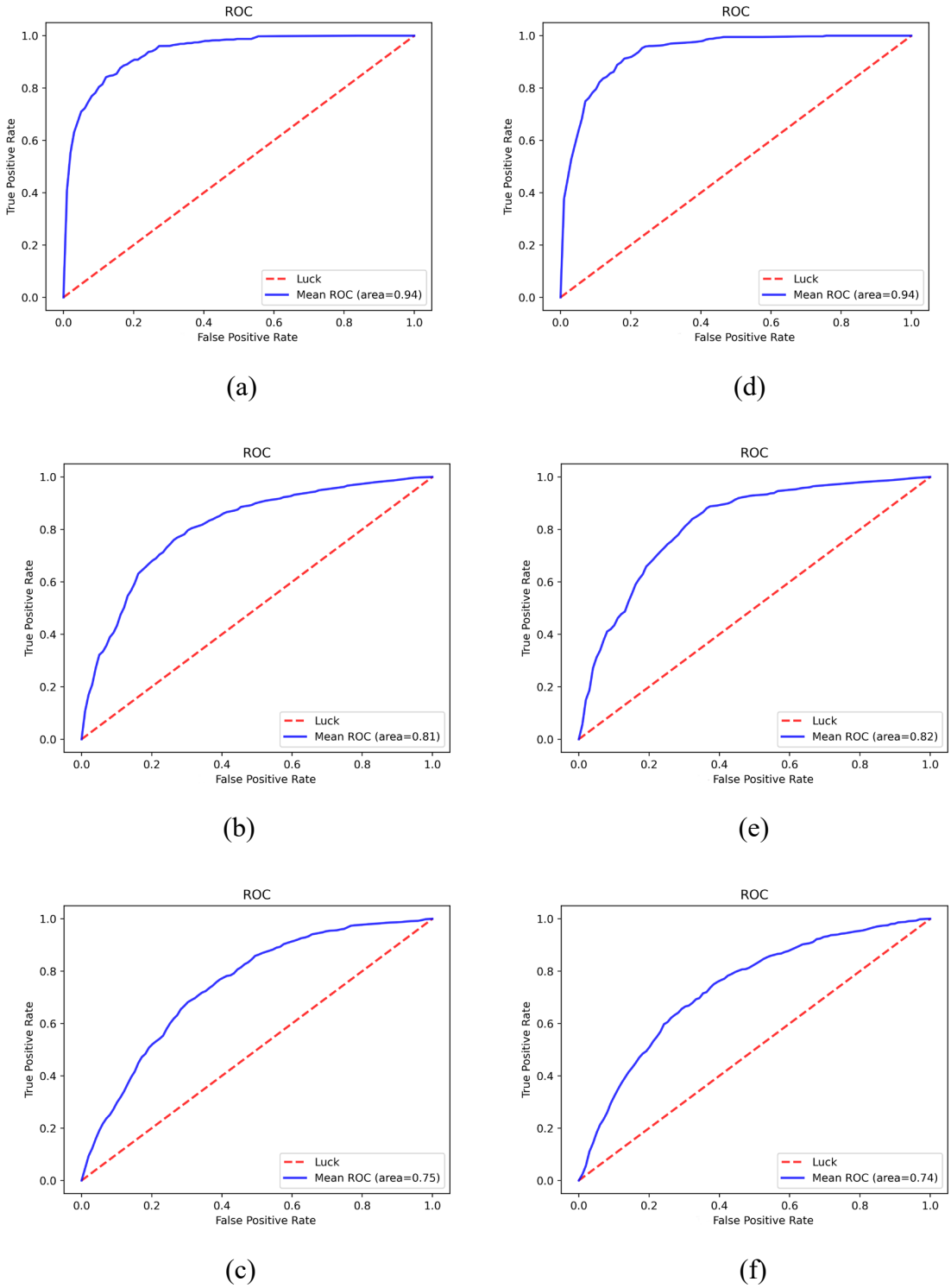


Figure 4-16 ROC results of EfficientNet-B4 model on different datasets of three tasks

(a, b, c: Image with clinical data ROC on detection, incidence and progression prediction tasks respectively. d, e, f: Image with clinical data plus al_by_cr ROC on detection, incidence and progression prediction tasks respectively)

Figure 4-16 shows ROC results of EfficientNet-B4 model on different datasets of myopia detection, incidence and progression prediction tasks. a, b, c: Image with clinical data ROC on detection, incidence and progression prediction tasks respectively. d, e, f: Image with clinical data plus al_by_cr ROC on detection, incidence and progression prediction tasks respectively

4.3 Summary & Discussion

4.3.1 Summary

Two different neural network models with distinct structures were trained and optimised on the retinal image dataset in this chapter. The Resnet model's activation images for three tasks were provided for clinical research. The EfficientNet model was modified and trained on image and clinical data. Additionally, we train the model in clinical datasets with and without the feature axial length to corneal radius ratio.

Resnet model performs slightly better in detection tasks than EfficientNet-B4 models, but performs slightly worse in incidence prediction tasks. Both models achieve the same result in the prediction of myopia progression task. AUC values of up to 0.95 (Resnet) and 0.94 (EfficientNet-B4) can be achieved in detection tasks, 0.80 (Resnet) and 0.82 (EfficientNet-B4) in incidence prediction tasks, and 0.75 (Resnet & EfficientNet-B4) in progression prediction tasks using deep learning. The GRAD-CAM visualisation of three tasks with distinct retinal images is plotted on Resnet models. Patients with myopia had activated regions around the macula or on the optic disc and its peripheral arteries for the detection task. The macula and its surrounding areas were mostly activated in non-myopic pictures. The activation area was only exhibited around the macula for patients with myopia incidence in the myope incidence prediction task. Patients with non-incidence of myopia, on the other hand, had their activated regions localised mostly on the optic disc and its peripheral blood vessels. In the myope progression prediction task, patients with myopia progression had their activation area concentrated in the small region of the macula, whereas patients without progression had their activation area concentrated in the small area of the optic disc or the small area of the macula.

On multi-modal datasets, the AUC of EfficientNet-B4 models is the same for myopia detection. Image with clinical data plus al_by_cr is 1% higher than Image with clinical data (0.80), but 1% lower than Image with clinical data (0.75).

4.3.2 Discussion

The following conclusions can be drawn from the results obtained from the experiment:

1. Compared with data mining models, deep learning models can obtain higher AUC results only by inputting retinal image data.
2. Compared with ResNet, EfficientNet-B4 can obtain the same quality results with fewer model parameters.
3. Model visualization shows that the neural network learns the information contradicts medical research and could lead to new insights on myopia diagnosis.
4. Based on the results of multi-modal datasets, it was inferred that clinical data had limited effects.

Chapter 5. Conclusion & Future Works

In this study, first, we use a data mining approach to evaluate the extracted features from retinal images and ophthalmic data in this work. The retinal image dataset was then used to train and optimize two separate neural network models with differing topologies. Finally, the EfficientNet model was tweaked and trained using clinical and image data.

5.1 Conclusion

5.1.1 Summary

First, we use a data mining approach to evaluate the extracted features from retinal images and ophthalmic data in this work. The retinal image dataset was then used to train and optimize two separate neural network models with differing topologies. Finally, the EfficientNet model was tweaked and trained using clinical and image data.

The following data are available regardless of the axial length to corneal radius ratio: The LGBM model produces the best results when combined with the extracted features and ophthalmic data. The values for Detection, Incidence, and Progression were 0.89, 0.71, and 0.70, respectively. LGBM model performs slightly better in the Detection task than Logistic Regression models, but slightly worse in the Incidence and Progression tasks. The results show that the LGBM model's output is statistically significantly different from the other models' output in all three tasks. The findings of data analysis using the feature axial length to corneal radius ratio are summarised as follows: The LGBM model produces the most precise outcomes. The values for Detection, Incidence, and Progression were 0.96, 0.81, and 0.74, respectively. The LGBM model's AUC increases by 7%, 10%, and 14% in the detection, incidence, and progression tasks, respectively, when compared to results that do not include this feature. The results of the pair-wise independent samples t-test on the AUC values for detection, incidence, and progression suggest that the output of all models in all three tasks is statistically significantly different from the output of the other models.

The Resnet model outperforms the EfficientNet-B4 model in detection tasks but falls short in incidence prediction. In the job of predicting myopia progression, both models attain the same outcome. Deep learning may reach AUC values of up to 0.95 (Resnet) and 0.94 (EfficientNet-B4) in detection, 0.80 (Resnet) and 0.82 (EfficientNet-B4) in incidence prediction, and 0.75

(Resnet & EfficientNet-B4) in progression prediction tasks. Resnet models are used to see the GRAD-CAM visualisation of three tasks with different retinal pictures. Patients with myopia had activated areas localised around the macula or on the optic disc and its peripheral arteries for the detection task. Non-myopic pictures revealed activity mostly in the macula and its surroundings. For the myope incidence prediction test, the activation area was only seen around the macula for patients with an incidence of myopia. Of contrast, the active areas in individuals who did not have myopia were primarily focused on the optic disc and its peripheral blood arteries. In patients with myopic progression, the activation area was focused in the small region of the macula, whereas patients without advancement were concentrated in the small area of the optic disc or the small area of the macula.

The AUC of EfficientNet-B4 models for myopia detection is identical on multi-modal datasets. Image with clinical data + al_by_cr is 1% more than Image with clinical data (0.80), but 1% less than Image with clinical data (0.80). (0.75).

5.1.2 Conclusion

The LGBM model's results differ significantly from those of the other methods. Obviously, the training model converges to distinct vertices for loss minimisation under the same sample space due to different loss functions of different models and different learning procedures. As can be seen, each model's outputs are distinct. The data mining model outperforms the Logistic Regression model in the myopia detection challenge but falls short in the myopia incidence and progression prediction tasks. The data mining model, in particular, can eliminate the need for time-consuming manual feature selection. In the detection, incidence, and progression prediction tasks, the LGBM model's AUC increases by 7%, 10%, and 14%, respectively, when compared to results that do not include the feature axial length to corneal radius ratio. This is in line with the findings of clinical studies. Deep learning algorithms, when compared to data mining methods, can achieve superior AUC results simply by inputting retinal image data. EfficientNet-B4 can get the same quality outcomes as ResNet with fewer model parameters. The findings contradict previous medical studies and may provide fresh light on myopia diagnosis. Clinical data was found to have limited effects based on the results of multi-modal datasets.

5.2 Future works

1. Since the number of samples for the data mining model training is less than that in the deep learning model, the same number of samples need to be used to retrain the deep learning model in order to make a better comparison between the two models.
2. This project is part of the Master of Information Technology's Master of Information Technology programme, which lasts only two trimesters. Due to time constraints, there is still some work to be done on this subject, such as the pair-wise independent samples t-test for deep learning model results. Subsequent work can include conducting a statistical significance test on the results following retraining with the same sample size.
3. On the Internet, a wealth of open source retinal image data with labels is available. Pre-training the model on these publicly available datasets may have a beneficial effect on the final result. It's worth a shot.
4. Additionally, BHVI has some private datasets on which the model can be tested and the model's robustness discussed further.

References

[1] D. Turbert, “Nearsightedness: What is myopia?,” in American Academy of Ophthalomology, 2020.

[2] Holden, Brien A., et al. "Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050." Ophthalmology 123.5 (2016): 1036-1042.

[3] Wong, Tien Y., et al. "Epidemiology and disease burden of pathologic myopia and myopic choroidal neovascularization: an evidence-based systematic review." American journal of ophthalmology 157.1 (2014): 9-25.

[4] Fricke, T. R., et al. "Global cost of correcting vision impairment from uncorrected refractive error." Bulletin of the World Health Organization 90 (2012): 728-738.

[5] Liu, Xiaolong, Zhidong Deng, and Yuhan Yang. "Recent progress in semantic image segmentation." Artificial Intelligence Review 52.2 (2019): 1089-1106.

[6] BOptom, Divya Jagadeesh. Predictive models for myopia in children: Detection, Incidence and Progression. Diss. School of Optometry and Vision Science Faculty of Science, The University of New South Wales, 2020.

[7] Dolgin, Elie. "The myopia boom." Nature News 519.7543 (2015): 276.

[8] Cheng, Sam CK, Carly SY Lam, and Maurice KH Yap. "Prevalence of myopia-related retinal changes among 12–18 year old H ong K ong C hinese high myopes." Ophthalmic and Physiological Optics 33.6 (2013): 652-660.

[9] Li, X. "Beijing Rhegmatogenous Retinal Detachment Study Group: Incidence and epidemiological characteristics of rhegmatogenous retinal detachment in Beijing, China." Ophthalmology 110 (2003): 2413-2417.

[10] Praveen, Mamidipudi R., et al. "Prevalence of cataract type in relation to axial length in subjects with high myopia and emmetropia in an Indian population." American journal of ophthalmology 145.1 (2008): 176-181.

[11] Wong, Tien Yin, et al. "Refractive errors, axial ocular dimensions, and age-related cataracts: the Tanjong Pagar survey." Investigative ophthalmology & visual science 44.4 (2003): 1479-1485.

[12] Machida, Shigeki, et al. "High prevalence of myopia in Japanese patients with idiopathic focal subretinal neovascularization." Retina 26.2 (2006): 170-175.

[13] He, Xiangui, et al. "Shanghai time outside to reduce myopia trial: design and baseline data." Clinical & experimental ophthalmology 47.2 (2019): 171-178.

[14] Ovens, Dannielle, et al. "Automated Grading of Refractive Error from Fundus Images using Deep Learning." Imaging Systems and Applications. Optical Society of America, 2020.

[15] James, Gareth, et al. An introduction to statistical learning. Vol. 112. New York: springer, 2013.

[16] Wilcox, K. Tyler. "An Introduction to Ensemble Methods for Machine Learning." (2016).

[17] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

[18] Strobl, Carolin, James Malley, and Gerhard Tutz. "An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests." Psychological methods 14.4 (2009): 323.

[19] Friedman, Jerome H. "Stochastic gradient boosting." Computational statistics & data analysis 38.4 (2002): 367-378.

[20] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25 (2012): 1097-1105.

[21] Recker, hawn, and Christiaan Gribble. “Real-Time, In Situ Intelligent Video Analytics: Harnessing The Power of GPUs for Deep Learning Applications.” DSIAC, 2 Nov. 2019, www.dsiac.org/resources/articles/real-time-in-situ-intelligent-video-analytics-harnessing-the-power-of-gpus-for-deep-learning-applications.

[22] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[23] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[24] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[25] McMonnies, Charles W. "Clinical prediction of the need for interventions for the control of myopia." Clinical and Experimental Optometry 98.6 (2015): 518-526.

[26] McCullough, Sara, et al. "Axial Growth and Refractive Change in White European Children and Young Adults: Predictive Factors for Myopia." Scientific Reports (2020).

[27] Lin, Haotian, et al. "Prediction of myopia development among Chinese school-aged children using refraction data from electronic medical records: A retrospective, multicentre machine learning study." PLoS medicine 15.11 (2018): e1002674.

[28] Devasia, Thresiamma, K. Poulose Jacob, and Tessamma Thomas. "Automatic early stage glaucoma detection using cascade correlation neural network." Smart Intelligent Computing and Applications. Springer, Singapore, 2019. 659-669.

[29] Hnoohom, Narit, and Anuchit Jitpattanakul. "Comparison of ensemble learning algorithms for cataract detection from fundus images." 2017 21st International Computer Science and Engineering Conference (ICSEC). IEEE, 2017.

[30] Levenkova, Anastasia, et al. "Automatic detection of diabetic retinopathy features in ultra-wide field retinal images." Medical Imaging 2017: Computer-Aided Diagnosis. Vol. 10134. International Society for Optics and Photonics, 2017.

[31] Govindaiah, Arun, Roland Theodore Smith, and Alauddin Bhuiyan. "A new and improved method for automated screening of age-related macular degeneration using ensemble deep neural networks." 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018.

[32] Varadarajan, Avinash V., et al. "Deep learning for predicting refractive error from retinal fundus images." Investigative ophthalmology & visual science 59.7 (2018): 2861-2868.

[33] Garner, L. F., et al. "The Nepal longitudinal study: predicting myopia from the rate of increase in vitreous chamber depth." Optometry and vision science 81.1 (2004): 44-48.

[34] Roth, Holger R., et al. "Anatomy-specific classification of medical images using deep convolutional nets." 2015 IEEE 12th international symposium on biomedical imaging (ISBI). IEEE, 2015.

[35] P. Ruiz, "Understanding and visualizing resnets," in Towards Data Science, 2018.

[36] Karim, Raimi. “Illustrated: 10 CNN Architectures.” Towards Data Science, 28 Nov. 2020, towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d.

[37] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." International Conference on Machine Learning. PMLR, 2019.

[38] Agarwal, Vardan. “Complete Architectural Details of All EfficientNet Models.” Towards Data Science, 31 May 2020, towardsdatascience.com/complete-architectural-details-of-all-efficientnet-models-5fd5b736142.

[39] Duchi, John, Elad Hazan, and Yoram Singer. "Adaptive subgradient methods for online learning and stochastic optimization." Journal of machine learning research 12.7 (2011).

[40] Geoffrey Hinton Neural Networks for machine learning nline course. https://www.coursera.org/learn/neural-networks/home/welcome

[41] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

[42] Müller, Rafael, Simon Kornblith, and Geoffrey Hinton. "When does label smoothing help?." arXiv preprint arXiv:1906.02629 (2019).

[43] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. 2017.

[44] Amid, Ehsan, et al. "Robust bi-tempered logistic loss based on bregman divergences." arXiv preprint arXiv:1906.03361 (2019).

[45] Bergstra, James, et al. "Algorithms for hyper-parameter optimization." Advances in neural information processing systems 24 (2011).

[46] Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." Proceedings of the IEEE international conference on computer vision. 2017.