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

Never Stand Still

Faculty of Engineering | CSE

Master Student: Zhao WANG

Supervisor: Prof. Arcot Sowmya

Co-supervisors: Prof. Padmaja Sankaridurg

Advisory Panel: Dr. Thomas John Naduvilath
& Dr. Divya Jagadeesh

OUTLINE OF THE PRESENTATION

- Background
- Literature Review
- Data Mining Models
- Deep Learning Models
- Summary

Myopia: A Global Epidemic

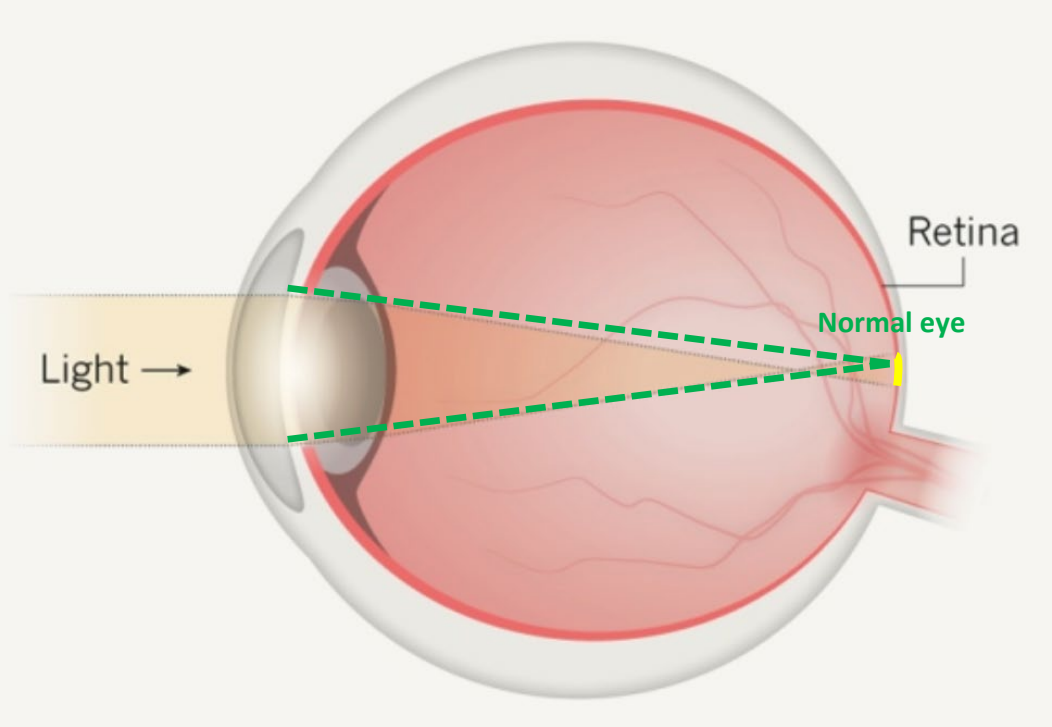


Figure 1: mechanism of myopia[1]

High myopia related pathologic changes :

- **Retinal Detachment**
- **Glaucoma**
- **Cataract**
- **Macular Degeneration**

[1] Dolgin E. The myopia boom. Nature 2015;519:276-8.

Myopia: A Global Epidemic

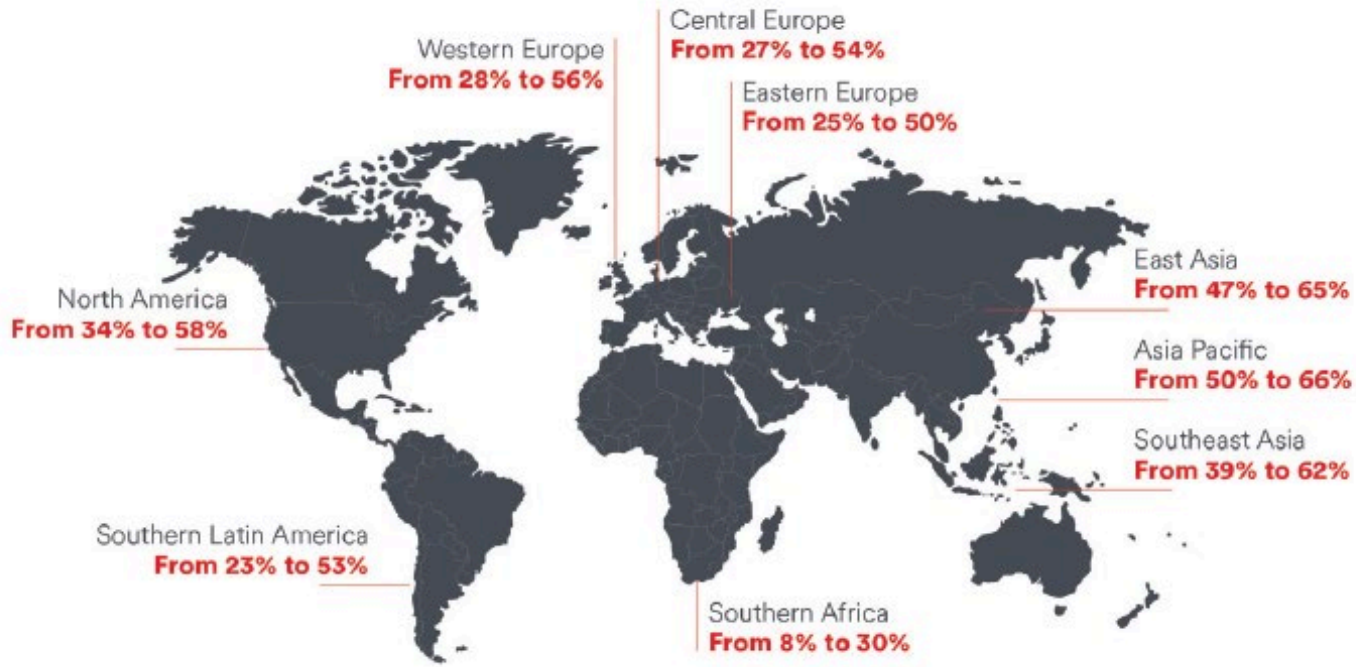


Figure 2: Current and projected 2050 myopia prevalence by region [2]

[2] Johnson & Johnson Vision and Holden BA, Fricke TR, Wilson DA, et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. Ophthalmology. 2016;123(5):1036-1042.

Myopia: A Global Epidemic

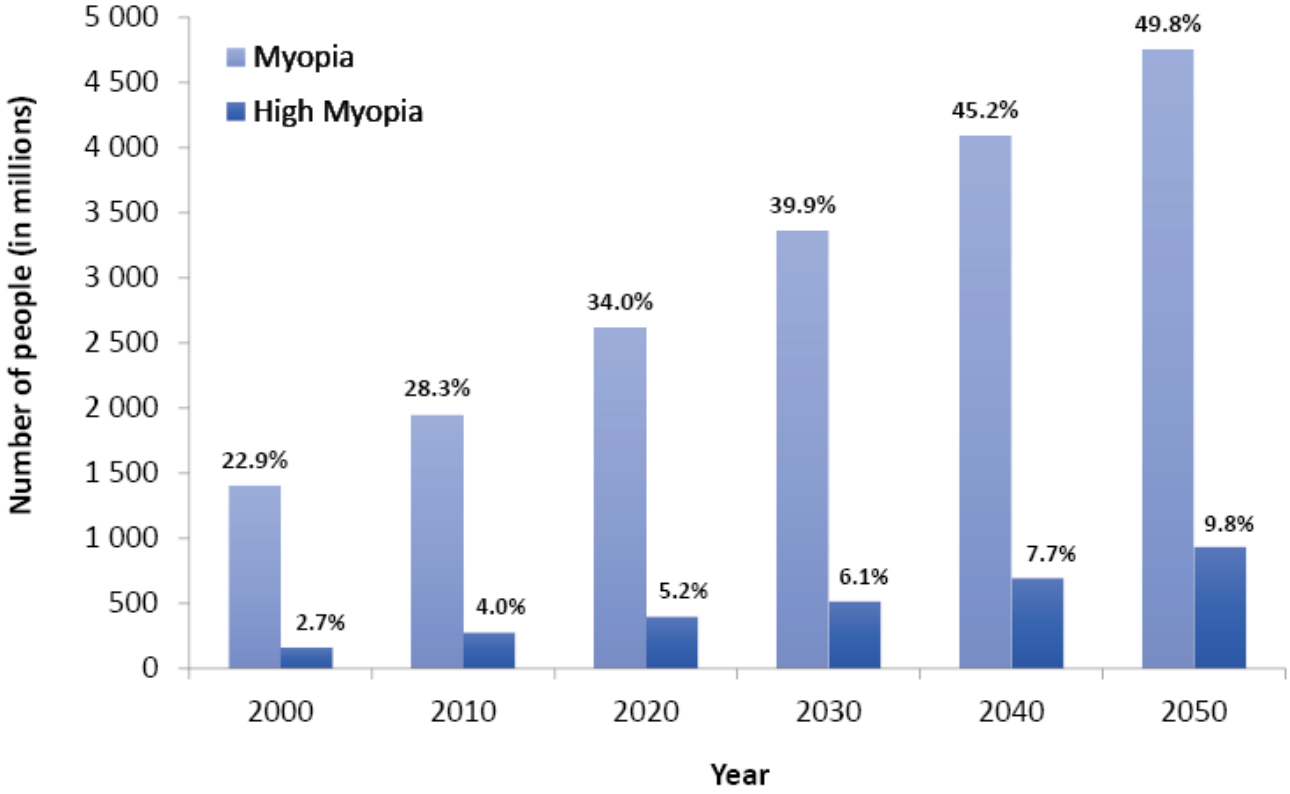


Figure 3: The estimated global prevalence of myopia and high myopia per decade from 2000 to 2050 based on current trends. [3]

[3] Holden BA, Fricke TR, Wilson DA, et al. Global Prevalence of Myopia and High Myopia and Temporal Trends from 2000 through 2050. Ophthalmology 2016.

Myopia Progression Prediction in Adolescents

- **Predict the progression of myopia in children**

△ Hypothesis:
A considerable amount of data has been collected. Therefore, the progression of myopia can be predicted with the implementation of statistical learning methods.

△ Aim:
To predict the progression of myopia in children using data mining methods and deep learning methods.

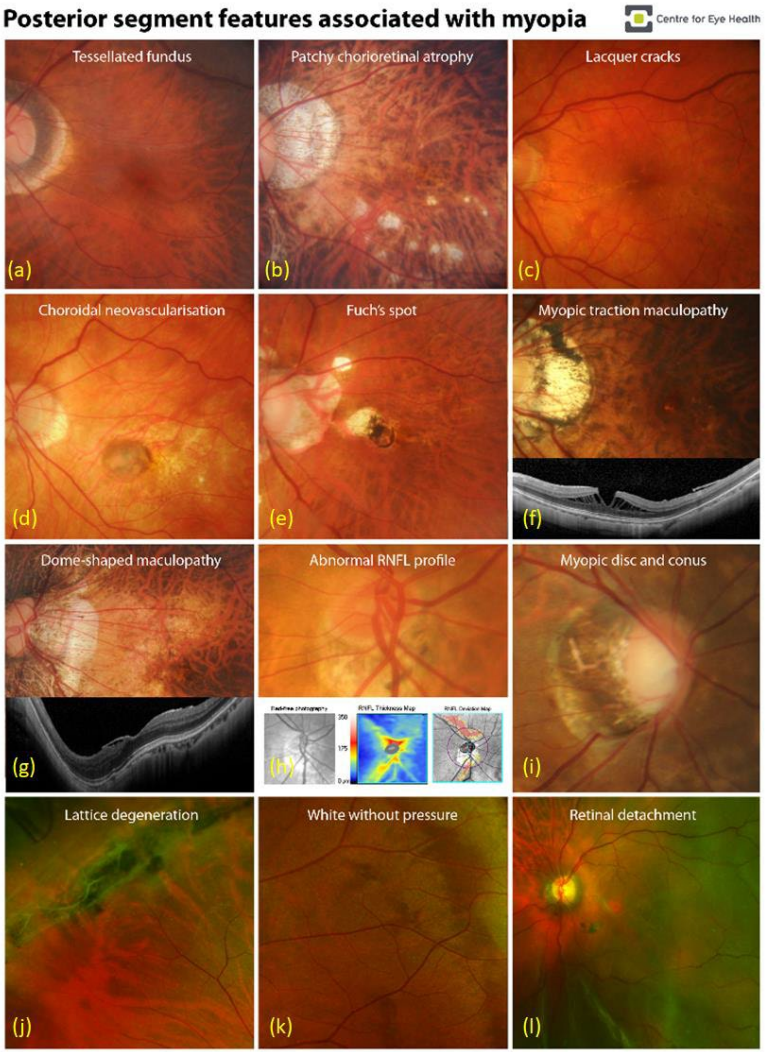


Figure 4: Posterior segment changes associated with myopia

Literature Review – Related Works

Table 1: Research on Ophthalmic Disease with Machine Learning

Researcher	Problem	Method	Reference
Langade Umesh et.al.	Eye diseases detection	Naive Bayes, KNN, SVM	(2014)ISSN: 2395 - 0056
R. Priya & P. Aruna	Diabetic retinopathy classification	Bayesian arrangement & SVM	(2016)ISSN: 2229-6956
Patitapaban Rath	Infected eye classification	Pattern recognition	(2017) DOI: 10.1146/14.356987
Maya Alsheh Ali et.al.	Glaucoma classification	Features analysis	(2014) DOI: 10.1117/12.2043098
Ohsugi Hideharu et.al.	Rhegmatogenous retinal detachment(RRD) classification	Deep learning	(2017) DOI:10.1038/s41598-017-09891-x
Varadarajan et al.	Refractive error predicting	Deep learning	(2018) doi.org/10.1167/iovs.18-23887

Literature Review – Clinical Works

$$\textit{Equation 4:1;}y = (0.149 \times Age\ (years) + (0.518 \times Parental\ myopia{:}\ 1) + (0.620 \times Parental\ myopia{:}\ 2) + (-1.029 \times Gender(male)) + (0.451 \times Grades\ of\ temporal\ crescent) + (0.025 \times pRNFL\ temporal - nasal\ difference\ (\mu m)) + (-0.016 \times Fovea\ to\ disc\ distance\ (pixels)) + (-0.005 \times Total\ macular\ thickness(\mu m)) + (0.913 \times presence\ of\ optic\ disc\ ovality) + (1.917 \times Baseline\ axial\ length\ (mm)) + (-31.84)$$

$$\textit{Equation 4:2;}y = (0.207 \times Age(years)) + (0.248 \times Parental\ myopia{:}\ 1) + (0.627 \times Parental\ myopia{:}\ 2) + (-0.576 \times Gender(male)) + (-0.011 \times short\ axis\ of\ the\ optic\ disc\ (pixels)) + (-0.012 \times Fovea\ to\ disc\ distance(pixels)) + (-0.056 \times Overall\ tessellations) + (-0.019 \times pRNFL\ thickness\ temporal - nasal\ difference(\mu m)) + (-0.005 \times Macular\ In\ thickness(\mu m)) + (0.006 \times Tilt\ from\ Vertical(°)) + (0.008 \times STRAA(°)) + (0.839 \times baseline\ axial\ length(mm)) - (5.93$$

$$\textit{Equation 4:3;}y = (-0.126 \times Age(years) + (0.0228 \times Parental\ myopia{:}\ 1, + (0.677 \times Parental\ myopia{:}\ 2) + (0.510 \times Gender(male) + (-1.762 \times Presence\ of\ baseline\ myopia) + (0.005 \times Tilt\ from\ Vertical(°)) + (-0.021 \times Temporal\ crescent\ Garde{:}\ 1 + (0.073 \times Temporal\ crescent\ Grade{:}\ 2) + (0.498 \times Temporal\ crescent\ Grade \geq 3) + (-0.004 \times Fovea\ to\ disc\ distance(pixels)) + (0.005 \times IAA(°)) + (0.205 \times Baseline\ axial\ length(mm)) - 3.401.$$

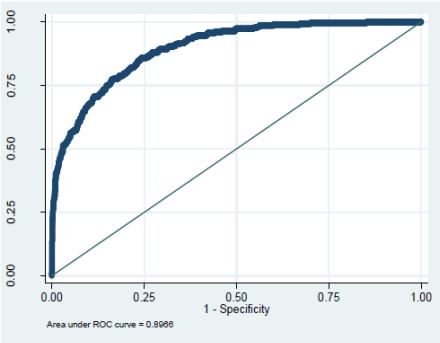


Figure 4.13 ROC curve for the predictive model to differentiate myopes and non-myopes.

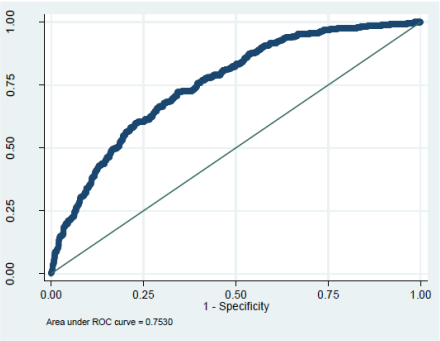


Figure 4.17 ROC curve for predictive model for incidence of myopia.

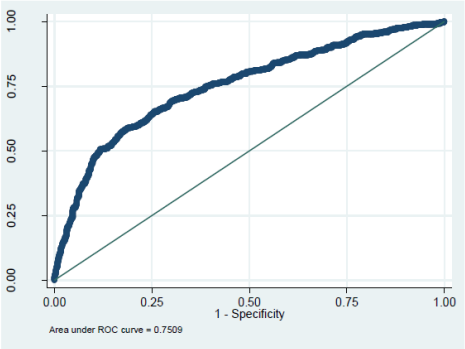


Figure 4.24 ROC curve for the predictive model for the progression of myopia

Figure 5: ROC curves of LR models from BHVI research

works for improvement:

1. **manually select features takes too much effort.**
2. **Results are over-fitting.**

Literature Review – Data Mining

Table 2: Comparison of different data mining models

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

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

Literature Review – Ensemble Learning

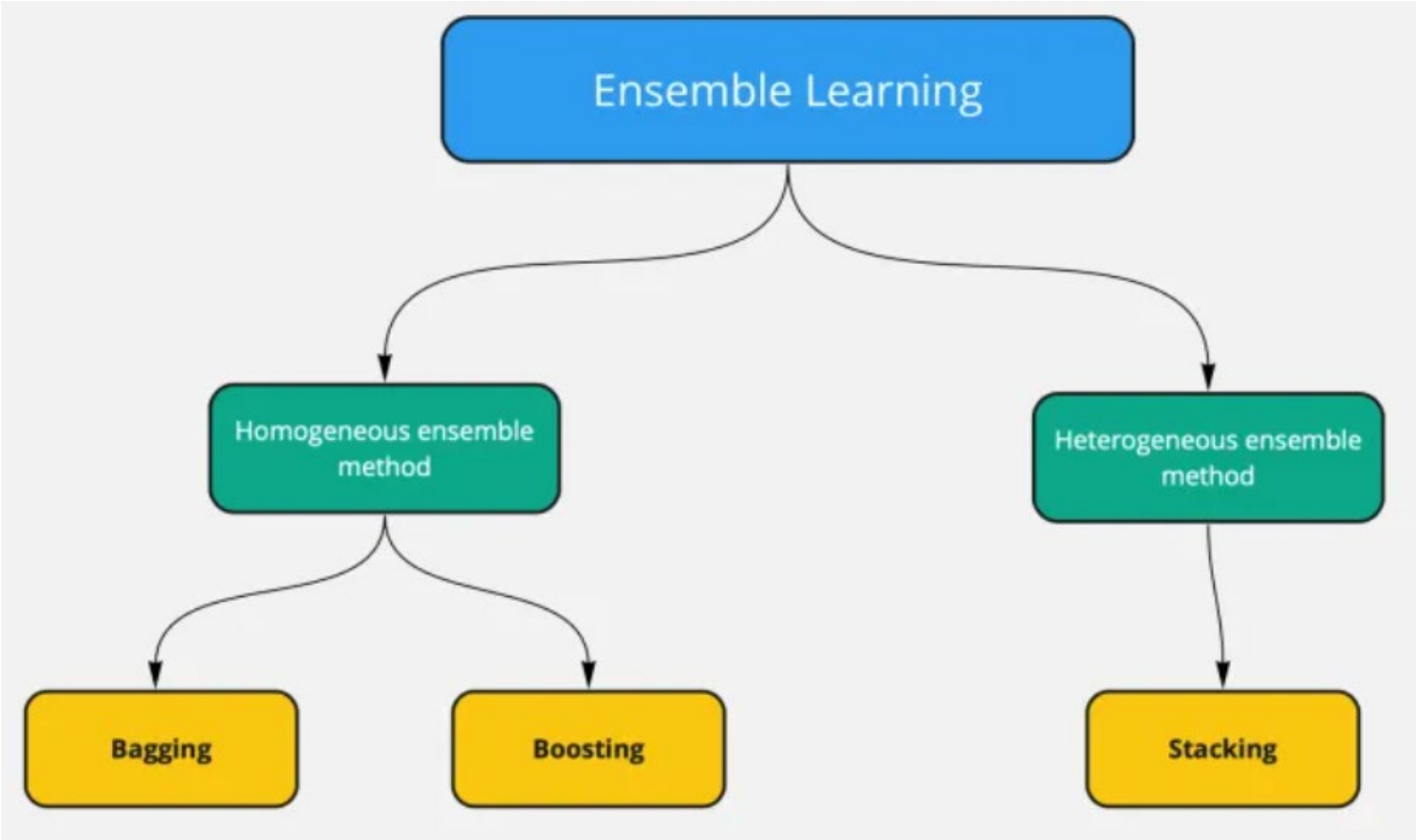


Figure 6: Ensemble Learning Category

Literature Review – Ensemble Learning

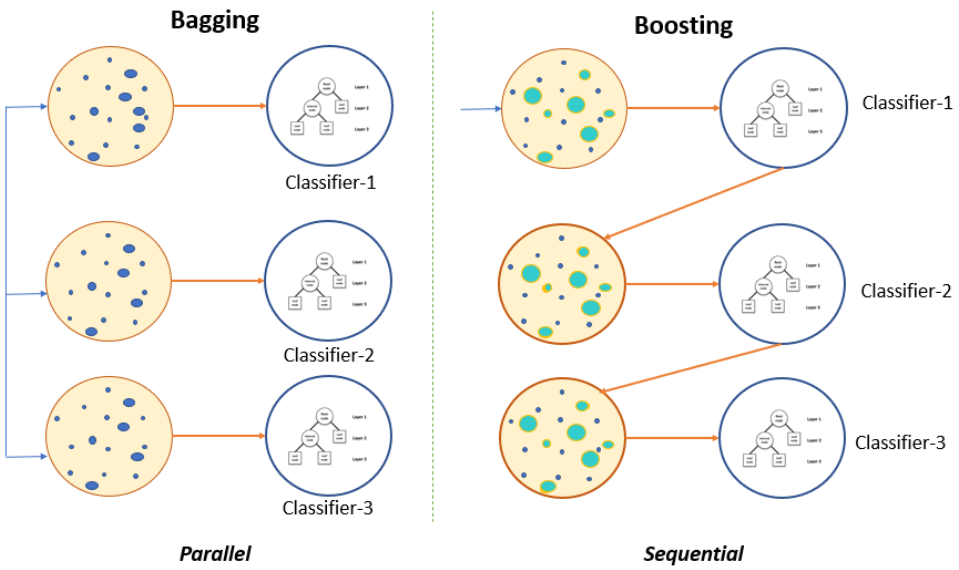
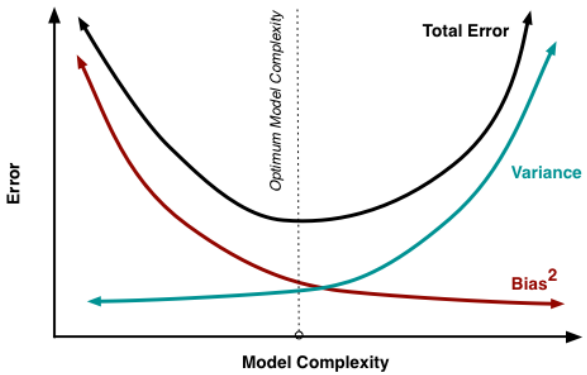


Figure 7: Bagging Vs Boosting

Literature Review – Ensemble Learning



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

Bagging

$$\begin{aligned} E(F) &= \sum_i^m r_i E(f_i) \\ &= m\frac{1}{m}\mu \\ &= \mu \end{aligned}$$

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

Bagging reduce the variance

Boosting

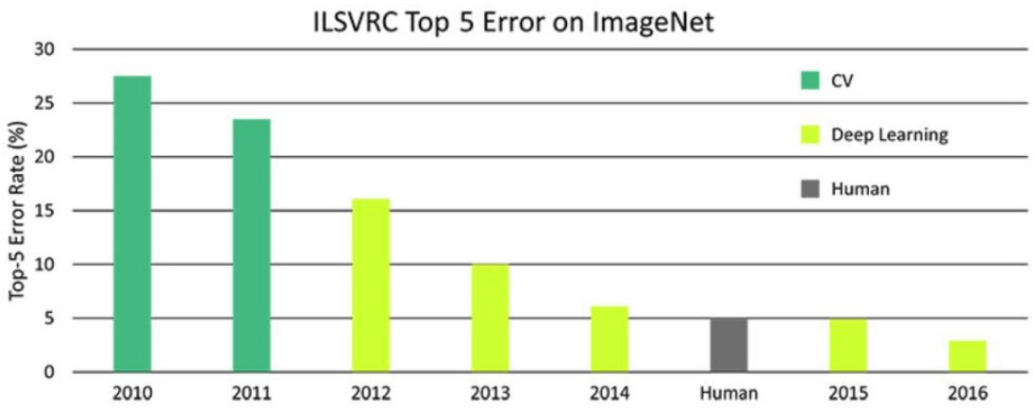
$$\boxed{E(F) = \sum_i^m r_i E(f_i)}$$

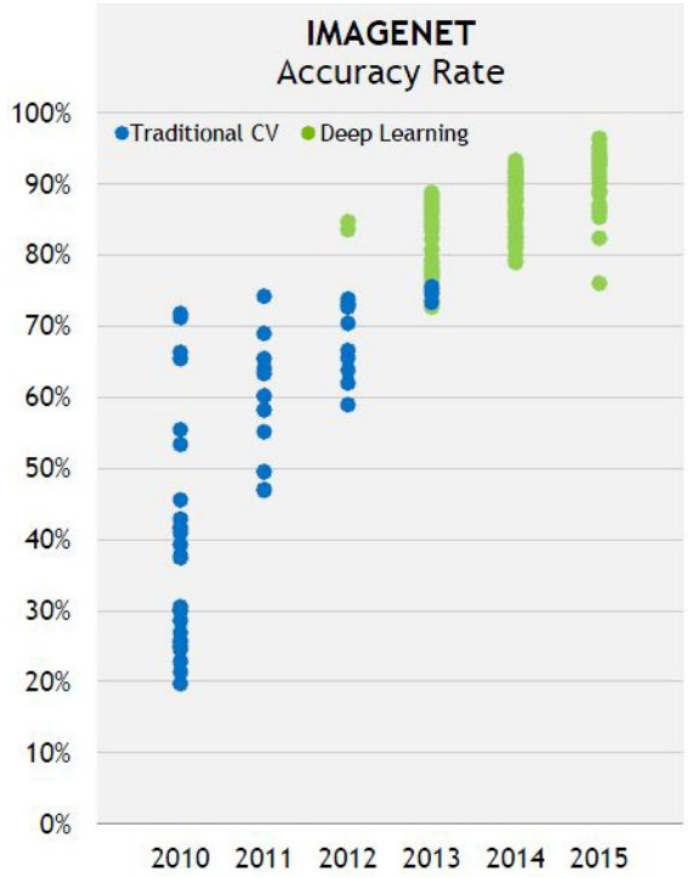
$$\begin{aligned} Var(F) &= mr^2\sigma^2(1-\rho) + m^2 r^2 \sigma^2 \rho \\ &= m\frac{1}{m^2}\sigma^2(1-1) + m^2\frac{1}{m^2}\sigma^2 1 \\ &= \sigma^2 \end{aligned}$$

Boosting reduce the bias

Literature Review – Deep Learning in CV

Deep Learning has produced significant performance improvements on a variety of important problems very quickly





Literature Review – Deep Learning

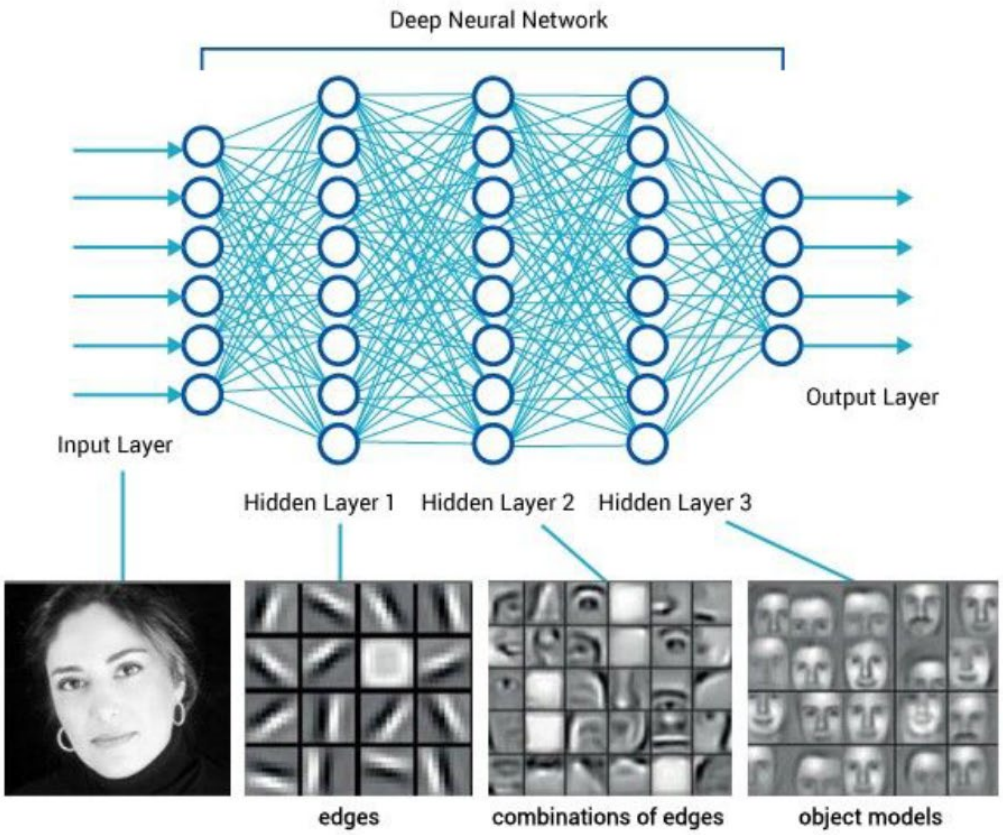
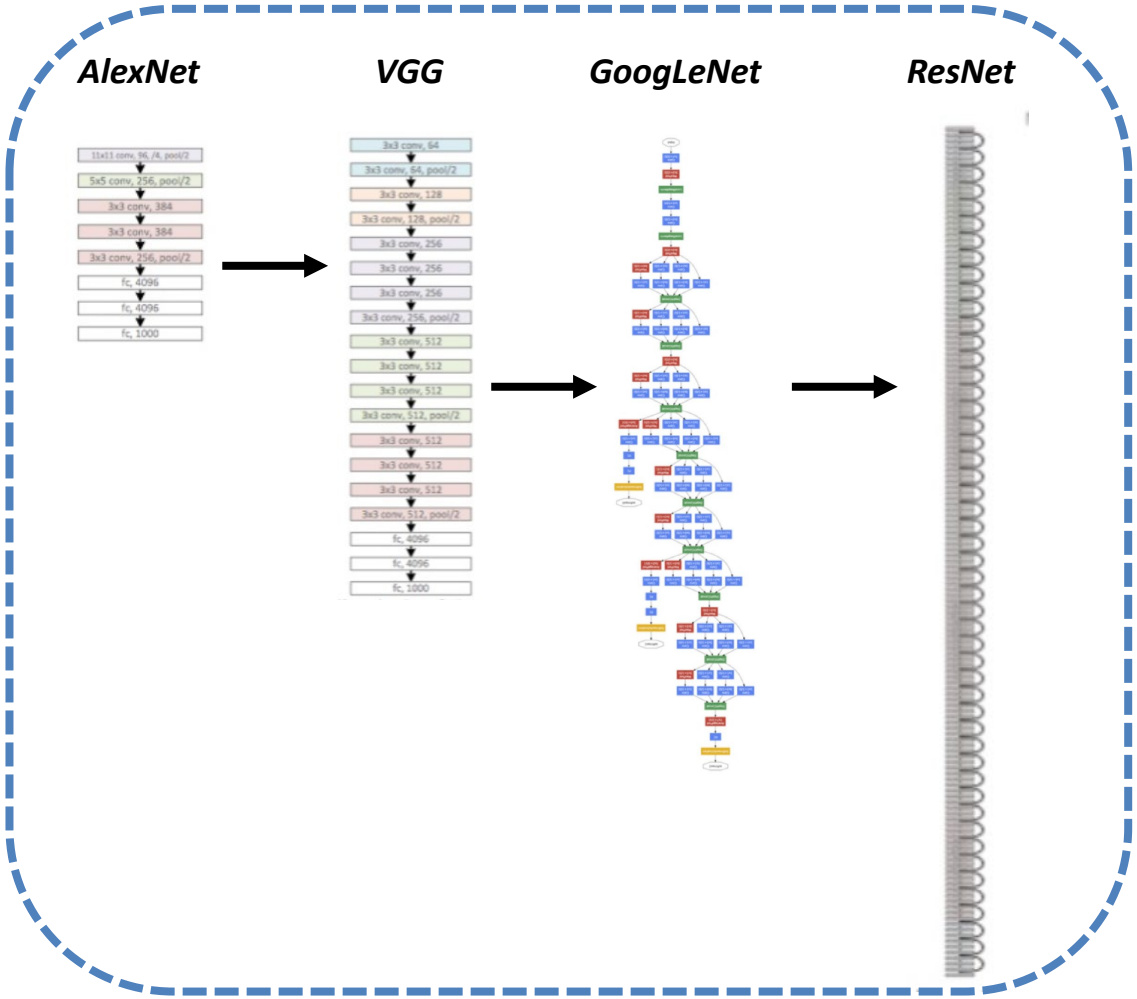


Figure 8: Feature Extraction from Deep Neural Network

Literature Review – CNN for classification

Figure 9: The development of neural network structures:



- **DEEPER**
- **WIDER**
- **HIGHER RESOLUTIONS**

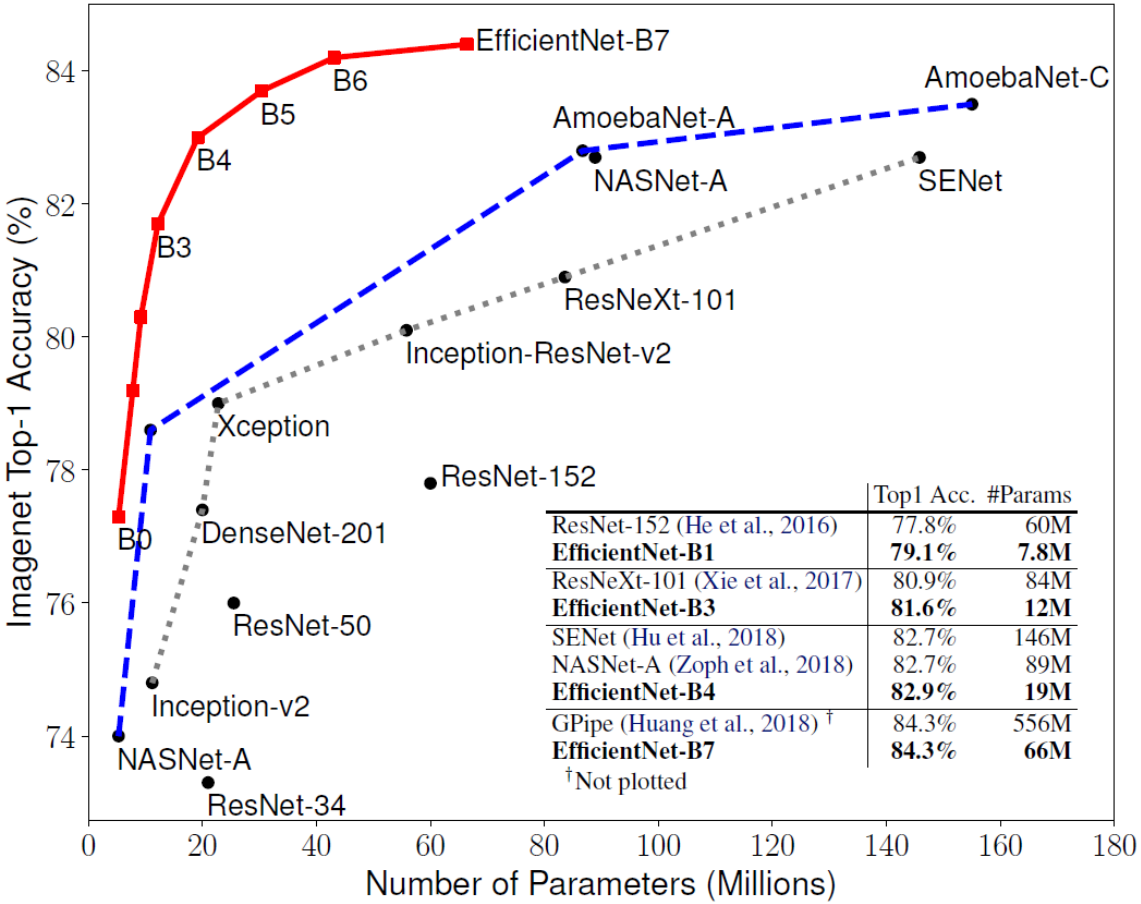
AlexNet: Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25 (2012): 1097-1105.
VGG: Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
GoogleNet: Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
ResNet: He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

Literature Review – CNN for classification

Table 3: 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

Literature Review – CNN for classification



	Top1 Acc.	#Params
ResNet-152 (He et al., 2016)	77.8%	60M
EfficientNet-B1	**79.1%**	**7.8M**
ResNeXt-101 (Xie et al., 2017)	80.9%	84M
EfficientNet-B3	**81.6%**	**12M**
SENet (Hu et al., 2018)	82.7%	146M
NASNet-A (Zoph et al., 2018)	82.7%	89M
EfficientNet-B4	**82.9%**	**19M**
GPipe (Huang et al., 2018) †	84.3%	556M
EfficientNet-B7	**84.3%**	**66M**

†Not plotted

Figure 10: Model Size vs Image-Net Accuracy

Literature Review – EfficientNet

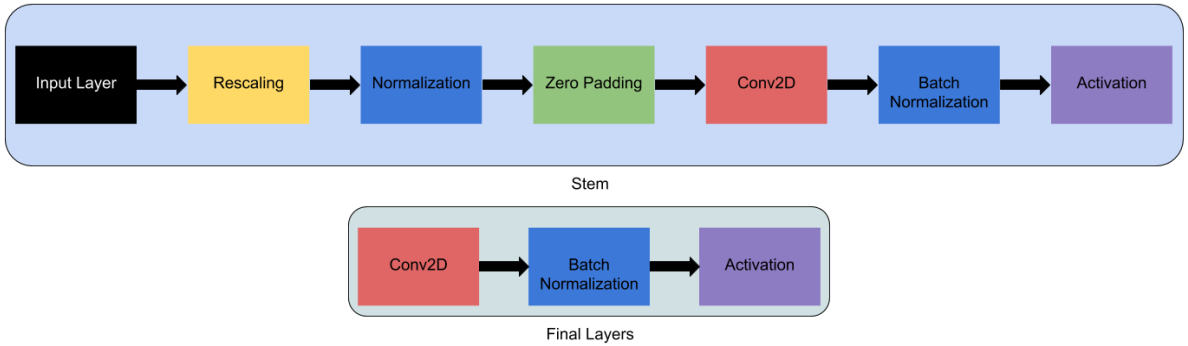


Figure 11:Stem and Final Layers in EfficientNet

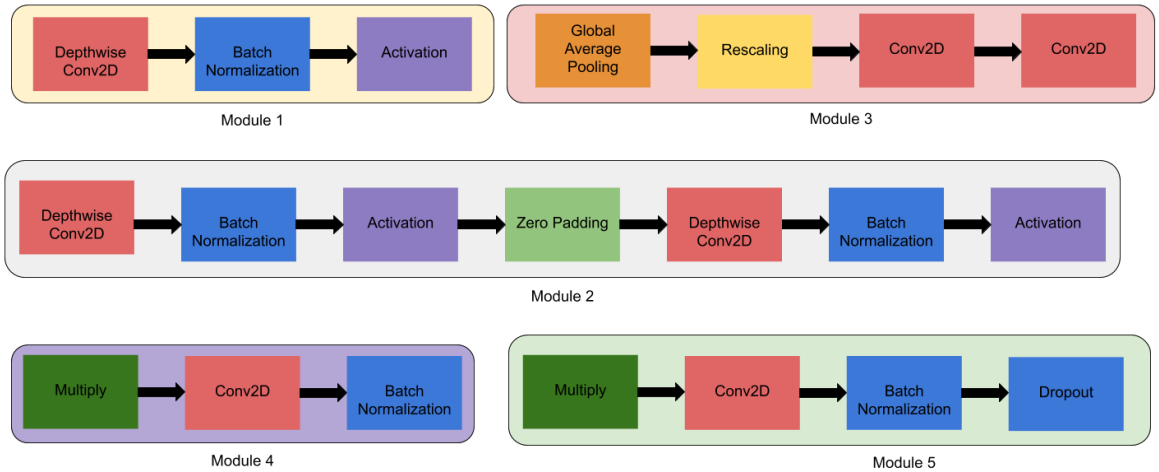
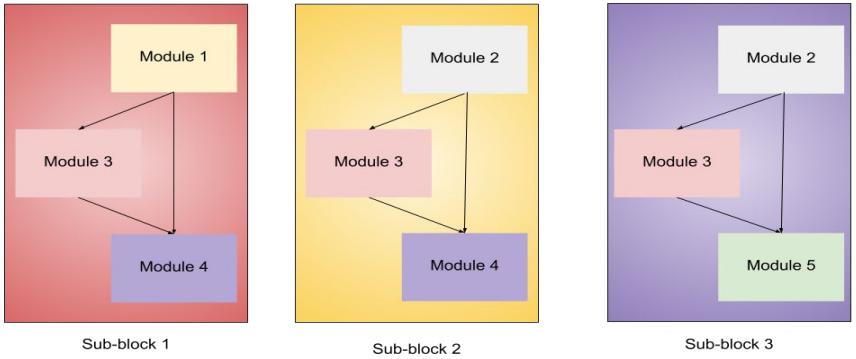


Figure 12: Five modules for architecture construction

- 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.

Literature Review – EfficientNet



- 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.

Figure 13: Sub-blocks for architecture construction

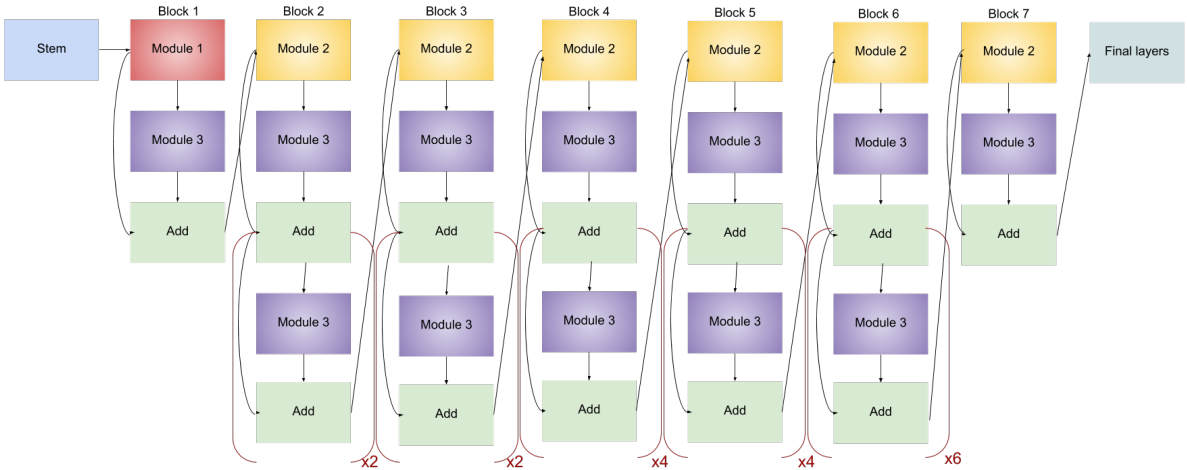


Figure 14: Architecture for EfficientNet-B4

Data Mining Models

Data Sets Introduction

2875 participants' features Semi-Auto extracted from retinal images [2169 × 38] and the other ophthalmic data [2169 × 189]

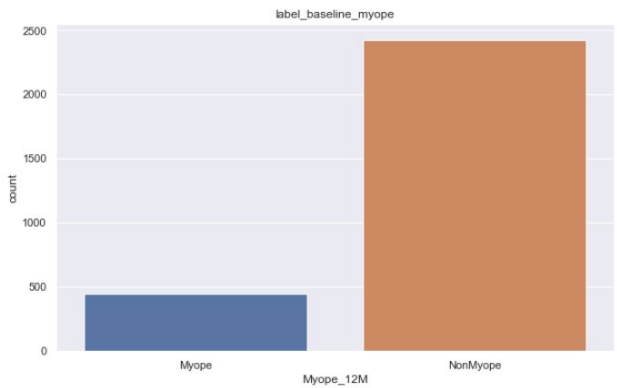
(such as age, parental myopia, axis length and other data from machine) in month 0 and their refractive error labels in month 0 and month 12.

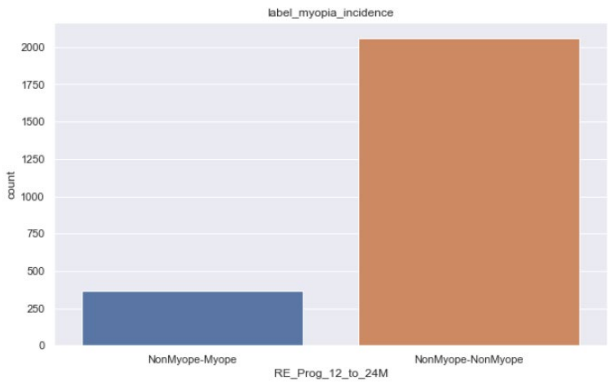
Data Set – Extracted Features

```
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 15: view of first 38 features of the dataset

Data Sets - Exploratory Data Analysis (EDA)





Therefore, This is a classification problem

Figure 16: Myope distribution in month 0 and Myope incidence distribution in month 12

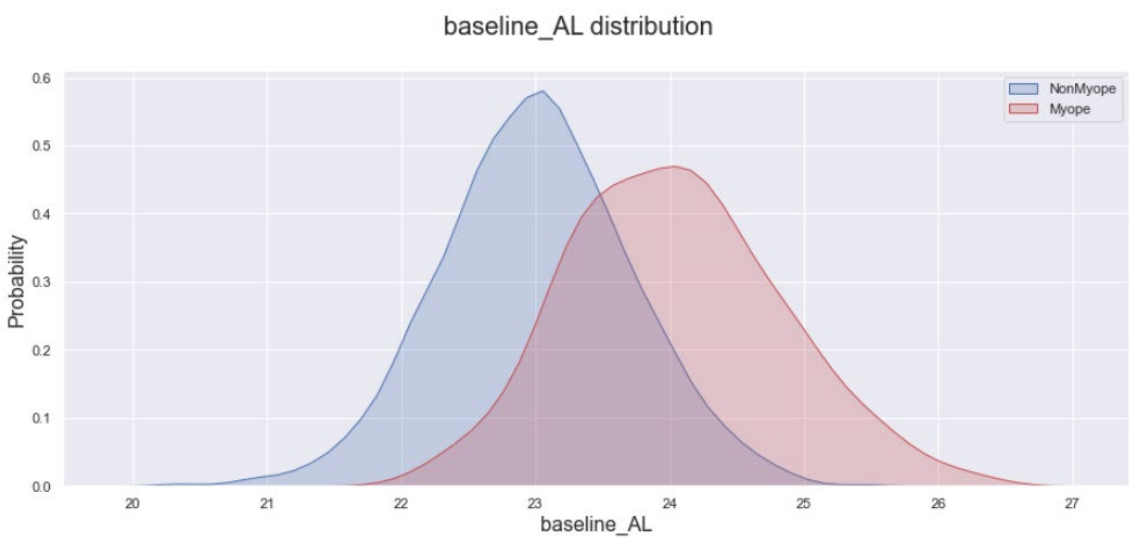


Figure 17: Axis Length distribution with Myope and Non-myope in month 0

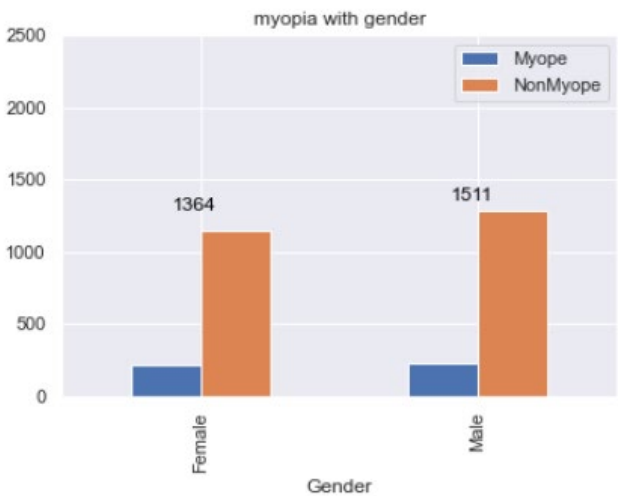


Figure 18: Myope distribution with gender in month 0

Data Mining Models

Table 4: Data Mining Models for datasets and tasks

Models \Training data set	Extracted features [2875]	Extracted features + ophthalmic data(intersection)[2169]
GBDT(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

Data Mining Results - AUC

Table 5 :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

Data Mining Results - ROC

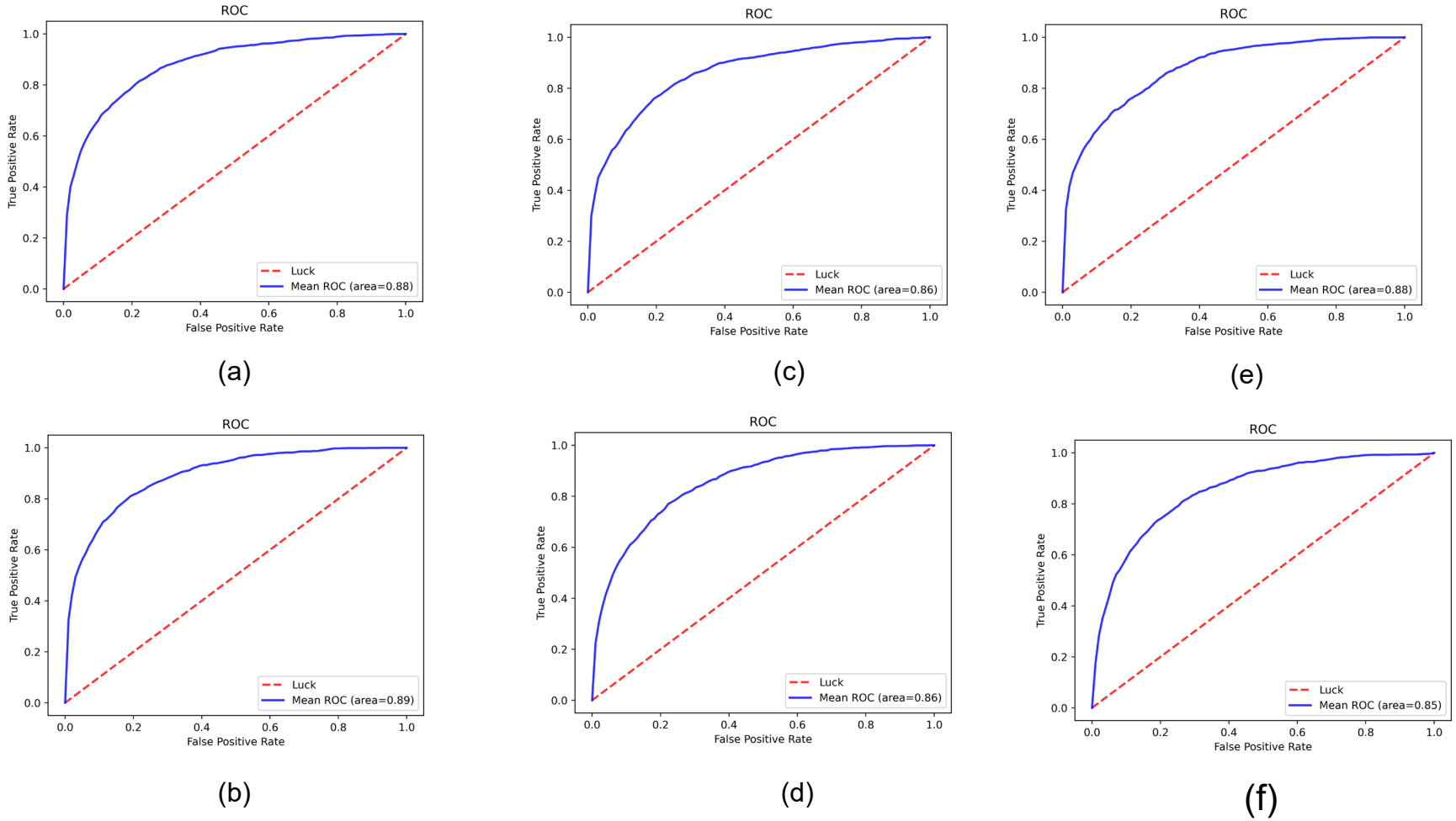


Figure 19 ROC curves of each detection model on datasets

(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)

Data Mining Results - ROC

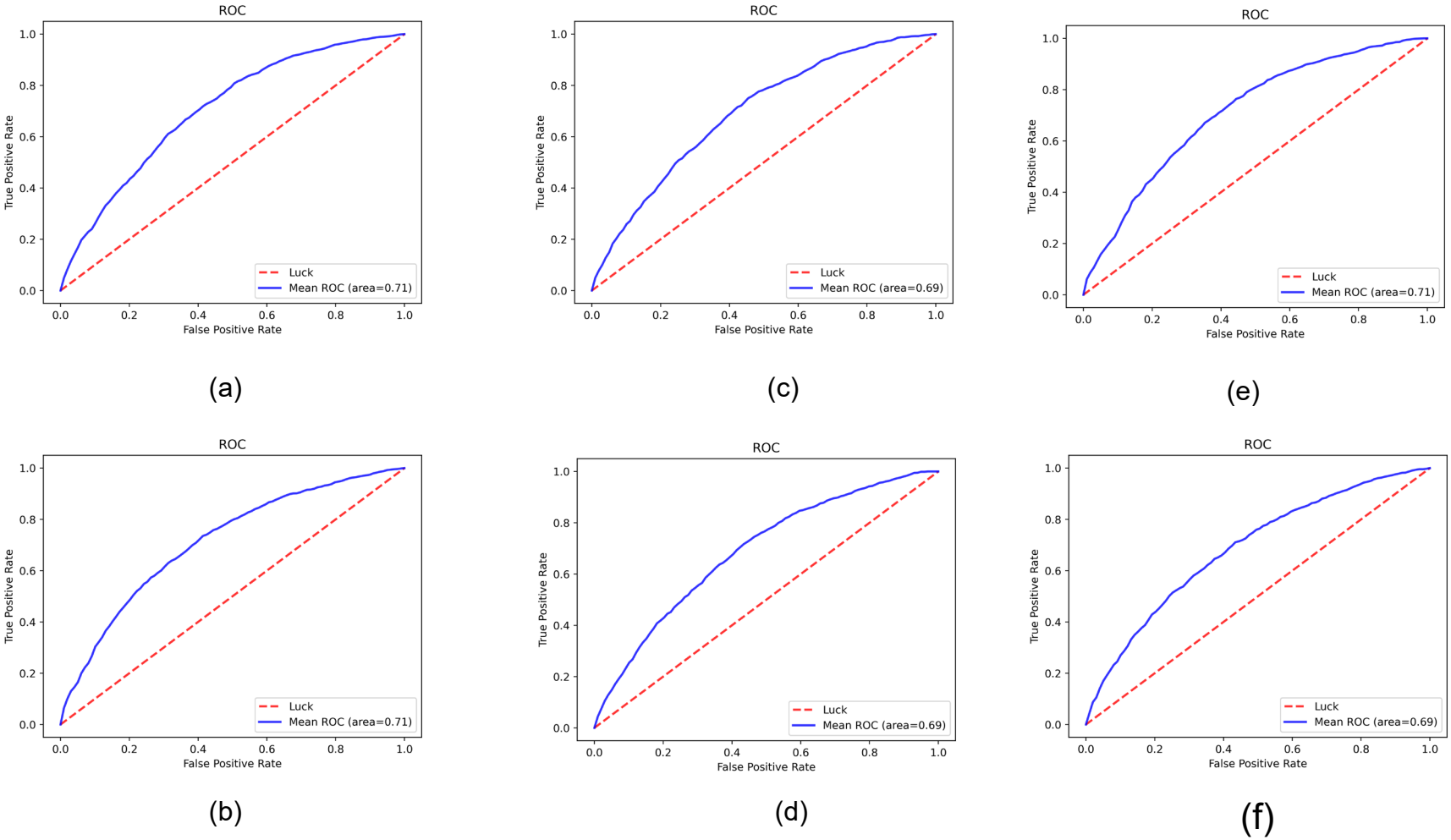


Figure 20 ROC curves of each incidence prediction model on datasets

(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)

Data Mining Results - ROC

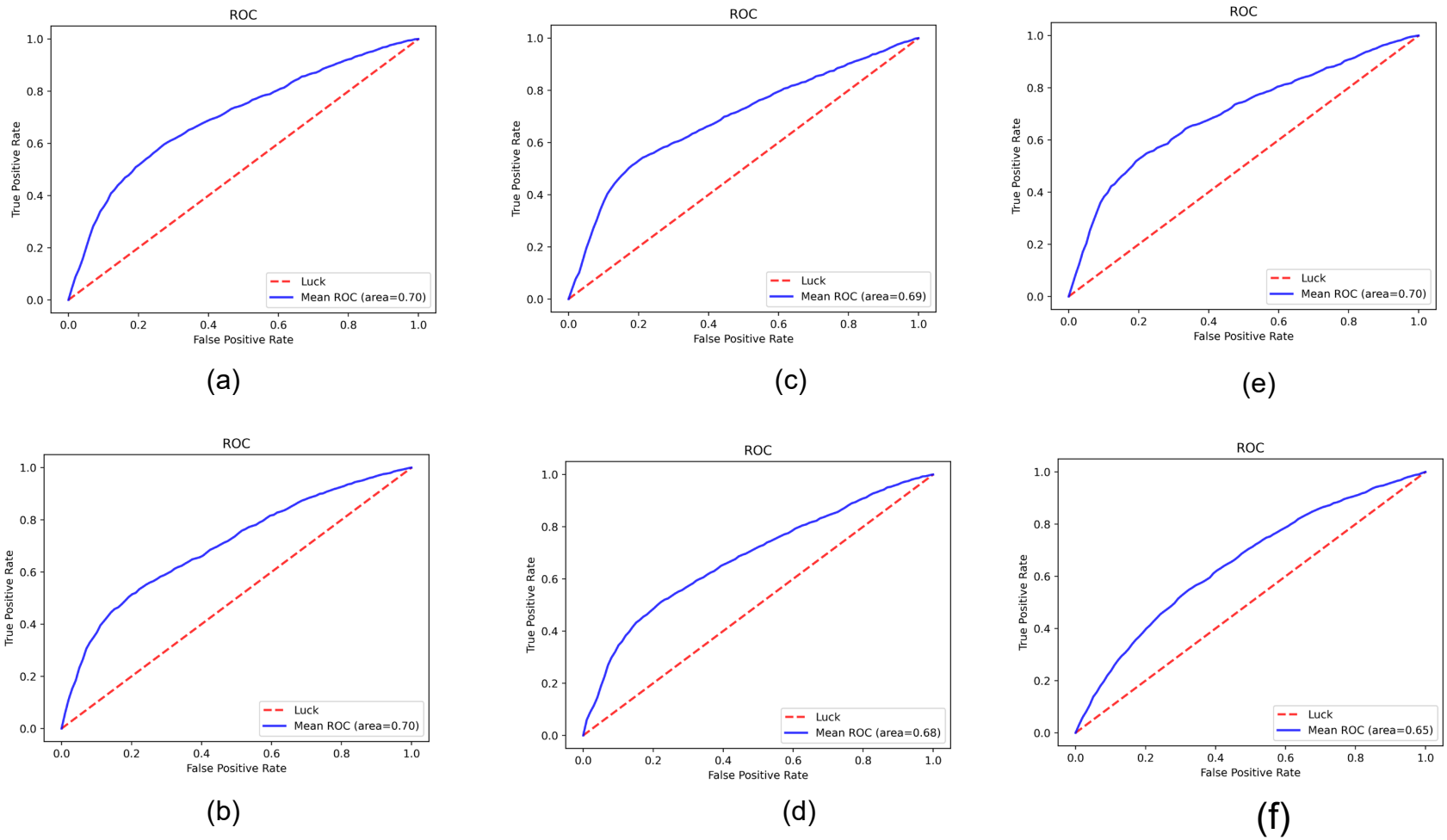


Figure 21 ROC curves of each progression prediction model on datasets

(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)

Data Mining Results – Pair-wised Independent t-test

Table 6 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 7 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 8 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

Data Mining Results – AUC with Axial length to corneal radius ratio

Table 9 AUC results of each model on datasets on three tasks with Feature Axial length to corneal radius ratio

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

Data Mining Results – Pair-wised Independent t-test

Table 10 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 11 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 12 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

Data Mining Conclusion

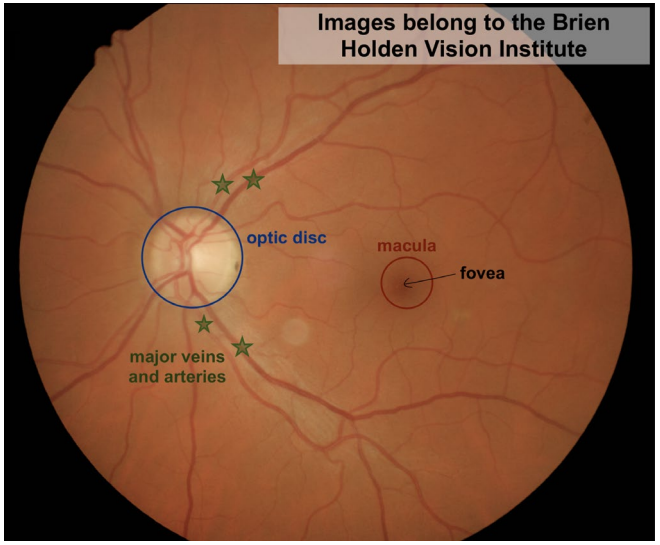
1. **The results of the LGBM model are significant different from the other methods. As can be seen, the outcomes of each model are distinct.**

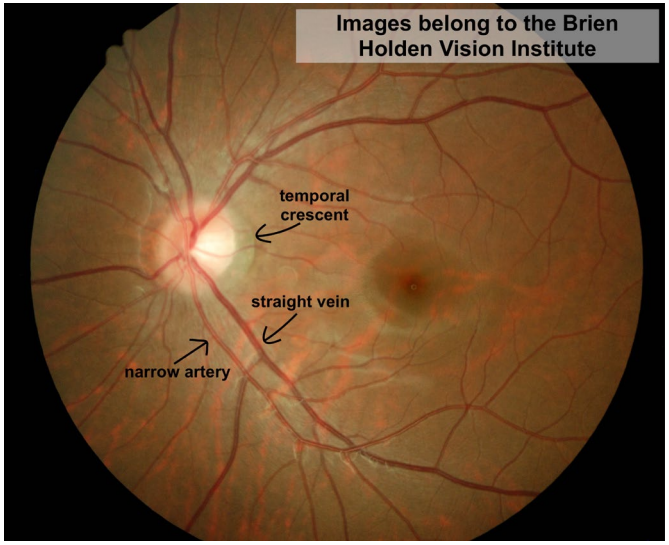
2. **The data mining model performs marginally better in the myopia detection task than the Logistic Regression model, but slightly worse in the myopia incidence and progression prediction tasks. Notably, the data mining model can eliminate the need for tedious manual feature selection.**

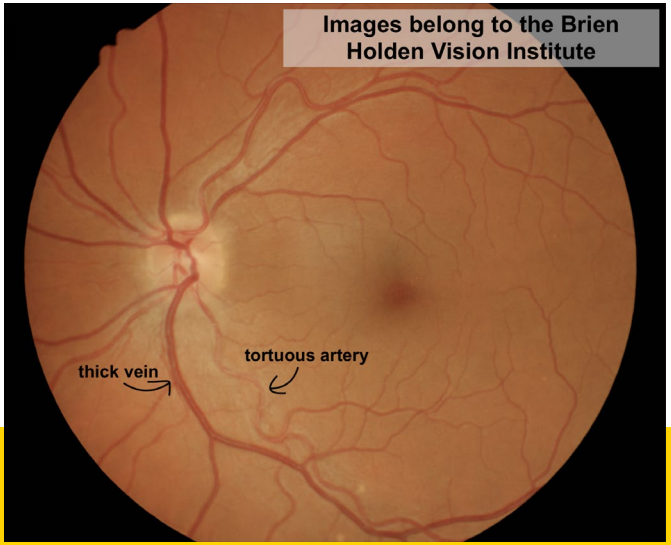
3. **In comparison to results that do not include the feature axial length to corneal radius ratio, the LGBM model's AUC increases by 7%, 10%, and 14% in the detection, incidence, and progression prediction tasks, respectively. This is consistent with the results of clinical research.**

Deep Learning Methods

Data Set – Retinal Images







2995 participants' retinal images in month 0 and their refractive error labels in month 0 and month 12. [2995 × 3]

Deep Learning Methods - Data Augmentation

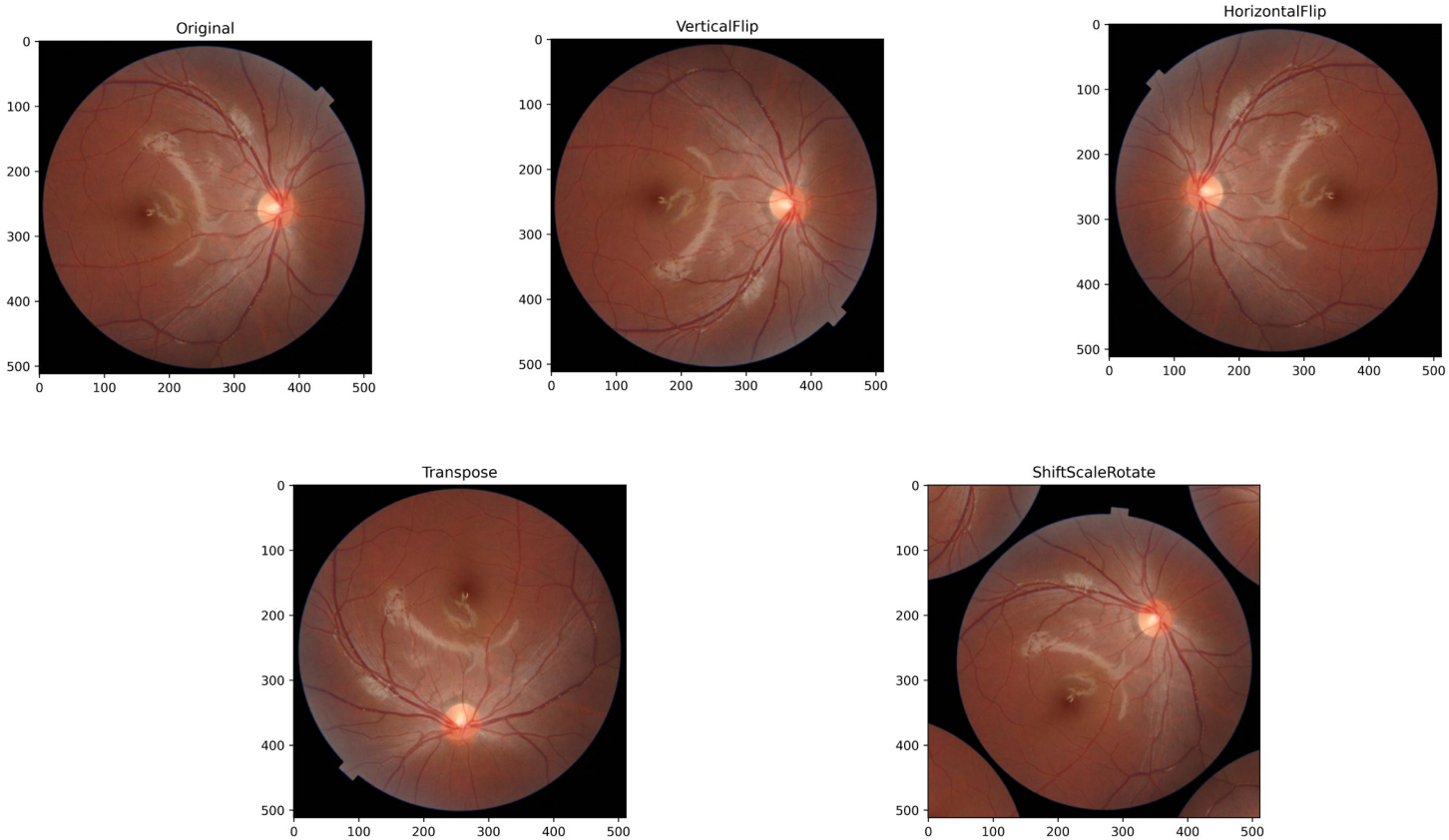


Figure 22 Retinal Images Geometric Transformation

Deep Learning Methods - Data Augmentation

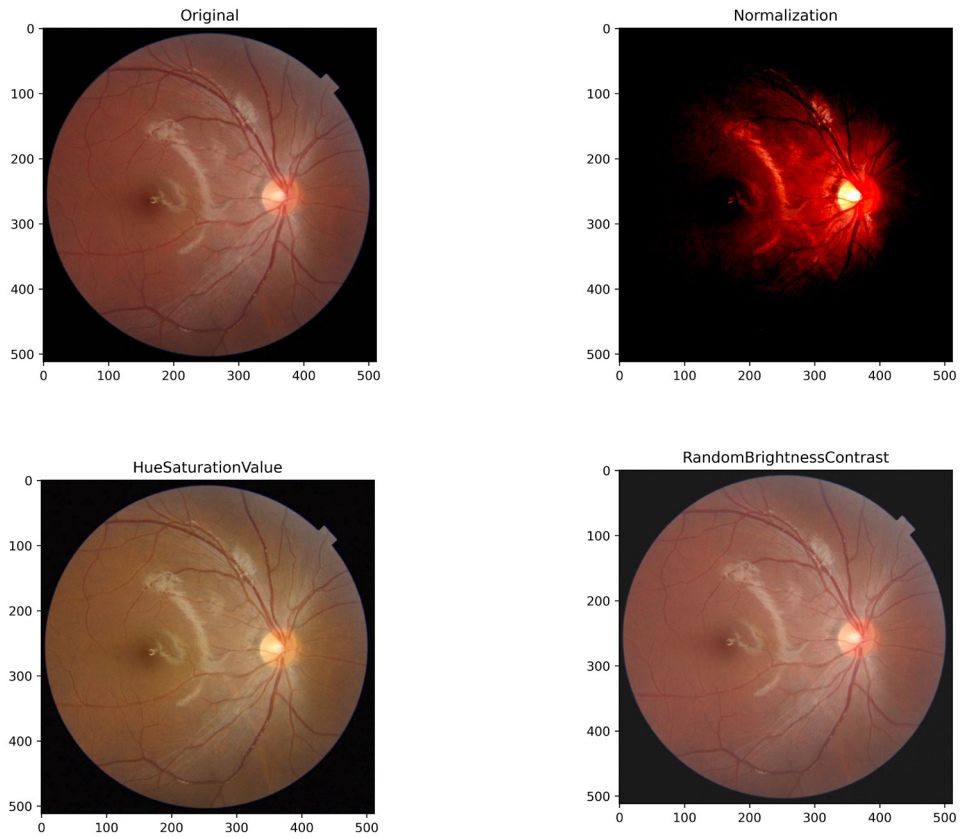


Figure 23 Retinal Images Colour Transformation

Deep Learning Methods - Data Augmentation

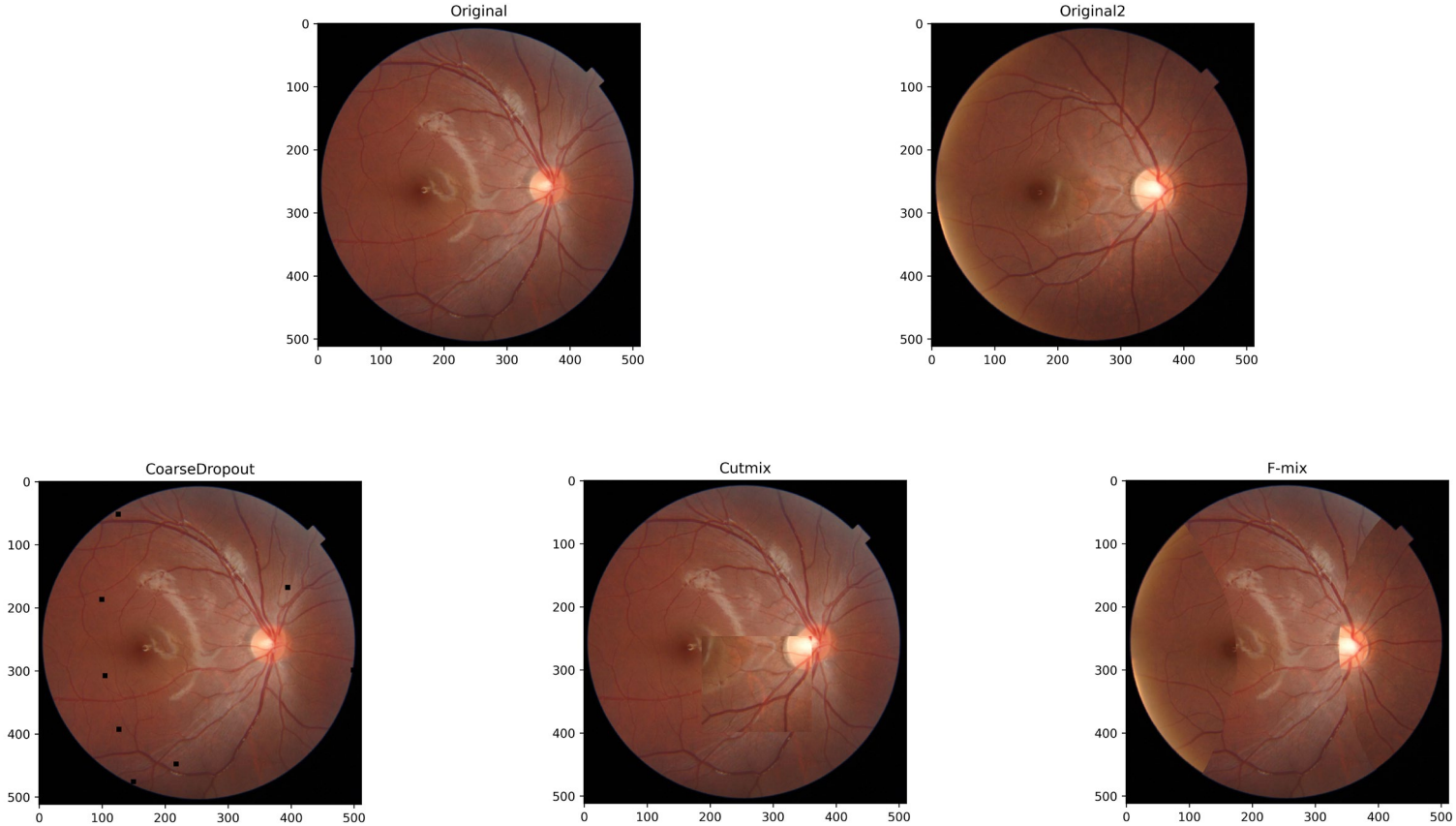


Figure 24 Retinal Images cut and mix operation

DNN image Classification

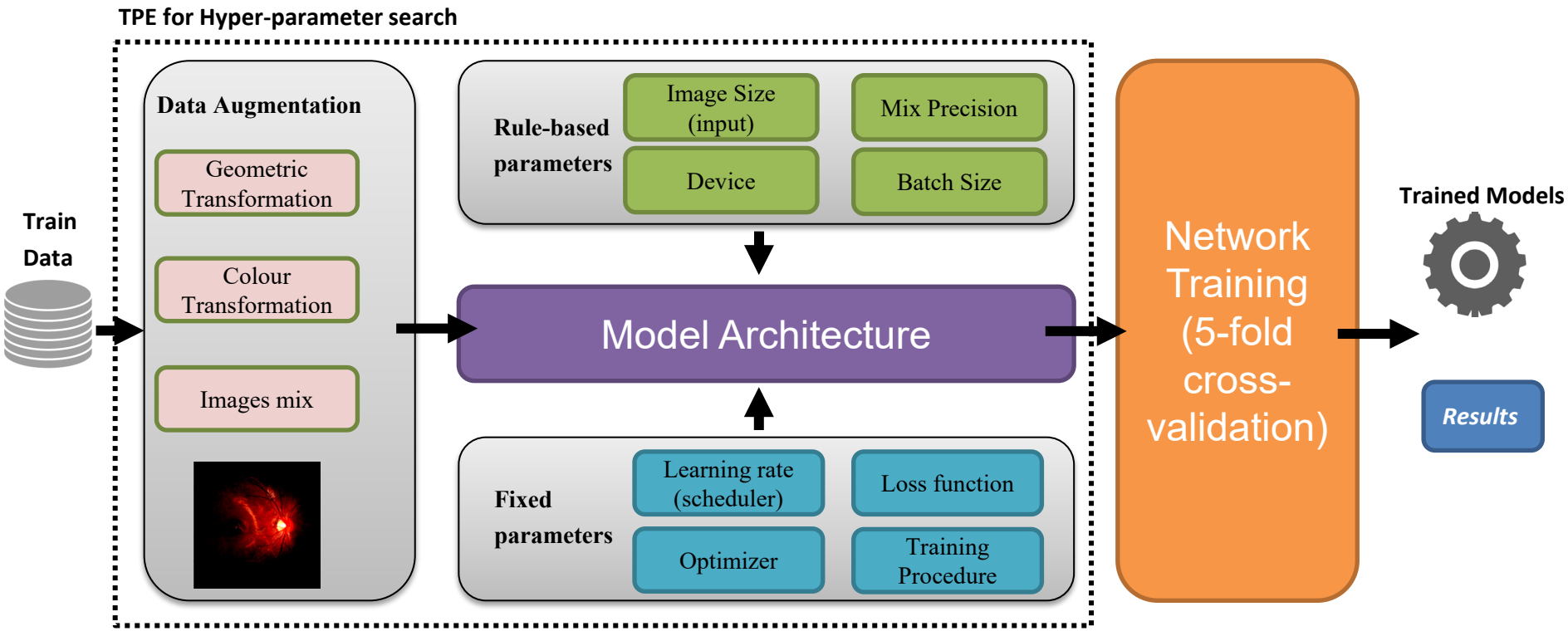


Figure 25 Pipeline Flow for Image Data Training

DNN image Classification

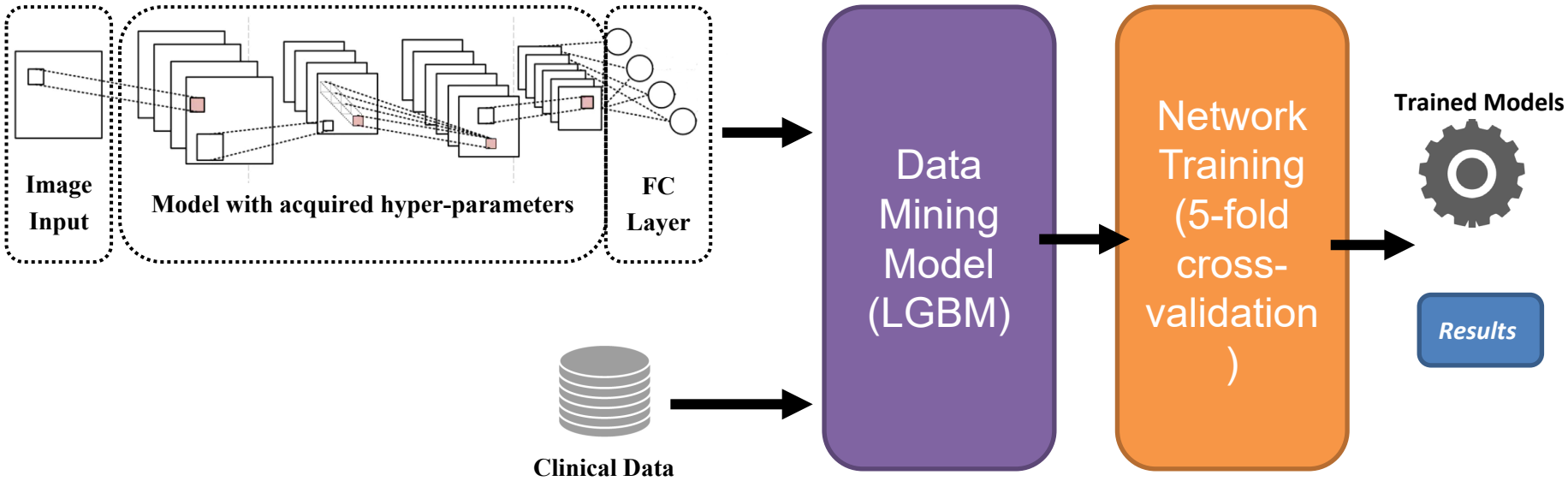


Figure 26 Pipeline Flow for Multi-modal Data Training

DNN image Classification

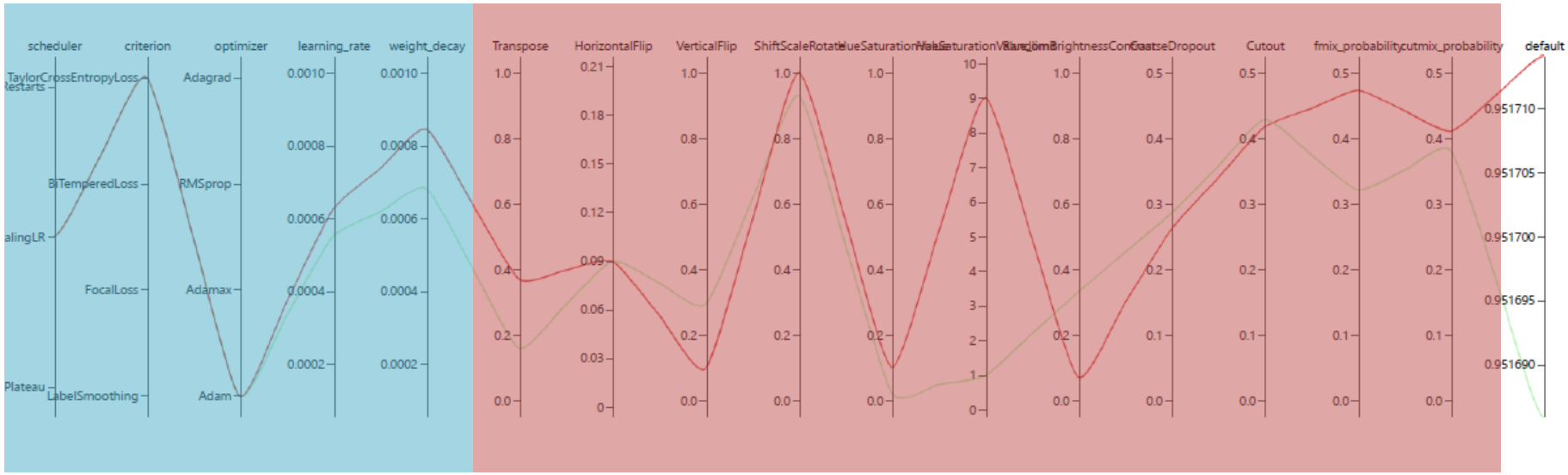


Figure 25: Hyper-parameter search with TPE Algorithm

Result – AUC for NN image prediction

Table 13: 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

Result – ROC for NN image prediction

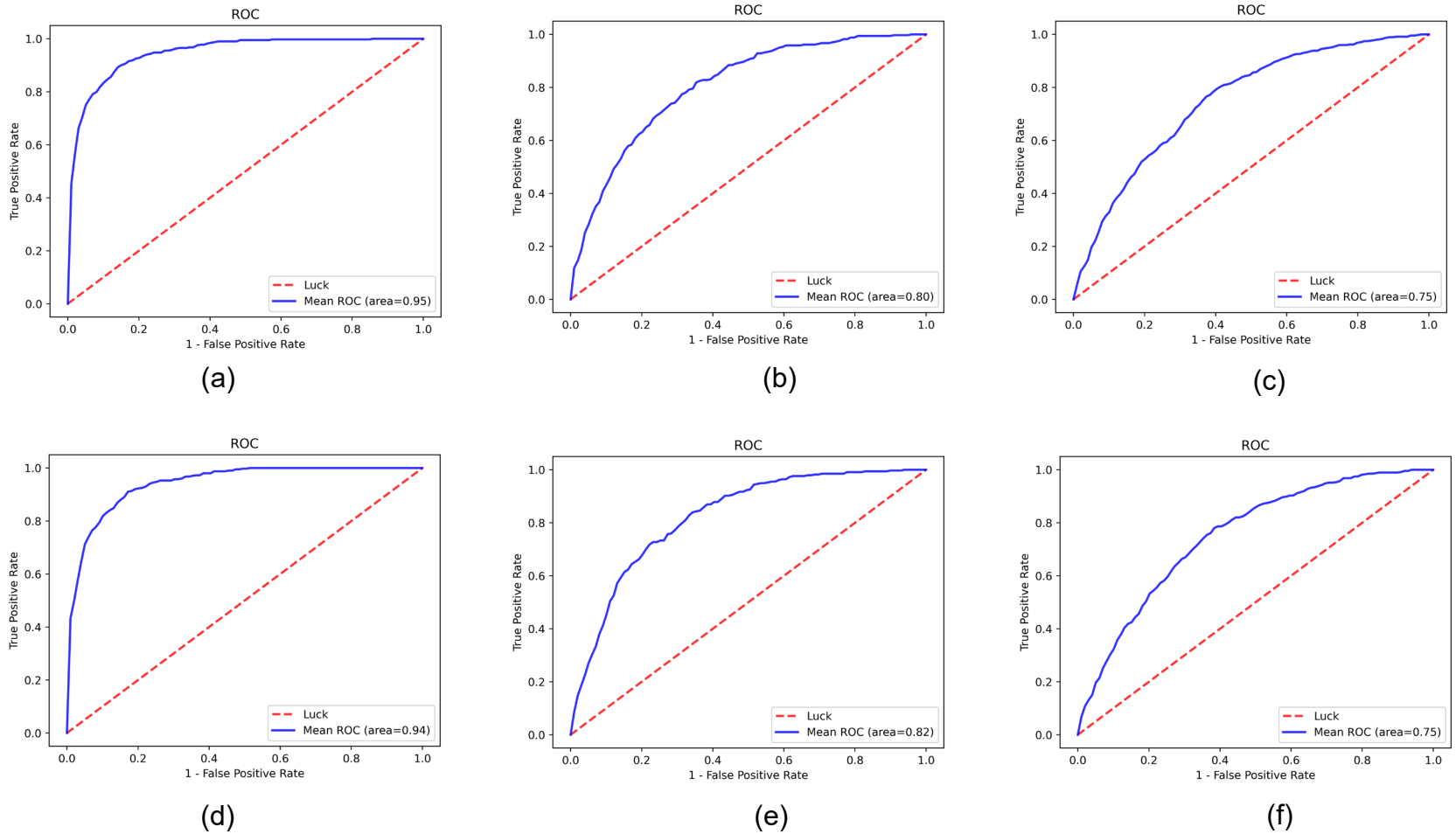
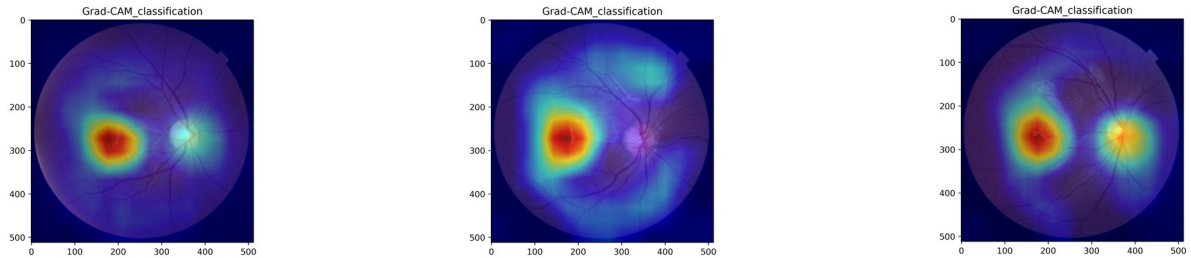


Figure 26: ROC curves of each model on image data of three tasks

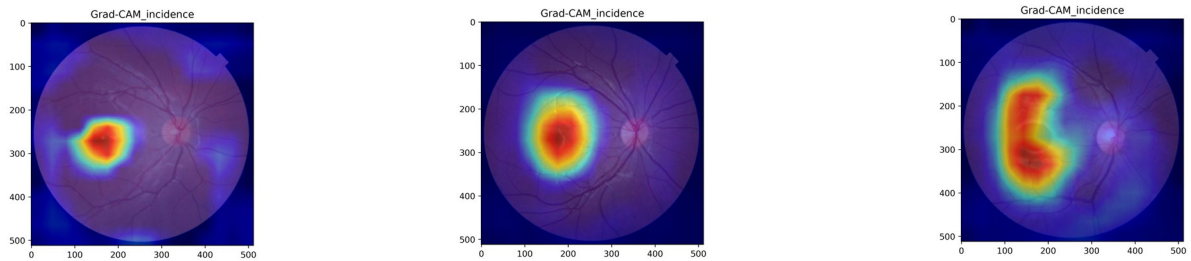
(a ,b ,c: Resnet models ROC on detection, incidence and progression prediction tasks respectively. d, e, f: EfficientNet-B4 models ROC on detection, incidence and progression prediction tasks respectively)

Result – Visualization – GRAD-CAM

Resnet Detection:



Resnet Incidence:



Resnet Progression:

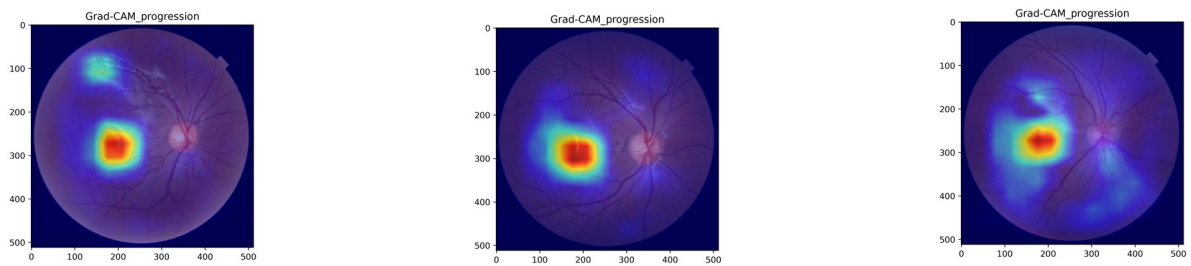


Figure 27: GRAD-CAM visualization on Resnet models of three tasks with different retinal images

Result – AUC for NN multi-modal data prediction

Table 14: 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

Result – ROC for NN multi-modal data prediction

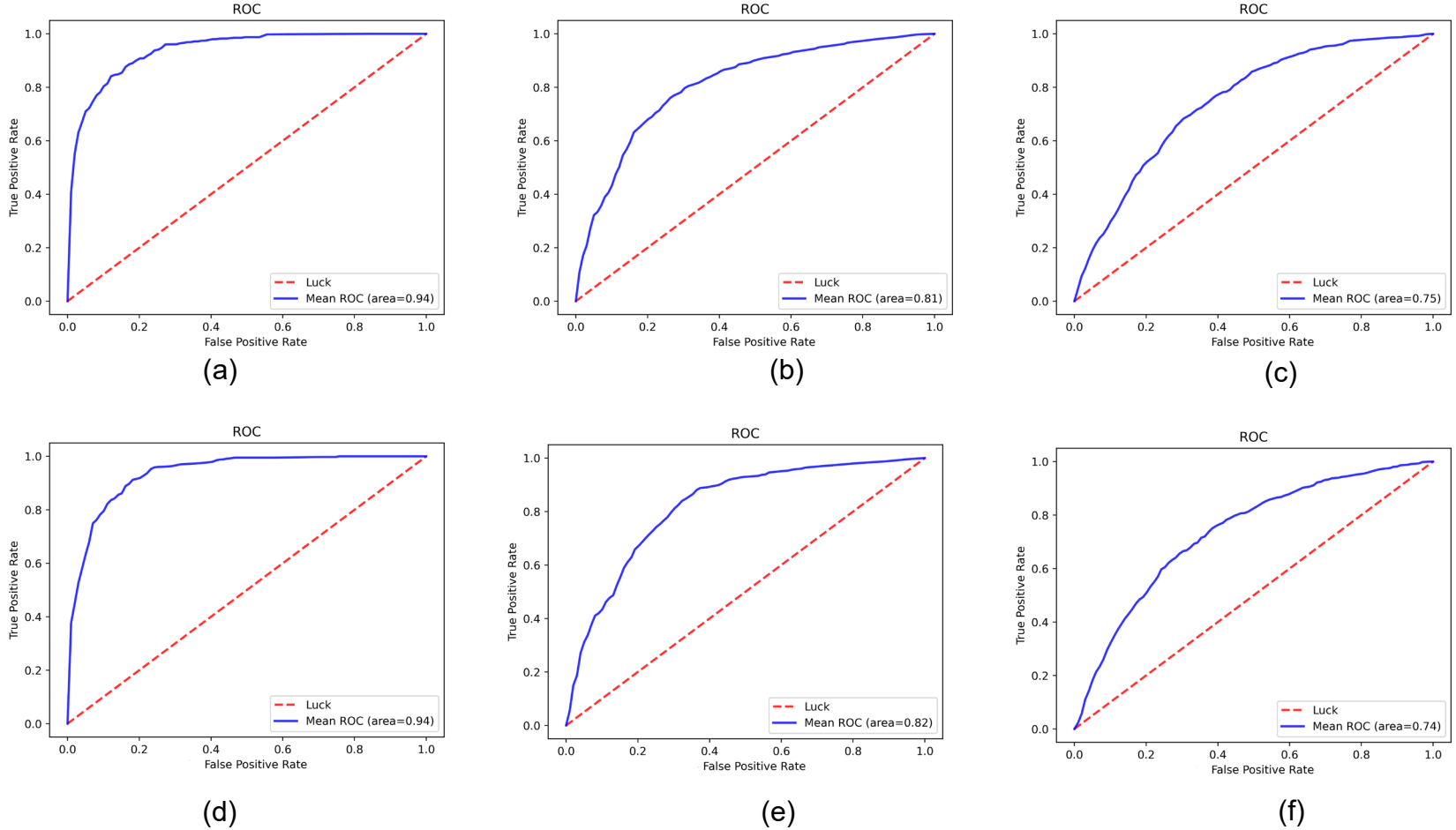


Figure 28: ROC results of EfficientNet-B4 model on different datasets of three tasks

(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)

Deep Learning Conclusion

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 proves that the neural network learns the information that aligns with the Optometry research**

4. **Based on the results of multi-modal datasets, it was inferred that clinical data had limited effects.**

Future Works

- **Retrain the Deep Learning Model with same dataset as Data Mining Models for comparison.**
- **T-test for Deep Learning results.**
- **Pre-train model with public dataset.**
- **Test data on different data set.**

Thank you for your kind attention.