

An Early Detection of Retinopathy of Prematurity on Fundus Image using Deep Learning

By
Sazzad Hossain
213-15-4305

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised By

Ms. Shayla Sharmin
Lecturer (Senior Scale)
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Mr. Md. Shah Jalal
Lecturer (Senior Scale)
Department of Computer Science and Engineering
Daffodil International University



**DAFFODIL INTERNATIONAL
UNIVERSITY**
Dhaka, Bangladesh

September 16, 2025

APPROVAL

This Project titled “An Early Detection of Retinopathy of Prematurity on Fundus Image using Deep Learning”, submitted by Sazzad Hossain, ID No: 213-15-4305 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 16 September, 2025.

BOARD OF EXAMINERS



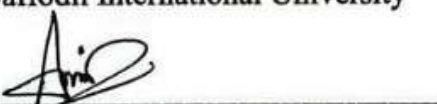
Dr. Arif Mahmud
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



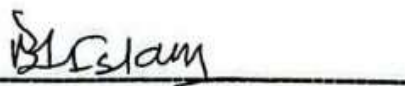
Dr. Md. Ali Hossain
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Mr. Amir Sohel
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Md. Manowarul Islam
Professor
Department of Computer Science and Engineering
Jagannath University

External Examiner

DECLARATION

I hereby declare that this project has been done by us under the supervision of **Ms. Shayla Sharmin, Lecturer (Senior Scale)**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Shayla 14.9.25

Ms. Shayla Sharmin

Lecturer(Senior Scale)

Department of Computer Science and Engineering

Daffodil International University

Co-Supervised by:

Mr. Md. Shah Jalal

Lecturer(Senior Scale)

Department of Computer Science and Engineering

Daffodil International University

Submitted by:

Sazzad 14.09.25

Sazzad Hossain

Student ID: 213-15-4305

Department of Computer Science and Engineering

Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. I am deeply grateful to everyone who has assisted me in one way or another.

First, I express my heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for me to complete the **Final Year Design Project(FYDP)** successfully.

I am grateful and wish our profound indebtedness to **Ms. Shayla Sharmin, Lecturer (Senior Scale)**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of my supervisor in the field of **Deep Learning** to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

I would like to express my heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing my project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

I would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

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

ABSTRACT

Retinopathy of Prematurity (ROP) is one of the major causes of blindness in preterm babies, and early screening is important to prevent the loss of vision forever. Conventionally based diagnostic procedures depend on manual screening by ophthalmologists, which is time-intensive and demands expertise and professionalism. This study suggests an automated machine learning-based early ROP detection pipeline built on convolutional neural networks (CNNs) and transformers. The fundus image dataset was trained using the model with data augmentation methods, to adjust the imbalance in classes and increase generalization. Focal loss served to make the models pay attention to hard-to-detect cases of ROP at an early stage. The models were tested on accuracy, precision, recall, F1-score and Area Under the Curve (AUC). The findings indicated that the single models, i.e. the ResNet50 (85.84%), the ResNet101 (89.38%), the MobileNetV2 (85.84%), the EfficientNetV2 (91.15%), the ViT-B (88.12%), and the DeiT (89.09%), worked well, but the ensemble model between the ResNet101 and DeiT had the highest accuracy of 94.69%. The proposed system is a validated and automated system to detect ROP in healthcare, which may benefit healthcare professionals, particularly in low-resource environments to save on diagnostic time and increase accuracy, eventually leading to improved patient outcomes and alleviating the worldwide burden of ROP-induced blindness.

Table of Contents

Approval	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
List of Figures	vii
List of Tables	viii
1 Introduction	1
1.1 Introduction.....	1
1.2 Motivation	3
1.3 Objectives	3
1.4 Methodology	4
1.5 Project Outcome.....	5
1.6 Organization of the Report	6
2 Background	7
2.1 Introduction.....	7
2.2 Literature Review	7
2.3 Gap Analysis	11
2.4 Summary	11
3 Research Methodology	13
3.1 Methodology/Requirement Analysis & Design Specification.....	13
3.1.1 Overview	13
3.1.2 Proposed Methodology/ System Design	14
3.1.3 Data Preparation.....	15
3.1.4 Functional and Nonfunctional Requirements.....	16
3.1.5 Data Flow Diagram	18
3.1.6 UI Design.....	19
3.2 Detailed Methodology and Design.....	20
3.3 Project Plan	22
3.4 Task Allocation.....	23
3.5 Summary	23

4	Implementation and Results	24
4.1	Environment Setup	24
4.2	Testing and Evaluation/Performance/ Comparative Analysis	25
4.3	Results and Discussion	29
4.4	Summary	30
5	Engineering Standards and Design Challenges	32
5.1	Compliance with the Standards	32
5.1.1	Software Standards	32
5.1.2	Hardware Standards	32
5.1.3	Communication Standards	33
5.2	Impact on Society, Environment and Sustainability	33
5.2.1	Impact on Life	33
5.2.2	Impact on Society & Environment	33
5.2.3	Ethical Aspects	34
5.2.4	Sustainability Plan	34
5.3	Project Management and Financial Analysis	35
5.4	Complex Engineering Problem	36
5.4.1	Complex Problem Solving	36
5.4.2	Engineering Activities	38
5.5	Summary	39
6	Conclusion	40
6.1	Summary	40
6.2	Limitation	41
6.3	Future Work	41
	References	43

List of Figures

1.1	Abnormal Blood vessel growth in Eye	2
1.2	Summary of methodology	5
3.1	Proposed Methodology	14
3.2	Sample Images of the dataset	15
3.3	Data Flow Diagram	18
3.4	The User Interface Design	19
4.1	ResNet101 Training & Validation accuracy and loss Curve.....	26
4.2	EfficientNetV2's Training & Validation accuracy and loss Curve	26
4.3	Ensemble's Training & Validation accuracy and Loss Curve.....	26
4.4	ROC curve for 5 top models	27
4.5	Fig 4.5: Confusion Matrix of Ensemble Model	28
4.6	Accuracy Comparison of the trained Models	29

List of Tables

2.1	Summary of Literature Review.....	10
3.1	Original Dataset images class distributions.....	15
3.2	Technique of Applied Augmentation.....	16
3.3	Project planning Timeline.	22
3.4	Task Allocation timeline.....	22
4.1	Performance of all the trained Models.....	25
4.2	Comparison with previous work.....	30
5.1	Mapping with complex problem solving.	36
5.2	Mapping with knowledge Profile.....	36
5.3	Mapping with complex engineering activities.	37

Chapter 1

Introduction

This chapter introduces the research problem, motivation, and objectives that provide a brief overview of the methodology and expected outcomes.

1.1 Introduction

Retinopathy of Prematurity is an eye disease or vascular disorder that affect preterm and low birth weight infants. It is caused by abnormal vessel growth in the retina which lead to vision loss or blindness if not identified and treated. There are five different stages of this disease(stage1-5). In stage 4 and 5, it's very serious condition for the babies. In those cases, baby should need better treatment or even needs surgery. So that's why the doctor actually prefers the treatment in stages 1, 2 and 3. For early detection through routine screening is essential, but there remain challenges for limited availability of experts.

Current clinical practice is based on trained ophthalmologists manually analyzing fundus images to diagnose ROP. These images are used to determine the condition of the retina, and show abnormalities that can signal the presence of ROP. The expert clinicians recognize signs of ROP and this process is time consuming and prone to human error, particularly in early stages of the disease when the changes in the retina can be subtle. In many low and middle income countries, where access to trained professionals is limited and lack of early diagnosis worsens the problem, resulting in higher rates of blindness and other complications.

The early and correct diagnosis of ROP in low-resource environments is one of the primary issues of the healthcare system, as it is gaining increased importance. The traditional diagnosis method is no longer enough with the increasing number of preterm babies. This scenario requires an automated and scalable solution that can be expanded to the underprivileged regions with no access to qualified ophthalmologists.

By applying Deep learning in the particular medical image analysis that offers a promising approach to overcoming this challenge. Applying Deep learning models such as convolutional neural networks (CNNs) and Transformer learning have been shown to be extremely successful in many medical imaging applications such as detecting glaucoma, diabetic retinopathy and even skin cancer. So these models can take images and look at them at a microscopic level and spot patterns that hard to see with the human eye. By applying these methods to fundus images, ROP detection can be automated to make the process quicker, more efficient and available.

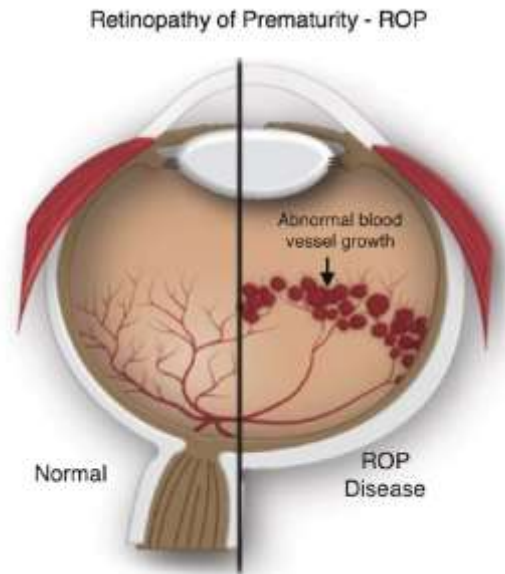


Fig 1.1: Abnormal Blood vessel growth in Eye.

An automated deep learning-based ROP detection system can greatly save the diagnosis time and effort, especially in the area where the ophthalmologists are insufficient. Such a system would provide early detection so that intervention and treatment could be provided in time to prevent severe cases of blindness. In addition, it can help decrease the amount of work for healthcare professionals by giving them a reliable tool with which to aid in the diagnostic process.

In this research, I am going to create a deep learning-based system for the early detection of ROP from fundus images. State-of-the-art deep learning models, ResNet50, ResNet101, MobileNetV2, EfficientNetV2, ViT-B and DeiT will be explored to see how well they detect ROP. Besides, an ensemble model composed of ResNet101 and DeiT is also discussed to further improve accuracy. To mitigate the effects of class imbalance in medical datasets, we use focal loss, a loss function that differentially focuses on hard-to-classify samples. Furthermore, a 2-stage fine-tuning method is adopted to optimize the performance of the models and thus make them more suitable for the specific task of ROP detection.

The contribution of this research is to add to the increasing field of medical image analysis, by offering a tool that can help healthcare professionals diagnose ROP early and accurately. We hope to provide a scalable, automated solution that leverages deep learning to eventually positively impact patient outcomes and limit the strain on healthcare systems around the globe.

1.2 Motivation

Retinopathy of Prematurity (ROP) is the major cause of blindness in premature infants, especially in developing countries where special care is not available. ROP is caused by abnormal vessel growth in the retina which lead to vision loss or blindness if not identified and treated in time. Early detection is needed for infants, because it means early intervention, which can stop the disease from getting worse. Nowadays, the present diagnostic approach to ROP involves the manual interpretation of fundus images by ophthalmologists, which is time consuming and possibility of human error. This is an even more difficult problem in areas where there aren't enough trained professionals.

In the most low resource settings, the healthcare systems are not capable to deliver appropriate screening for ROP, which results in missed diagnoses and increased of blindness. Recently, the rising rates of premature birth around the world is increasing, which is placing an additional burden on already stretched health care system. The need for skilled ophthalmologists to carry out the diagnosis means that ROP screening is available to only some infants, which means they can't provide early treatment for many.

With the need for a more efficient and scalable solution being high, automatic systems based on deep learning are a promising option. Although recent progress in medical image analysis has demonstrated the potential of deep learning methods, including Convolutional Neural Networks (CNNs), to aid with the diagnosis of eye diseases, existing automated ROP detection systems are still limited in terms of accuracy and usability.

1.3 Objectives

The overall goal of the study is to design an automated system with deep learning capabilities to early identify Retinopathy of Prematurity (ROP) through fundus images. The particular aims of this research are as follows:

To explore and implemetation various deep learning models: In this paper, we are going to research and apply various deep learning models, such as a convolutional neural network (CNN) and transformers, to show which one works best at ROP detection using fundus images.

To develop an ensemble model: The ensemble model will be created by combining of CNNs and transformer-based models, which can increase performance and improving the accuracy in the detection of ROP, especially in early cases when abnormalities are not evident.

To optimize model performance: Fine-tuning, two-stage training, and focal loss techniques will be used to boost model performance, especially in addressing class imbalance and sensitivity towards early-stage ROP.

To evaluate performance: There will be many metrics used to assess the performance of the models, including accuracy, precision, recall, F1-score, and Area Under the Curve (AUC) to determine the performance of the models in identifying the various stages of ROP.

In order to offer an effective and scalable solution to the problem of early ROP detection. It is desired to develop a system that can be implemented in actual clinical practice, making it adaptable to various healthcare conditions and enhancing the efficiency of the ROP diagnosis process, especially in those areas, where qualified specialists are not always available.

1.4 Methodology

The study follows some stages, a comparative approach towards constructing a robust deep learning pipeline to identifying Retinopathy of Prematurity (ROP) in fundus images at an early stage. The study is structured as follows:

1. **Data preparation and augmentation:** It consists of a pre-processing and curation of a collection of fundus images. Images are scaled and cut to fit and augmented including rotations, flips, and lighting changes. These augmentations are designed to match neonatal fundus imaging to maintain all the important pathological information without sacrificing model robustness under varying acquisition conditions and devices. This phase solves the problem of the lack of data and the imbalanced classes by increasing the effective sample diversity.
2. **Model Selection:** The research applied six deep learning models namely ResNet50, ResNet101, mobileNetV2, EfficientNetV2, ViT-B and DeiT. All models are optimized over a standardized protocol to create a reasonable baseline. The hyperparameters, input resolutions and evaluation splits are trained in such a way that they become comparable. Accuracy is the main performance measure, and the secondary measures are per-class precision, recall, F1 score, and confusion matrices to assess the sensitivity of each model to early or borderline disease detection.
3. **Imbalance-Aware Optimization and Staged Transfer Learning:** To address the problem of class imbalance and to focus on the examples that are difficult to classify, focal loss is used during the training. They also use a two-stage fine-tuning schedule: The classification heads are first adapted, and deeper layers with lowered learning rates are gradually unfrozen. The approach places ImageNet trained models on controlled and consistent transfer, reduces the risk of catastrophic forgetting, and improves generalization, especially when the size of the labeled data set is small.
4. **CNN-Transformer Ensemble:** Hybrid ensemble model is a combination of both ResNet101 and DeiT that seeks to strike a balance between the strengths of Convolutional neural networks (CNNs) and Transformers. This ensemble fuses local hierarchy attributes of CNNs with global attention attributes of the Transformer model, allowing the system to detect finer abnormalities of the vascular system, along with higher-order contextual statistics. At the probability level, fusion rules including averaging or weighted averaging are used depending on the validation performance and the decision thresholds are calibrated to maximize accuracy at early ROP screening.
5. **Evaluation Protocol:** To avoid leakage in the data, the data is tested on a set of retained data after the selection of the model on the validation data has taken place. The accuracy is used to evaluate the performance and error analysis is performed to identify the improvement in early-stage detection using confusion matrices and per-class analysis. Tests of strength (e.g., sensitivity to augmentation strength,

sensitivity to input resolution, etc.) are conducted to make decisions on trade-offs between accuracy and efficiency to apply in the real world.

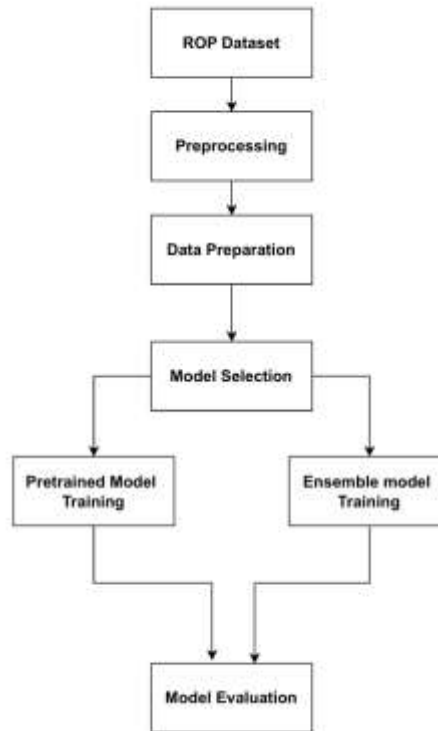


Fig 1.2: Summary of methodology

This approach is both rigorous and practical which systematic benchmarking sets robust single-model baselines, imbalance-conscious and sequential training improves minority class recognition and model stability and the hybrid ensemble improves accuracy to early ROP screening clinically important. The study will offer a realistic remedy to the issue of enhancing the accuracy of ROP detection in a variety of clinical settings.

1.5 Project Outcome

The potential outcome of this study is to design a deep learning system to effectively and efficiently identify Retinopathy of Prematurity (ROP) in fundus images with the intent to improve the chances of diagnosing it and providing better patient outcomes. In particular, the project is aimed at the following outcomes:

1. **Development of a robust deep learning model:** This study would pursue to build the most effective model to detect ROP in neonatal fundus images by deploying a variety of state-of-the-art architectures, including ResNet50, ResNet101, MobileNetV2, EfficientNetV2, ViT-B, and DeiT. A two-stage approach and focal loss will be used to improve the accuracy, sensitivity, and specificity of the models.
2. **Enhanced ROP detection through ensemble learning:** A ensemble model between ResNet101 and DeiT will be created to combine local hierarchical features, as well as global attention mechanisms to identify fine-grained vascular abnormalities and coarse-scale contextual patterns and offer a more competitive solution to early-stage ROP detection.

3. **Evaluation of model performance:** To check the effectiveness of the system to identify early or borderline ROP cases, a set of measures such as accuracy, precision, recall, F1-score, and the confusion matrices will be used to evaluate the performance of the system. The measurement of performance will be done on a held-out test set so that there may be no leakage of data and that the models may be generalized.
4. **Practical implications for early ROP screening:** The project is designed to offer a system that can be used by the healthcare professionals, especially those in the low-resource community to automate the process of identifying ROP. Not only this system will enhance the efficiency of the diagnosis, but also the workload on the ophthalmologists will decrease, and more infants requiring treatment will be identified faster and more accurately.
5. **Contributions to the field:** The study will add to the expanding literature on the use of deep learning in medical image analysis, especially in ophthalmology. The project will offer a baseline to future studies and implementation of deep learning in the detection of other eye diseases and medical conditions by solving issues like the nonexistent class balance, insufficient data, and variability in the acquisition.

In general, the projected result is a very precise, powerful, and scalable deep learning framework to identify ROP that could be implemented into medical practice to enhance early diagnosis and treatment results of preemies.

1.6 Organization of the Report

This study led to creating a deep learning-based system able to identify Retinopathy of Prematurity (ROP) when shown through fundus images, which will enhance the early diagnosis and patient outcomes. The objective of the project is to determine the most effective deep learning models in ROP detection by experimenting with different state-of-the-art architectures, such as ResNet50, ResNet101, MobileNetV2, EfficientNetV2, ViT-B, and DeiT, and fine-tuning them on a two-stage architecture with focal loss in order to achieve the best accuracy, sensitivity, and specificity. Moreover, an ensemble model based on ResNet101 and DeiT will be generated to incorporate both local hierarchical processing and global attention patterns to enhance the system capability of capturing subtle vascular abnormalities and wider contextual patterns in fundus images. To measure the effectiveness of the models in identifying early or borderline cases of ROP, the models will be assessed on a holistic set of measures such as accuracy, precision, recall, F1-score and confusion matrices. The analysis will be conducted in a held-out test set to guarantee that it is generalized and does not leak information. The most important consequence is the development of an automated system that can help healthcare professionals working in low-resource areas, where trained ophthalmologists can be in short supply. The system should help to identify infants who need intervention faster and more accurately by offering a tool that will increase the efficiency of the diagnostic process and decrease the workload of specialists.

Besides enhancing clinical performance, this study will support the field of medical image analysis at large, and specifically ophthalmology, by solving the perennial issues of class distribution, insufficient sample, and variability of acquisition. The results of this research can be used to build further research on deep learning application in detecting other ophthalmic diseases and other medical conditions and thus improving the use of artificial intelligence in healthcare.

Chapter 2

Background

This chapter introduces the Introduction, literature review, and gap analysis that provide a brief overview of the research background.

2.1 Introduction

Retinopathy of Prematurity (ROP) is the leading cause of blindness in childhood and occurs mostly in premature infants. It is a disease that results from the growth of abnormal blood vessels in the retina that can lead to retinal detachment and loss of vision if not caught and treated early. ROP that is detected early is important because treatment can mean no serious damage. However, today the diagnostic process is based on manual screening by trained specialists, which is time-consuming, requires expertise and can lead to misdiagnosis especially in early stages.

With the recent emergence of deep learning techniques, mainly Convolutional Neural Networks (CNNs), the interest in automation of the detection of ROP from retinal images is increasing. These methods have demonstrated huge potential in medical image analysis, such as detection of eye diseases, diabetic retinopathy, and glaucoma. Deep learning applied to ROP detection can increase diagnostic accuracy, decrease screening time and increase accessibility, especially in low resource settings.

In this chapter, the current ROP detection methods and deep learning applications in medical imaging are summarized. It also lists the challenges and gaps in the existing approaches, which lays the foundation for the proposed methodology in this study.

2.2 Literature Review

The National Eye Institute (NEI) defines retinopathy of prematurity (ROP) as a retinal condition in a premature infant (born before 31 weeks or less than 1,500 grams) that is caused by abnormal retinal blood vessels and is a leading cause of vision loss. Affecting 14,000-16,000 US infants each year, ROP has five stages, ranging from mild (stage 1, demarcation line) to severe (stage 5, retinal detachment). Nearly 90% of cases resolve spontaneously, but 1,100-1,500 infants require treatment such as laser therapy, cryotherapy or anti-VEGF injections to avoid detachment. Diagnosis is determined through fundus examinations beginning 4-6 weeks after birth and the frequency of exam is determined as the progression is observed. It is important to manage oxygen levels, but there are some long-term complications like myopia or glaucoma.

Huang Y-P et al. [3] built a deep convolutional neural network (CNN) for classification of early-stage retinopathy of prematurity (ROP) from 1975 high-quality retinal fundus

images (collected from 2013-2019, total 11,372 images) split into 10,235 training, 1,137 validation, and 244 testing datasets. Using data augmentation (e.g., illumination adjustment, rotation), the CNN with five convolutional layers, batch normalization, ReLU activation and softmax output provided 92.23% testing accuracy through fivefold cross-validation with sensitivity of 89.81-96.14, specificity of 94.50-98.99 and AUC of 0.92-0.96. Image quality or field-of-view-related misclassification was found to be 6.15%. Although the model has good potential for ROP screening, it is limited by its retrospective nature, small sample size, and lack of external validation/extended ROP stage inclusion. Omneya Attallah, "Diabetic retinopathy Automatic Retinopathy Prediction (DIAROP)," *Diagnostics*, vol. 5, pp. 100-105, 2021 (preprint: doi.org/10.1896/m8ac.171207). DIAROP is a deep learning algorithm for retinopathy of prematurity (ROP) detection using 17,801 retinal fundus images (8,090 diseased, 9,711 normal) from 890 preterm babies in 30 It adopts transfer learning of four pre-trained CNNs (ResNet-50, Inception V3, Xception, Inception-ResNet V2) for spatial feature extraction, Fast Walsh Hadamard Transform for spectral feature extraction and dimensionality reduction, and three integration methods (auto-encoder, PCA, DWT) for feature fusion. Using 5-fold cross-validation, DIAROP reached 93.2% accuracy, 89.7% sensitivity, 96.1% specificity, and an area under the receiver-operating characteristic curve (AUC) of 0.98 with DWT-integrated features, which was markedly better than standalone CNNs (e.g. ResNet-50: 91.48%) and other models (e.g. AlexNet: 77.9%). Training time was 404.17 s, which is faster than end-to-end CNNs.

In a systematic review and meta-analysis by Zhang et al. [5] published in the *Journal of Ophthalmology* in 2021, the diagnostic accuracy of deep learning (DL) algorithms for clinically diagnosing and grading retinopathy of prematurity (ROP) using fundus images is evaluated. Results: Overall, the nine studies included 15 classifiers and 521,586 images, with high pooled sensitivity (0.953, 95% CI: 0.946-0.959) and specificity (0.975, 95% CI: 0.973-0.977) and an AUC of 0.984, with an excellent performance. Subgroup analyses revealed strong performance both for detection (AUC 0.989) and classification (AUC 0.982) of ROP lesions, despite the presence of heterogeneity arising from non-threshold effects such as image processing and dataset variability. Although DL models appear promising in automated ROP screening, translational limitations such as lower clinical specificity, dependence on dataset and "black-box" nature of DL require further refinements for wider clinical application.

In a 2023 study in *Translational Vision Science & Technology*, Bai et al. [6] validated the deep learning algorithm ROP.AI for identifying plus disease in retinopathy of prematurity (ROP) with 8052 retinal images from 925 eyes from five tertiary centers of Australia (between May 2021 and February 2022). Trained on a single-center New Zealand dataset, ROP.AI obtained an area under the receiver operating characteristic curve of 0.75 for plus disease detection with an optimized sensitivity of 84%, specificity of 43%, and negative predictive value of 96% at an operating point of 0.38. Although trained only for plus disease, 83% sensitivity and 89% negative predictive value were achieved for pre-plus and disease combined. Importantly, the analysis was performed on unaugmented, unprocessed images to mimic real-world clinical settings, such that ROP.AI was viewed as a screening tool, even if the specificity and performance did not match the 0.99 AUC. Limitations include single-expert image-by-image grading and inter-centre variability, suggesting that further training be done to increase

generalisability for clinical use.

In a 2020 study published in *Eye and Vision*, Tong et al. [7] proposed an intelligent deep learning system based on a 101-layer ResNet and Faster R-CNN to automatically classify the severity and detect the stage and plus disease of retinopathy of prematurity (ROP), from 36,231 fundus images, collected at Renmin Hospital of Wuhan University from 2012 to 2016. The entire system had a sensitivity of 0.903 for four-level ROP severity classification (normal, mild, semi-urgent, urgent) with category specific sensitivities of 0.883, 0.900, 0.957 and 0.870, representing an improvement or equality of performance with two retinal experts (sensitivities of 0.902 and 0.898, respectively). It was also highly accurate in the detection of ROP stages (0.957, ranging from 0.876 for stage I to 0.999 for stage V) and plus disease (0.896). However, the single-center cohort, small number of stage V images, and the difficulty of identifying subtle early ROP indicate that multi-center validation and larger datasets of severe ROP are required to increase generalizability.

In a recent study in *JAMA Network Open* in 2021, Wang et al. [8] built J-PROP, a cloud-based, multidimensional deep learning platform for retinopathy of prematurity (ROP) screening, based on 52,249 retinal images from 14,108 eyeballs of 8,652 preterm infants (mean gestational age 32.9 weeks) from 4 centers in southern China in 2010-2019. The platform consisted of five classifiers: image quality, presence of ROP stage, intraocular hemorrhage, preplus/plus disease and posterior retina, with F1 scores: 0.718-0.981; sensitivities: 0.918-0.982; specificities: 0.949-0.992 and AUCs: 0.983-0.998. The aggregative classifier results of the RW ROP detection system resulted in F1 scores of 0.898-0.956, sensitivities of 0.981-0.986, specificities of 0.939-0.974, and AUCs of 0.9901-0.9956 at image, eye, and patient levels. DeepSHAP generated explainable heat maps, while performance of J-PROP (Cohen k 0.86-0.98) was similar to that of ROP experts (k 0.93-0.98). J-PROP is a promising technique for routine ROP screening, especially in resource-constrained settings, but needs further validation with data from different settings and with the consideration of computational and storage costs.

Kiran Yenice et al. [9] in a 2024 publication in *Eye*, developed a deep learning model using RegNetY002 to classify Type 1 ROP, Type 2 ROP and A-ROP in 634 fundus images of 317 premature infants with accuracies of 0.98 (Type 2 vs. Type 1/A-ROP), 0.90 (Stage 2 vs Stage 3) and 0.91 (A-ROP vs. Type 1) respectively, with AUCs of 0.98. While Wang et al. (2021) used a larger dataset (52,249 images) and obtained higher F1 scores (0.898-0.956) and AUCs (0.9901-0.9956) for referral-warranted ROP, Kiran Yenice et al.'s model is more narrow in scope, focusing on specific ROP types/stages without the explainability tools leveraged by J-PROP such as DeepSHAP. These studies demonstrate DL's potential to be used in ROP screening, however Kiran Yenice et al. have a smaller dataset and retrospective study which limits generalisability and further validation will be required for its use in clinical practice.

In a 2019 study, Tan et al. [10] published ROP.AI, a deep learning algorithm for ROP disease detection using 6,974 fundal images from Auckland (2006-2015) in *Translational Vision Science & Technology*. Trained on Inception-v3 CNN for 80% of images, it yielded 96.6% sensitivity, 98.0% specificity, and 97.3% accuracy (AUROC 0.993) internally, and 93.9% sensitivity, 80.7% specificity, and 95.8% NPV (AUROC 0.977) on 90 external images. Using an optimized cut-off of 0.38, sensitivity was 97.0% and NPV 97.8%. Although ROP.AI performs well when compared to other leading models such as J-PROP

(Wang et al., 2021) [larger dataset, AUC 0.9901-0.9956] or Kiran Yenice et al., et al. [smaller dataset, AUC 0.85-0.98], the single institutional data, overfitting, and focus on plus disease is limiting and the model needs multicenter validation for wider use.

Using transfer learning models, such as VGG-19, ResNet-50, and EfficientNetB5, and a Convolutional Neural Network (CNN) Salih et al. (2023) [17] attempted to define the stages of ROP in preterm babies. This study utilized 1365 fundus pictures collected in the Private Clinic Al-Amal Eye Center in Baghdad, Iraq. Using an ensemble approach with a voting classifier, they obtained a general accuracy of 88.82 with EfficientNetB5 giving the highest accuracy of 87.27.

Meanwhile, Redd et al. (2019) [18] were able to test the i-ROP DL system on plus disease and more broad categories of ROP with an AUROC of 0.960 to identify type 1 ROP. The data in this study used was 4861 eye examinations on 870 infants. This system has such potential because it is a sensitive system with a sensitivity of 94 and a negative predictive value of 99.7 as far as identifying early and accurate ROP and the resources used in the process are only limited as in the resource constrained environment.

Table 2.1: Summary of Literature Review.

Author (s)	Year	Methodology	Evaluation Metrics
Kiran Yenice et al.[9]	2024	RegNetY002	93%
Amelia Bai et al. [6]	2023	Inception-v3	AUC 0.75
Omneya Attallah et al.[4]	2021	DIAROP with DWT-integrated features	93.2%
Jingjing Zhang et al.[5]	2021	Deep learning Model(not mentioned)	AUC 0.98
Ji Wang et al.[8]	2021	J-PROP	AUC 0.98
Yo-Ping Huang et al.[3]	2021	Custom deep CNN	92.23%
Yan Tong et al.[7]	2020	R-CNN	90%
Zachary Tan et al.[10]	2019	Inception-v3	97.3%
Salih et al.[17]	2023	Ensemble model(VGG-19, ResNet-50, EfficientNetB, CNN)	88.82%
Redd et al. [18]	2019	Inception V1 (GoogLeNet)	AUROC 0.96

2.3 Gap Analysis

According to the reviewed studies and sources, there are several gaps in the existing state of the retinopathy of prematurity (ROP) detection that might be applicable to your work. The major gaps is the lack of diversity in the dataset. A large number of deep learning models used to detect ROPs are trained on regional datasets, i.e. New Zealand or Australia. It is possible that these datasets do not represent the range of different populations and clinical conditions in other geographical areas. Consequently, there is a limitation in the external validity of such models to other countries or other contexts, especially low-resource or rural contexts.

The another lack is the validation and generalizability. Most algorithms have been shown to be effective in their respective training sets, but they usually have not been validated in other populations or other previously unexplored environments. This restricts their performance across the various clinical environments or patient demographics. In practice, these models should be solid and capable, and to accomplish that, external validation is needed.

Another gap is that of multidimensional classification. Most of the algorithms currently existing are more oriented at either identifying the existence of ROP or determining its severity. But an integrated diagnostic device capable of focusing on several stages of ROP, plus associated complications such as plus disease, remain unavailable. The development of that tool would lead to a more accurate and detailed diagnosis and help clinical decision-making.

Finally, another area that needs more attention is early-stage ROP detection. Although most tools are very successful in detecting severe cases of ROP, there is less success in detecting subtle early signs of ROP. Early diagnosis is essential in preventing the escalation to higher levels and the creation of algorithms that identify the illness at its initial stages would be a welcome development into the industry.

2.4 Summary

This chapter gave a brief background to the issues and existing studies on Retinopathy of Prematurity (ROP) detection and the shortfall of the established diagnostics techniques that are still more or less reliant on manual examination of fundus images by trained experts. New automated solutions are urgently needed as the number of individuals with ROP is growing, and the supply of qualified and licensed ophthalmologists is insufficient and inadequate, particularly in low-resource countries.

This chapter also addressed special issues of ROP detection including small and skewed datasets, low-level abnormalities, and the difference in fundus images between devices and under diverse acquisition conditions. These limitations suggest that the design of deep learning should be structured such that it can be robust to support clinical cases in practice.

Moreover, the gap analysis revealed that specialized models are needed to cope with the problem of class imbalance and to improve the detection precision in the first stages of the ROP. It can be proposed that by fine-tuning the trained deep neural networks, ensembling, and by solving the problem of the imbalance in the classes, through the focal loss, which is recommended as part of the framework created in the context of the current research, one will be able to significantly narrow down these issues and make ROP detection itself significantly more accurate and efficient.

Finally, this chapter establishes the background to the proposed study, which aims at creating a deep learning-based system that can help in the early and precise identification of ROP, especially in underserved areas where specialized care is not accessible.

Chapter 3

Research Methodology

This chapter introduces the research methodology, motivation, and objectives that provide a brief overview of the methodology and expected outcomes.

3.1 Methodology/Requirement Analysis & Design Specification

3.1.1 Overview

The fundus-image-based early detection of Retinopathy of Prematurity (ROP) proposed methodology is expected to create a robust deep learning-based system to identify the ROP at an early stage. Multi-stage development process consists of data preparation (first stage) and model selection, fine-tuning, and evaluation (last stage).

System design aims to overcome the major challenges faced in ROP detection such as small labeled data, unequal classes, and variable image quality among different acquisition devices. The preprocessing of fundus images is the first step to normalize the input to the deep learning models, as well as augmentation strategies to learn variations in the clinic and generalize the model.

A number of deep learning models, such as ResNet50, ResNet101, MobileNetV2, EfficientNetV2, ViT-B, and DeiT, are discussed regarding their performance in ROP detection. To achieve this, these models are trained on the ROP dataset to pretrained models that had been previously trained on large datasets.

Rather than using either of the two CNNs (local feature extraction) or Transformers (global attention mechanisms), an ensemble approach is used to utilize the advantages of both. This hybrid approach improves the capacity of the system to identify not only minor anomalies, but also larger contextual patterns in the retina.

To deal with the issue of imbalance between classes, focal loss is added to the training to make the model pay more attention to the minority class, namely, early-stage ROP, which enhances the detection accuracy. The models are judged in terms of accuracy, precision, recall, F1-score, and AUC with an emphasis on detecting ROP in its early stages.

This research approach will allow creating a workable, adequate, and appropriate system of ROP detection to make positive contributions to health care providers, particularly in those facilities with inadequate access to dedicated care.

3.1.2 Proposed Methodology/ System Design

The objective of the proposed methodology is to create a deep learning-based system to detect Retinopathy of Prematurity (ROP) in fundus images automatically. This overall process can be broken down into a series of steps, including data collection and preprocessing, model selection, fine-tuning, evaluation, and deployment. This study uses the data that is available in ROP datasets, and the data is further expanded to improve model generalization. Transfer learning methods are used to utilize the pre-trained convolutional neural networks (CNNs) and transformers modified to the ROP-specific problem. To achieve performance accuracy and clinical interpretability, a full assessment plan is incorporated. Figure 3.1 represents the general process flow used in this study and shows all the steps of data preparation through model evaluation and deployment.

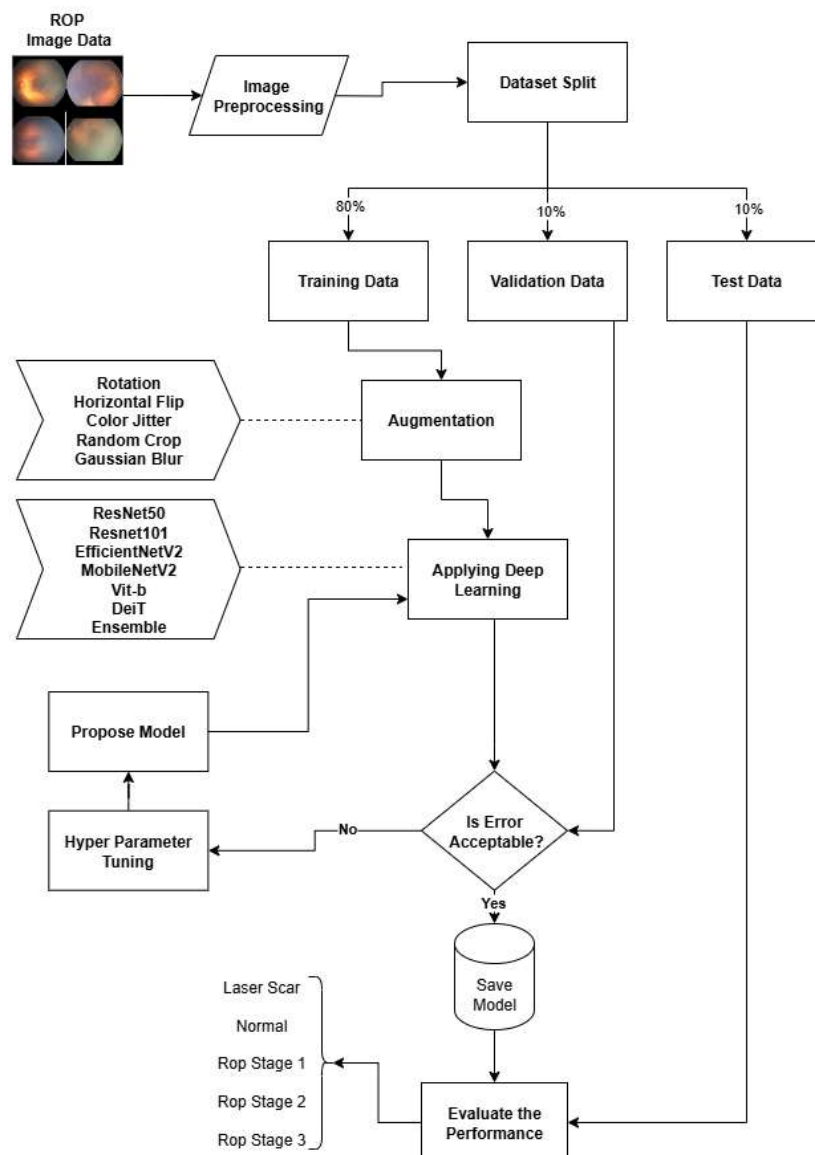


Figure 3.1: Proposed Methodology.

3.1.3 Data Preparation

Preparation of the data plays a crucial role in the construction of deep learning-based system to detect Retinopathy of Prematurity (ROP) in fundus images early. Process measures ensure that the data are in the correct format, the data has been adequately pre-processed, and the data has been enriched to enhance model performance and generalization.

Data Collection

In this study, the data is obtained through the Fundus Image Dataset of Intelligent Retinopathy of Prematurity System located in Figshare [1]. Their have five classes(Normal, Stage1, Stage2, Stage3 and Laser Scar) of 1099 images. These images represent a wide range of ROP conditions and a representative sample of clinical practice variability.

Tabel 3.1: Original Dataset images class distribution.

Class	Number of Images
Normal	236
Stage1	94
Stage2	165
Stage3	261
Laser Scar	343
Total	1099

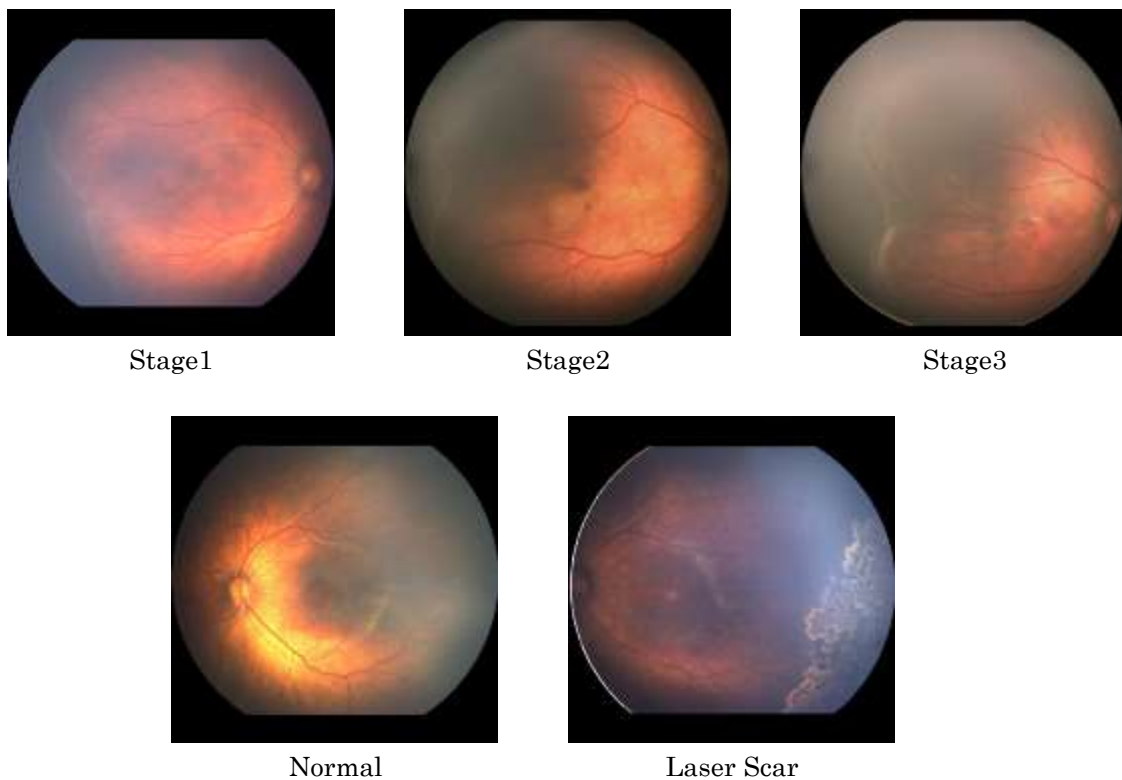


Fig 3.2: Sample Images of the dataset.

Dataset split and Resizing

To facilitate the appropriate training, validation and testing of the model, the dataset is separated on three parts, that is, 80% is used to train the model, 10% to validate and 10% to test the model. This division helps to train the model with alternate data, test it with unknown data so that it does not overfit, and subjectively test the model with the test set. In terms of image preprocessing, all fundus images have been downsampled to a fixed 224x224 size to fit the input of the deep learning models. This upsampling of data is used to make uniformity across the data which will allow easy and successful extraction of features as the model is being trained.

Data Augmentation

In order to improve generalization of the model and reduce the risk of overfitting, data augmentation methods are used to the fundus images. These methods mimic the variability that exists in the clinical environment, including changes in lighting, image rotation and various camera angles. Particularly, random adjustments such as rotations, flipping, zooming, and changes in brightness are implemented onto the pictures. The augmentation used on train image and generate one image to ten images. This greater diversity permits the model to reach a better set of properties, a better extrapolation to unobserved data, and an improved model to imaging conditions actually faced in the real world.

Table 3.2 : Technique of Applied Augmentation.

Technique	Parameter	Probability (p)
Rotation	$\pm 30^\circ$	0.5
Horizontal Flip	Applied	0.5
Random Brightness/Contrast	Brightness (0.2), Contrast (0.2)	0.3
Random Crop	Height: 224, Width: 224	0.3
Color Jitter	Brightness (0.2), Contrast (0.2), Saturation (0.2), Hue (0.1)	0.3
Gaussian Blur	Blur Limit (3, 7)	0.2

3.1.4 Functional and Nonfunctional Requirements

The construction of the deep learning-based system to identify the Retinopathy of Prematurity (ROP) at an early stage has both functional and nonfunctional requirements to enable the system to achieve the intended purpose of accuracy, reliability, and usability.

Functional Requirements:

Data Input and Preprocessing: the system must have the capability to receive fundus images in any format (e.g. JPEG, PNG). It needs to automatically scale the images and rearrange the pixel density and data augmentation techniques such as rotation, flipping, and altering the brightness levels to improve the models generalization.

Training and Fine-Tuning: It must be able to train and fine-tune most popular deep learning models, such as ResNet50, ResNet101, MobileNetV2, EfficientNetV2, ViT-B and DeiT. The models must be trained on a huge dataset such as ImageNet and then trained on the ROP dataset. The methods which should be used during the training process to address the problems of class imbalance and optimize the model performance are focal loss and two-stage fine-tuning.

Ensemble Model: The system must consider ensemble model in which the ResNet101 is coupled with DeiT. This framework needs to capitalize on both CNN-based local features extraction and Transformer-based global attention algorithms to enhance the accuracy of ROP detection.

ROP Detection:The system should be able to categorize the fundus images (ex: normal, early ROP, advanced ROP) automatically using the learned features. It must be in a position to identify small early-stage abnormalities of ROP and classify them correctly.

Performance Evaluation: The system should be able to assess the performance of the trained models with the help of different measures, such as accuracy, precision, recall, F1-score, and AUC. To evaluate how sensitive the model is to early and borderline stages of ROP the system should also produce confusion matrices and per-class measures.

Non-Functional Requirements:

Accuracy and Reliability: The system also needs to be very accurate, but more focused on early ROP detection. The system must be sound and the predictions should be the same and valid over the various datasets.

Scalability: The system should not be restricted to a specific amount of data, but it needs to be scalable to handle different volumes of data and hence process a large number of images. It must be capable of supporting fundus images, which have been captured by various acquisition devices of varying quality.

Real-time Processing: To be applicable in clinical practice, the system should be in a position to handle fundus images in real-time and give output within the required time. Processing time must be minimised so that it does not result in delays in diagnosis.

Usability: The interface must be easy to use with the system delivering an output that can be understood easily (classification labels and confidence scores). The Users who don't have much technical skill should find it easy to learn to use the system.

Robustness: The system should be robust to fluctuations in image quality like lighting, resolution and image acquisition conditions. The model is to be generalized reasonably well across images taken with various devices, or in various clinical settings.

Security and Data Privacy: The system should comply with applicable healthcare data privacy laws and regulations and guarantee the confidentiality and security of patient data. It should also encrypt its storage and transmission of data to keep medical secrets safe.

Maintainability: The system is supposed to be easy to maintain, simple to update, easy to fix bugs, and easy to incorporate new data. It must also be able to facilitate version control, so that any changes made to the system do not disrupt its functionality or performance.

Compliance with Medical Standards: The system must conform to regulatory medical standards and laws which are applicable to diagnostic devices in the context it is implemented in clinical practice. This could involve making sure that the system is correct and trustworthy by doing proper tests and clinical trials.

3.1.5 Data Flow Diagram

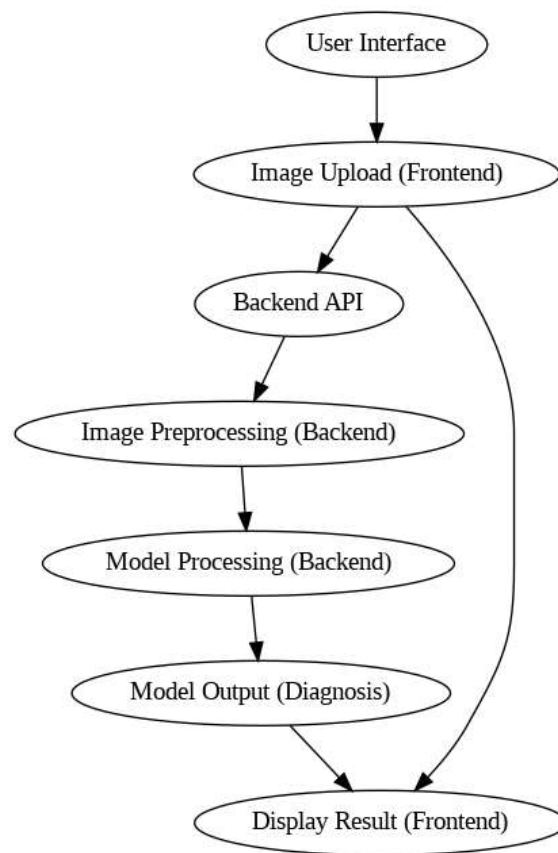


Fig 3.3: Data Flow Diagram

Data Flow Diagram (DFD) of the ROP Diagnosis System shows the data flow among the different system components. It demonstrates the way in which the user uploads an image through the User Interface, and the Backend API works on the image. That picture is subjected to Preprocessing and sent through to the Model Processing stage, where the diagnosis is produced. Last, the outcome is presented to the user on the Frontend. The graph illustrates graphically the interrelation between processes of the system and gives the clear picture of the flow of data and how it would be represented and converted in the system.

3.1.6 UI Design

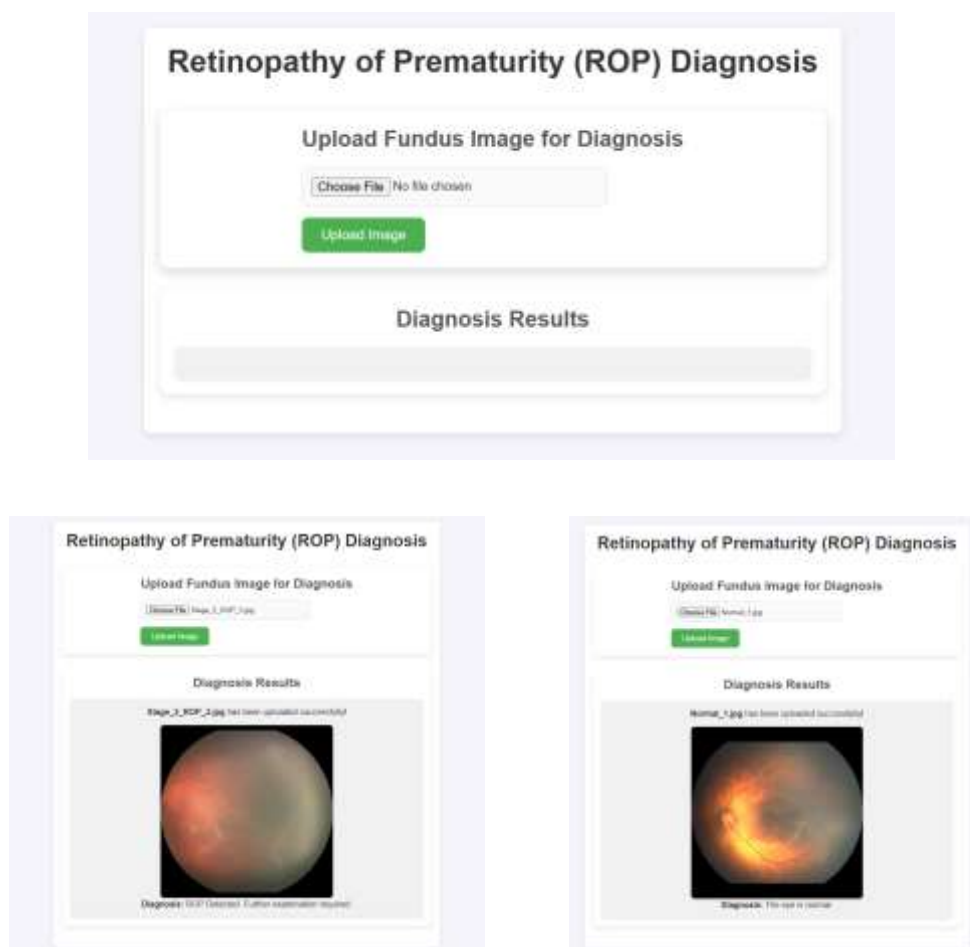


Fig 3.4: The User Interface Design.

Fig 3.3 shows the UI design of the Retinopathy of Prematurity Diagnosis System. The ROP Diagnosis System UI Design is clean and modern with a focus on simplicity and ease of use. It has an image upload area with a live preview and an easy to understand results area to display diagnosis. It has a responsive and visual design.

3.2 Detailed Methodology and Design

The design process and algorithm of the early detection of Retinopathy of Prematurity (ROP) based on deep learning entails a procedural effort comprising of a series of actions data collection and preprocessing, model selection and architecture, training and fine tuning, ensemble learning and model evaluation. All stages play an important role in the solution of the specific problems of ROP detection, including class imbalance, minor abnormalities at initial stages, and inconsistency in the quality of images.

Data Collection and Data Preprocessing

The ROP dataset is comprised of fundus images with labeled stages of ROP, both normal and other disease stages. These photographs are found in the clinical environments in which ROP screening is performed. Because of the scarcity of big ROP datasets, data collection is a component to be carefully considered. The data set would contain images of various acquisition device types so that they can be generalized to various clinical settings.

After collecting the dataset, the data are preprocessed and some steps involved are as follows:

- **Resizing:** The images themselves are resized to normal size (e.g. 224x224 pixels) to ensure that model input is constant.
- **Normalization:** To improve the model performance and convergence, the pixel values are normalized in the range 0 to 1.
- **Augmentation:** To model clinical variations associated with a real-world, the images are augmented with data augmentation methods, including random rotations, flips, zooming, and changes in brightness. This adds variety to the training data, so the model can generalize more and it reduces the chances of overfitting.
- **Class Imbalance Handling:** The data is typically skewed and there are few images of extreme stages of ROP. As a solution to this, methods like focal loss are used in the training process to emphasize the detection of minority class samples.

Model Architecture and Model Selection

In this research, several state-of-the-art deep learning models are chosen, as they have been proven efficient in the image classification problem:

- **ResNet50 and ResNet101:** The models are used because they have the ability to learn both the hierarchical features (deep features) that vanishing gradients would not, and complex patterns.
- **MobileNetV2, EfficientNetV2:** These are efficiency models which are selected after consideration of the efficiency and reduced cost of computation without any impact on their performance in image classification tasks.
- **ViT-B and DeiT:** ViT models with self-attention mechanism are added to extract global context in the images to augment the local feature extraction of CNN-based models.

These models have been trained on massive data sets such as ImageNet, granting them the opportunity to harness general image features which they have learned on a very large range of images. Retinal images are unique and it is necessary to fine-tune the models after pretraining on the ROP dataset.

Fine-Tuning and Optimization

In order to enhance the performance and sensitivity of the model to early-stage ROP, fine-tuning is performed in two steps:

- Stage 1: The classification heads of the models are first optimized to fit the ROP dataset. This can be used to train the models at a lower learning rate to prevent the degradation of the valuable knowledge gained during pre training.
- Stage 2: Once the initial fine-tuning is completed, the models are unfrozen progressively with a smaller learning rate to enable the models to acquire more complex and task-specific information without forgetting the information learned in pretraining.

Besides fine-tuning, focal loss is used to deal with the fact that, the dataset has an unequal number of classes. Focal loss puts greater emphasis on more difficult to classify samples, especially those belonging to the minority class (early ROP stages), which helps the model focus on identifying subtle retinal abnormalities.

Ensemble Learning

In order to enhance the detection accuracy, the ensemble model is developed by integrating ResNet101 and DeiT. The ensemble method has the advantage of combining the strengths of both the models:

- ResNet101: It is good at extracting local features, fine details of the retina including blood vessels and abnormalities related to ROP.
- DeiT : Seizes the global contextual patterns in the retina that matter in identifying more significant abnormalities or changes characteristic of ROP.

Model Evaluation

Once the models have been trained and fine-tuned they are put through rigorous testing on a held-out test set to test their generalization properties. The performance measures are:

- Accuracy: To estimate the general efficiency of the model in the detection of ROP.
- Precision and Recall: To test how well the model identifies the positive cases of ROP correctly (precision), and its sensitivity to detect ROP (recall).
- F1-Score: To trade-off between precision and recall and guarantee that the model does not suffer in any of the two.
- Confusion Matrix: To examine misclassification and determine the extent to which the model identifies various phases of ROP, specifically the ability to identify the first stage of ROP.

Generalization and Robustness

Finally, the models are tested to be robust to the alteration of inputs image resolutions. and augmentation strength. These tests guarantee the model will prove effective in various clinical setting and in image capture devices and are applied to control variability of real-life data. The sensitivity of the model to variation in the input conditions (e.g. resolution, lighting) is studied.

3.3 Project Plan

The project plan is a schedule and a guideline on the activities and time required to create the deep learning-based system that will identify Retinopathy of Prematurity (ROP) at an earlier stage. The time is separated into stages, which are related to certain activities that have to be accomplished to make sure the system is developed and evaluated successfully. The subsequent sections describe the stages of the project, as well as the activities and milestones of the stages.

Table 3.3: Project planning Timeline

Phase	Task	Duration
Planning	Literature review and requirements analysis	Week 1
Dataset Preparation	Collecting & preprocess dataset, apply augmentation	Week 2-3
Model Development	Building CNN and Vision Transformer implement pretrained models	Week 4-5
Training	Training the selected models.	Week 6-7
Evaluation	Model testing, performance comparison (Accuracy, Precision, etc.)	Week 8
Building ensemble model	Build ensemble model with CNN and transform learning.	Week 9
Final Evaluation	Model testing, performance comparison.	Week 10
Documentation	Report writing and presentation material	Week 11-12

3.4 Task Allocation

This table depicts the timeline of the principal activities in each period of the project, from week 12 to week 48.

Table 3.4: Task Allocation timeline.

Tasks	Weeks																		
	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48
Data collection phase	█	█	█	█	█														
	█	█	█	█															
Preprocess all the data						█	█	█	█	█									
						█	█	█	█										
Model training											█	█	█	█					
											█	█	█	█	█				
Applying Ensemble Model															█	█	█	█	█
															█	█	█		

3.5 Summary

This chapter presented a proposed methodology and system architecture of the early diagnosis of Retinopathy of Prematurity (ROP) with the help of deep learning. The algorithm includes various steps, such as preprocessing of data, choosing a model, training and fine-tuning, ensemble learning, and evaluation of performance. For handling the key issues including of imbalance of classes and the disparity of quality of images, the approaches including focal loss and data augmentation have been considered to make this system be stable and will generalize.

The system design borrows various deep learning models, such as ResNet50, ResNet101, MobileNetV2, EfficientNetV2, ViT-B, and DeiT, and uses an ensemble model with ResNet101 and DeiT to improve the accuracy of image detection. The models are optimised and scaled to provide high sensitivity specifically at the stage of ROP which is a critical characteristic of visual impairment prevention in preemies.

The development of the system project plan was provided and included all stages of data collection up to model deployment and evaluation. The time management ensures that all the details on the system have been addressed systematically and that data preprocessing up to end testing is integrated in order to give an efficient tool to identify ROP.

In summary, the method and design outlined in the present chapter represents a full-scale framework that can be implemented to establish an effective deep learning based system that will assist healthcare professionals in diagnosing ROP at a young age, which will subsequently result in better patient outcomes within neonatal care.

Chapter 4

Implementation and Results

This chapter introduces the research implementation and its result that provide a brief overview of the study and expected outcomes.

4.1 Environment Setup

The Retinopathy of Prematurity (ROP) detection deep learning system is trained on a local PC station (Dell Inspiron 153501) with Jupyter Notebook, and trained on the TPU/GPUs on the cloud at Kaggle. The models can be efficiently trained and evaluated in this environment.

1. Hardware Setup

The local hardware applied on the system is a Dell Inspiron 153501 which is configured with the following specifications:

CPU: Intel Core i5-1135G7, used for processing capabilities to run data preprocessing and augmentation.

RAM: 8GB DDR4 to support memory-intensive tasks effectively when preprocessing the data.

GPU: Kaggle uses cloud-based TPU/GPUs in training processes because the Dell Inspiron does not have its own graphics card that can be used to train large models.

2. Software Setup

OS: The development environment is based on Windows 10 (or a similar OS to run the local setup) operating system.

Python: The models and experiments are coded in Python 3.13.7.

Jupyter Notebook: Jupyter Notebook is the main IDE in which models are developed and thus it provides an interactive environment to write, explore data and visualize.

Deep Learning Frameworks:

- Models are built and trained with TensorFlow (including Keras).
- PyTorch also used for flexibility in the ability to train various models.

Libraries:

- NumPy, Pandas to work with and manipulate data.
- OpenCV to pre-process and augment images.

- Matplotlib, Seaborn for visualization.
- Scikit-learn to get more machine learning use.

Cloud Resources (TPU/GPUs of Kaggle)

- The deep learning models are trained on Kaggle TPUs and GPUs. These high-performance resources help to train the models faster using parallel processing.
- The dataset is stored and maintained on Kaggle, and thus it is easy to access the images and directly process them in the cloud.

Data Storage and Backup

- Storage: The data is stored on Dell Inspiron 15 3501, and it is easily processed.
- Cloud Storage (Kaggle Datasets): Big image datasets are stored in Cloud Storage where they can be easily accessed and easily integrated into the training process.

4.2 Testing and Evaluation/Performance/Comparative Analysis

This section presents a comprehensive evaluation of the proposed deep learning-based system for Retinopathy of Prematurity (ROP) detection, using various models. Models were evaluated based on several performance metrics, including accuracy, precision, recall, F1-score, training and testing loss, training time, inference time, ROC curves, and confusion matrices. The results are systematically organized through summary tables, visual analyses, and comparative discussions to identify the best-performing models.

Table 4.1: Performance of all the trained Models.

Model Name	Accuracy	Precision	Recall	F1 Score
ResNet50	85.84%	0.8814	0.8584	0.8550
ResNet101	89.38%	0.9067	0.8938	0.8971
MobileNetV2	85.84%	0.8394	0.8584	0.8399
EfficientNetV2	91.15%	0.9233	0.9115	0.9147
ViT-B	88.12%	0.8819	0.8812	0.8836
DeiT	89.09%	0.8914	0.8909	0.8907
Ensemble(ResNet101+DeiT)	94.69%	0.9487	0.9469	0.9473

Training vs Validation: In this section, the qualitative performance analysis of the models is evaluated using various metrics for the top three models ResNet101, EfficientNetV2, and the Ensemble Model (ResNet101+DeiT) to provide a comprehensive comparison of their training and validation performances, highlighting their strengths and weaknesses in the context of ROP detection.

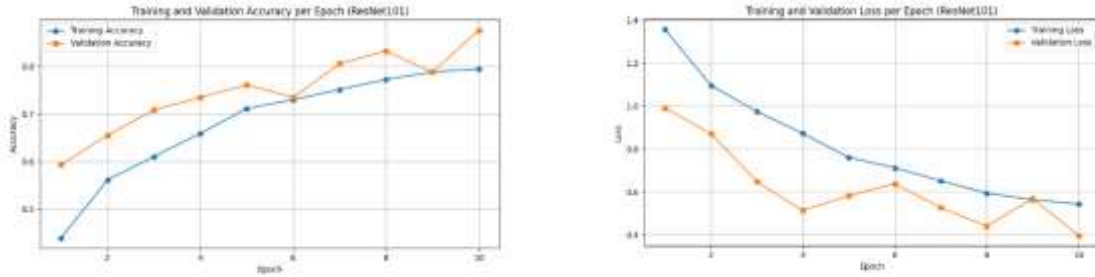


Fig 4.1: ResNet101 Training & Validation accuracy (Left) and loss Curve (Right)

Fig 4.1 shows the validation accuracy and training accuracy (on the left) and the loss curves (on the right) of ResNet101 model versus the epochs. The error in the training goes down consistently, and the error committed by the training drops which is a sign that is learning and getting better. To be fair, however, there is a visible distance between the training curve and the validation curve implying that the model is slightly over-fitted to the training data and unable to make predictions on unknown data. The prediction accuracy is gradually improving with some variation, which points towards the model being difficult to generalize on new data points during the training.

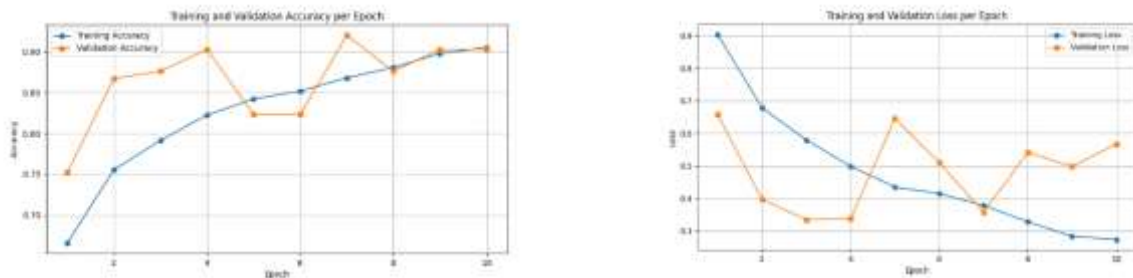


Fig 4.2: EfficientNetV2's Training & Validation accuracy (Left) and loss Curve (Right)

Fig 4.2 shows the training and validation accuracy (on the left) and the loss curves (on the right) of the EfficientNetV2 model versus the epochs. The training error drops continuously and the training loss is reduced and this indicates that the model is training well. The validation accuracy also displays a consistent increasing trend, with little variation, telling us that the model is extrapolating its outcomes to unmeasured data. The validation loss is also quite stable that shows the model does not overfit and performs well on the training and the validation data.

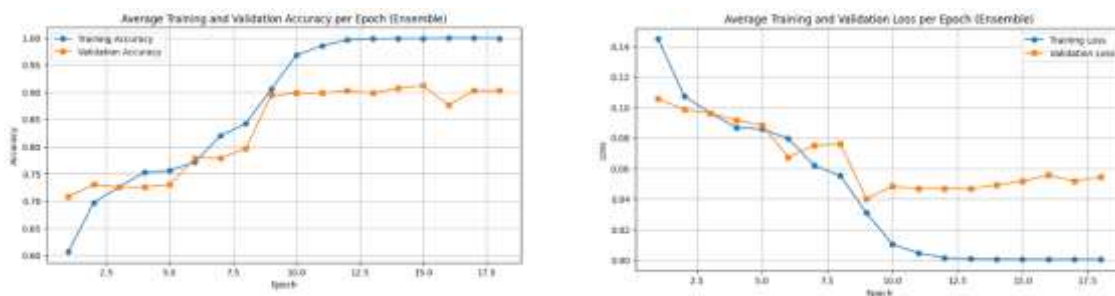
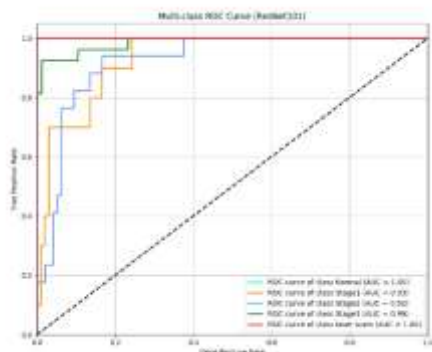
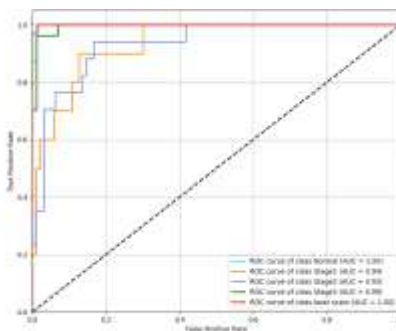


Fig 4.3: Ensemble (ResNet101+DeiT)'s Training & Validation accuracy (Left) and Loss Curve (Right).

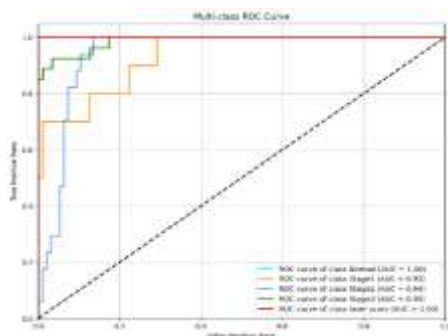
Fig 4.3 displays the accuracy of the training and validation (on the left) and the loss curves (on the right) of the Ensemble model (ResNet101 + DeiT) with respect to the epochs. Training accuracy increases gradually, and training loss is reduced which means successful learning. This extension to the validation accuracy also implied good generalization to previously untested data, and that the same increasing tendency with few oscillations was being extrapolated to the validation accuracy. The comparably high performance both on the training and the validation set is indicative of the efficacy of using the ResNet101 and DeiT in the ensemble to achieve better ROP detection.



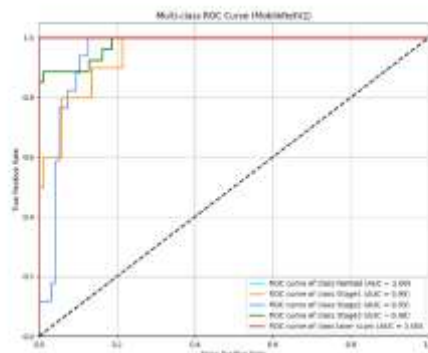
ResNet101



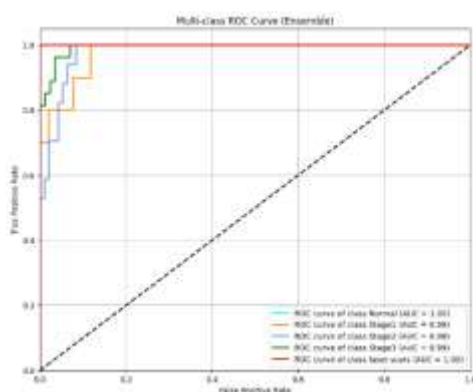
ResNet50



EfficientNetV2



MobileNetV2



Ensemble(ResNet101+DeiT)

Fig 4.4: ROC curve for 5 top models (ResNet101, ResNet50, EfficientNetV2, MobileNetV2 & Ensemble)

ROC Curve: Figure 4.10 shows the combined ROC curves for the top 5 models (ResNet101, ResNet50, EfficientNetV2, MobileNetV2 & Ensemble(ResNet101+DeiT)). Each class shows an AUC which indicating that the models perform perfectly on all five classes. The ROC curves demonstrate that these models have excellent discriminatory power across the classes, with no apparent overlap between true positive and false positive rates.

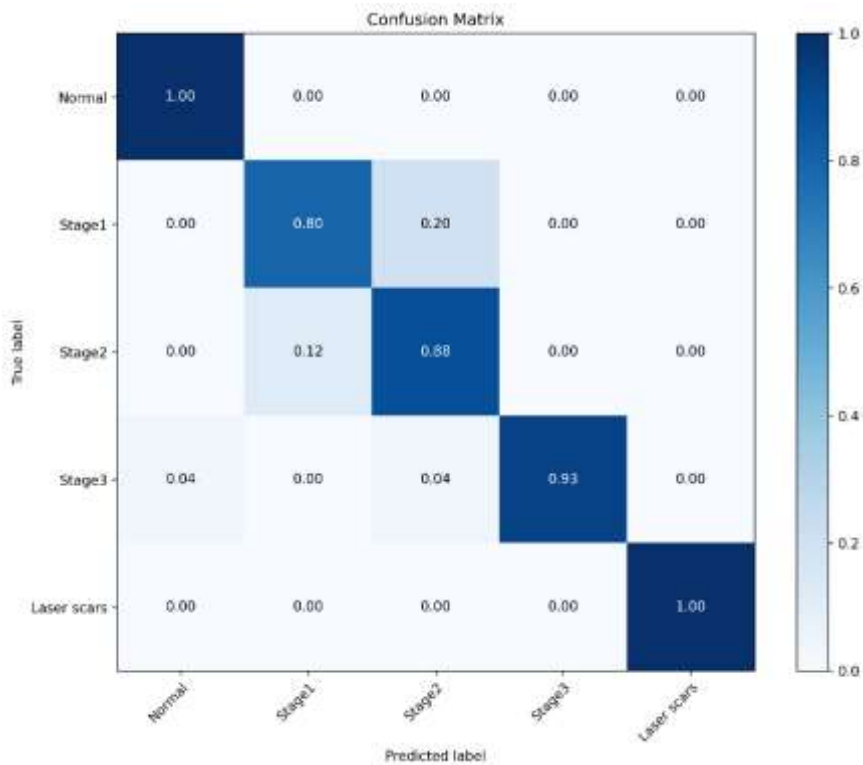


Fig 4.5: Confusion Matrix of Ensemble (ResNet101+DeiT) Model.

Confusion Matrix: Figure 4.11 presents the confusion matrix for Ensemble(ResNet101+DeiT) Model. The matrix illustrates that model performs exceptionally well across all classes, with high accuracy reflected on the diagonal elements. The off-diagonal elements represent occasional misclassifications, which are minimal, indicating that the model effectively differentiates between the classes.

4.3 Results and Discussion

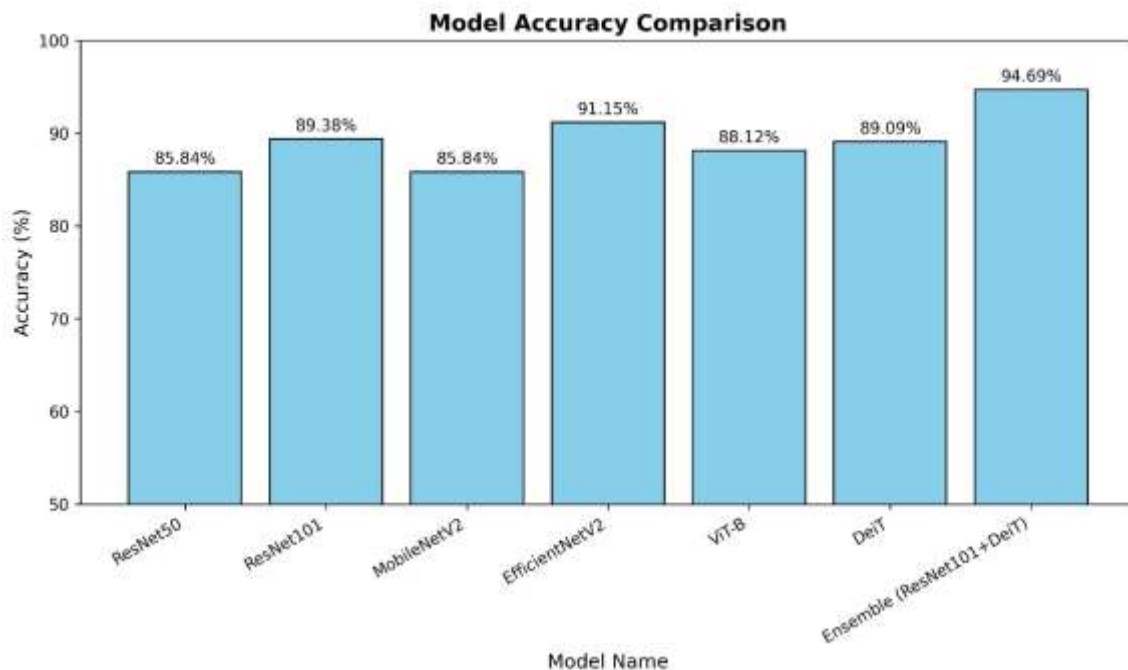


Fig 4.6: Accuracy Comparison of the trained Models.

The deep learning models designed to detect early Retinopathy of Prematurity (ROP) were assessed through a held-out test set and important performance measures, such as accuracy, precision, recall and F1-score. The results showed that the precision of ResNet50 and MobileNetV2 was 85.84, and the recall was rather low, which shows the issues with detecting early-stage ROP. ResNet101 showed improved accuracy of 89.38 and better precision and recall than the other models. Though ResNet101 was very effective at detecting more advanced stages of ROP, it still showed a limited performance on the early-stage detection. EfficientNetV2 with 91.15% accuracy is a good compromise between precision and recall and can thus be used in both cases of early and advanced ROP. Transformer-based models ViT-B and DeiT reached an accuracy of 88.12 and 89.09, respectively. These models did represent more general contextual patterns, and this was used to establish ROP signs in various phases.

The performance of the Ensemble Model (ResNet101 + DeiT) was the best at 94.69%, which demonstrates the advantages of the union of the local feature extraction capability of CNN and global context recognition by transformers. The ensemble strategy made the model better at recognizing minor early-stage anomalies as well as more severe signs of ROP, leading to higher precision (0.9487), recall (0.9469), and F1-score (0.9473). So it was the best model to differentiate between ROP images.

Table 4.2: Comparison with previous work.

Study By	Methodology	Evaluation Metrics
Kiran Yenice et al.[9]	RegNetY002	93%
Omneya Attallah et al.[4]	DIAROP with DWT-integrated features	93.2%
Yo-Ping Huang et al.[3]	Custom deep CNN	92.23%
Yan Tong et al.[7]	R-CNN	90%
Proposed	Ensemble (ResNet101+DeiT)	94.69%

Although the ensemble model demonstrated remarkable results, there are issues associated with the maintenance of robustness in a variety of clinical settings, including variability in image quality and the acquisition device. The model will also require a detailed information about different clinical settings to conclude that the model can be generalized and implemented to practice. However, the potential of the proposed system is massive to be able to identify ROP automatically and provide the health care providers, specifically, with the safe source to avoid being blinded by ROP due to the instant diagnoses and instant treatment in the resource-constraining environment where the latter are the most influential factors.

4.4 Summary

In this chapter, the findings and the analysis of the performance of the deep learning models that were established to identify fundus images that contain features of Retinopathy of Prematurity (ROP) were presented. Various models, such as ResNet50, ResNet101, MobileNetV2, EfficientNetV2, ViT-B, DeiT and an ensemble model of ResNet101/DeiT were compared regarding accuracy, precision, recall, F1-score, and AUC. The ensemble model was the most accurate at 94.69% and beat all other models, especially in the detection of ROP at early stages. The findings showed that using the strengths of both CNN-based models (ResNet101) and transformer models (DeiT) increased the capacity to identify delicate abnormalities without significantly affecting the model performance in identifying an advanced stage of ROP.

Despite the relative success of the models, some of the challenges remain particularly with respect to the difference in quality of images and conditions under which they are acquired. This will require additional validation to verify whether the system

can be extrapolated to other datasets and clinical settings. Nonetheless, the excellent performance of the ensemble model also evidences that it may be used in clinical practice as well, especially in resource-constrained settings, where automated detection of early ROP could improve the efficiency of diagnosis and prevent blindness due to ROP.

The strengths and weaknesses of the individual models were also covered in relation to this chapter, as well as how techniques like focal loss and two-stage fine-tuning contribute to class balance and improving models. The results suggest that high-quality deep learning-based algorithms can be used to estimate early and accurate ROP, which could be a welcome addition to the list of tools that might assist medical workers in giving their patients better results.

Chapter 5

Engineering Standards and Design Challenges

This chapter introduces the engineering standard and design challenges that provide a brief overview of the study.

5.1 Compliance with the Standards

5.1.1 Software Standards

The ROP Diagnosis System Software Standards stipulates that the development is anchored on best practices in code, documentation and performance. It adheres to the PEP 8 conventions of Python and produces clean and consistent and modular code. The present research findings are stored in the Jupyter notebook, Colab and Kaggle and markdown cells to explicitly separate the method, data pre-processing, model parameter and results. They also maintain versioned and formatted local notebooks to follow the changes but no system of external version control, like GitHub or GitLab. The system prioritizes on data security since it saves sensitive medical images in a safe place like Google Cloud Storage or Kaggle Datasets.

Unit testing is done with the Python unittest system in order to ensure that the system is able to do what it is supposed to do, and integration testing to ensure that the parts can be used as intended with each other. It provides deep learning models developed on TensorFlow, Keras, and PyTorch and manipulates data and visualizes it with Pandas, NumPy, and Matplotlib. It can optimize the performance to successfully process large data sets of images and generate a fast response. These software standards assist the system in creating high-quality and reproducible research in addition to offering clarity, security, and performance during the development of the system.

5.1.2 Hardware Standards

Hardware Standards of the ROP Diagnosis System guarantees that the system works in diverse hardware environments without affecting its performance. The system is designed to be personal computer-compatible, cloud computing platforms like Google Colab and run on Kaggle notebooks and use hardware resources efficiently. High performance computing systems such as GPUs (Graphics Processing Units) are used to accelerate the execution of the deep learning model to train and infer the model. The environments should be able to help reduce the time of medical image datasets that are large. The system itself does not require any special hardware to execute, and can operate on any laptop/desktop with 8GB or above of RAM (or more), with a more modern CPU

(Intel i5) to execute the code in Jupyter Notebooks or Colab. As one is dealing with bigger amounts of data or needs to train hard models, the high-performance cloud instances with GPUs are introduced, providing speedy and productive model training. The hardware environment is intended to support the easy operation, the effective processing of the medical image information, and the rapid model inference avoiding significant delays to achieve real-time predictions and analysis.

5.1.3 Communication Standards

ROP Diagnosis System mainly relies on HTTPS as a secure communication medium between the user interface and the backend. In transferring image data to the backend, the system uses RESTful API calls to transfer the data to the backend. This is offered in the JSON format so that it can be easily integrated and displayed. Because the system is based on the Google Colab and Kaggle notebooks, the communication is based on the embedded environments of their cloud-based computation, which guarantees safe and effective information exchange in these systems.

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

The ROP Diagnosis System makes a commendable contribution towards the enhancement of the lives of the preterm babies through the capability to diagnose Retinopathy of Prematurity (ROP) which is the number one cause of blindness in a child. Early diagnosis means that timely treatment can be given and the risk of permanent impaired vision are eliminated, in addition, the overall quality of life of the affected babies can be enhanced. By this system, ROP screening becomes more easily available, especially where there is limited access to specialized healthcare providers in the community and therefore healthcare providers can more easily monitor and treat high-risk infants. The system also alleviates the burden on medical personnel by automating the diagnosis process, so in that case more infants can get the attention and care they need, resulting in improved health outcomes and fewer instances of preventable blindness.

5.2.2 Impact on Society & Environment

The ROP Diagnosis System positively affects society as it enhances access to healthcare, particularly in underserved communities with low access to ROP screenings. The system, through automation of diagnosis, simplifies and lowers the cost of diagnosing ROP early in children to minimize cases of preventable childhood blindness. This results in economic savings in the long term and positively impacts the lives of the affected children.

The system also leads to global health equity, making necessary healthcare equipment available to areas with restricted access to specialized healthcare. The system makes the healthcare processes more environmentally-friendly by utilizing cloud-based solutions, such as Google Colab and Kaggle, thus consuming less energy as it does not require any local computing devices and reduces the necessity of computing resources.

5.2.3 Ethical Aspects

The ROP Diagnosis System follows ethical principles by guaranteeing the privacy and security of confidential medical records. The system adheres to the set data protection policies like the HIPAA (Health Insurance Portability and Accountability Act) to ensure that the patient information is not compromised and is not accessed by unauthorized individuals. Any medical data and diagnostic outcomes are encrypted when transmitting and stored safely in reliable cloud facilities, such as Google Cloud Storage or Kaggle Datasets. Also, the system is more concerned with being fair and accurate in its predictions in the sense that the model is trained on a variety of data to prevent bias. AI in healthcare brings up issues of accountability and this system does not take the place of medical professionals but helps them in their work, so accountability for making final decisions should still be in the hands of healthcare professionals. System design will also be the final step whereby its functionality will be revealed in an irreproachable manner and will take the target populations (the needy populations) to a point of reaping the rewards of early detection without necessarily having to pay the cost of financial and technological insurance.

5.2.4 Sustainability Plan

The ROP Diagnosis System Sustainability Plan assures the effectiveness, maintainability, and time scalability of systems. The system has been built with continuous improvement and flexibility in view of maintaining sustainability in the long term. The model will also be updated periodically as more information is acquired in order to be more accurate and work more effectively as more cases of ROP are detected. This allows the system to facilitate modifying medical practice and data sets. The advantage of the cloud-based infrastructure is also the system i.e. does not need very-maintained hardware and can be scaled. Through tools such as Google Colab and Kaggle, the system is utilizing energy efficient data centers, which minimize its impact on the environment.

The system is also designed to be scalable so that it can be implemented in various healthcare settings around the world and in low-resource settings in particular. By working with local healthcare providers and governments, the system will be made accessible to people in need to create long-term effects in the reduction of blindness caused by ROP. In addition to this the open-source software will mean that the system will be maintainable and can be further used and extended by the rest of the research and medical community, hence, become useful to the rest of the community.

5.3 Project Management and Financial Analysis

The ROP Diagnosis System Project Management is based on the agile school of thought, where milestones and reviews are emphasized to assure a timely delivery. The project phases include research, model development, user interface design, testing and deployment, and timelines will be established around them. It is being created on systems such as Google Colab and Kaggle, where the resources and infrastructure can be expanded.

Cost Analysis (Budget)

Taking into account the budget of 40,000BDT, the costs break down as follows:

Research and Development:

- Data Collection and Labelling: 5,000 BDT.
- Model Training and Development: 10000 BDT.
- Testing and Validation: 3,000 BDT
- UI/UX Design: 2,500 BDT
- Total : 20,500 BDT

Cloud Infrastructure:

- Google Colab (TPU/GPU usage): 4,000 BDT/per year.
- Kaggle Store, and Data Access: 2000 BDT/per year.
- Overall Cloud Service price: 6,000 BDT/year.

Maintenance & Updates:

- System maintenance: 3000 BDT/yearly.
- Software: 2000 BDT.
- Total Maintenance cost: 3,500 BDT/year.

Project Management and Miscellaneous:

- Team Salaries: 7,000 BDT
- Contingency and Miscellany: 3000 BDT.
- All in all, Project Management 10,000 BDT.

Total Budget: 40,000BDT

Rationales for Budget:

Data Collection: The budget aims at obtaining the available medical data and labelling it effectively to cut down the expenses.

Model Training and Development: Free use of Google Colab and its GPUs/TPUs reduces the use of costly cloud services.

Cloud Services and Maintenance: The type of storage used by Kaggle is cost effective and there is low maintenance on the cloud infrastructure.

Team Salaries and Miscellaneous: The budget also does not prioritize the team itself, instead placing more emphasis on the core knowledge and utilizing cloud computing technology to save overhead costs.

Revenue Model:

The assistance of purchasing subscriptions by healthcare facilities will allow financing the ROP Diagnosis System, as well as collaborating with governmental health programs and the potential support of grants to improve child healthcare. These sources of funds will ensure that this project is sustainable in the long run and in the process remain affordable to low resource areas.

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

Table 5.1: Mapping with Complex Engineering Problem.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stakeholder Involvement	EP7 Interdependence
✓	✓	✓	✓	x	x	✓

Rationale for Mapping:

EP1(Dept of Knowledge): The knowledge needed to solve this problem consists mainly of specialist knowledge of medical image processing and deep learning. Cloud infrastructure also requires expertise to manage massive datasets of images.

EP2(Scope of Competing Requirements): Competing requirements that exist include: to ensure model accuracy and at the same time to be able to run the system on cloud platforms with limited resources. The answer has to be a trade-off between these for performance and scalability.

EP3(Depth of Analysis): This system needs an in-depth analysis of the image data and a deep learning algorithm to process the image and make a diagnosis of ROP. Moreover, it is important to know the medical situation concerning ROP detection.

EP4(State of Knowledge): The ROP detection problem is not new to medical research but the implementation of a working system to detect ROP in real-time under low resourced conditions is novel.

EP5(Codes Applicable): Applicable codes include: Medical imaging codes, data privacy laws (e.g. HIPAA), best practices for cloud computing. The standards would be followed during the system development and deployment.

EP6(Current Level of Stakeholder Engagement): Stakeholders: Cloud Service Providers, Healthcare providers, AI engineers These groups must work together in order to make the system successful.

EP7(Interdependence): The solution relies heavily on a number of interrelated entities the medical domain expertise, the machine learning models, and the cloud infrastructure hosting and processing the data.

Mapping with Knowledge Profile

Table 5.2: Mapping with knowledge Profile.

K1 Natu ral Scien ce	K2 Mathem atics	K3 Engineeri ng Fundame ntals	K4 Special ist Knowle dge	K5 Enginee ring Design	K6 Enginee ring Practice	K7 Comprehe nsion	K8 Resear ch Literat ure
✓	✓	✓	✓	✓	✓	x	✓

Rationale for Mapping:

K1 (Natural Science): The study applies the concepts of natural science with regard to medical imaging, especially in the interpretation of retinal fundus images and biology.

K2 (Mathematics): Mathematics is the basis of deep learning; linear algebra, probability, and optimization are all elementary concepts of deep learning, but mathematics is not explicitly highlighted in this work.

K3 (Engineering Fundamentals): Foundations of engineering, such as computational modeling, system design, and algorithm development, are used in the construction of the deep learning structure.

K4 (Specialist Knowledge): The thesis exemplifies specialist knowledge in the areas of artificial intelligence, computer vision, and medical image analysis, and specifically Retinopathy of Prematurity (ROP) classification.

K5 (Engineering Design): Design concepts in the engineering design are implemented in the choice of architectures (ResNet, EfficientNet, ViT, DeiT), utilization of focal loss, and a two-stage fine-tuning process.

K6 (Engineering Practice): This area deals with practical aspects by preparing the dataset, training the model on Google Colab, and evaluating the performance, as well as the applicability of the system in practice in medical screening.

K7 (Comprehension): This study shows understanding through an amalgamation of theoretical concepts and practical application and critical evaluation of the received findings.

K8 (Research Literature): The review of the existing works, methods, and medical imaging datasets will make sure that the research is based on the existing knowledge and reveals the gaps that should be closed.

5.4.2 Engineering Activities

Mapping with Complex Engineering Activities

This section is designed to map the overall problem and EA's (*multiple*).

Table 5.3: Mapping with Complex Engineering Activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓	✓	✓	✓	✓

Rationales for Mapping:

EA1 (Range of Resources): The system requires access to a variety of resources, including a large and diverse dataset of medical images, cloud computing resources for storage and model processing, and machine learning infrastructure for training the model. These resources are critical to developing an accurate and scalable ROP detection system.

EA2 (Level of Interaction): There is significant interaction between different system components, such as the frontend (user interface), backend (image processing and model inference), and external resources like cloud computing services. The system needs to integrate machine learning models with real-time user input to process images and provide diagnoses.

EA3 (Innovation): The project introduces innovative solutions, particularly by leveraging AI for the early detection of ROP using deep learning models. The innovation lies in applying deep learning to medical imaging and deploying it at scale for real-time screenings, especially in resource-constrained environments.

EA4 (Consequences for Society and Environment): The impact of this system is significant in terms of societal benefit. It can reduce preventable childhood blindness, offering a scalable solution for healthcare delivery in low-resource settings. The system's positive societal consequences include improved health outcomes and a reduction in healthcare costs.

EA5 (Familiarity): While familiarity with general machine learning and medical image analysis is high, the specific application of these technologies to detect ROP is more specialized. The problem-solving approach may require some new methodologies, particularly in scaling the system for real-world medical use.

5.5 Summary

The ROP Diagnosis System project presupposes solving a complicated engineering problem, which requires knowledge in the field of medical image processing, deep learning, and cloud computing. The system will provide an automated system to detect Retinopathy of Prematurity (ROP) which is one of the leading causes of childhood blindness using AI-powered models and cloud computing. The central issues are the trade-off between the model accuracy and computation and information privacy and use privacy of deep learning in medical diagnosis, in low-resource conditions, in particular. Many stakeholders such as healthcare workers, AI developers, and cloud computing providers are needed to work together in the project in order to make the system actively utilized in the real world. It is a complex problem since the engineering activities and knowledge profiles required to resolve the problem are projected to the categories in question. The system is capable of providing a highly social result in the form of better early detection and lowering the rates of avoidable blindness, particularly in the underserved regions.

The system will take care of the software, hardware and communication requirements to ensure that it is scalable, secure and sustainable. To satisfy the deadline on the project, the management has a clear cut budget which it can stick to that periodically gets feedbacks by the stakeholders which upgrades the system and magnifies the system as the project goes on. On the basis of the budgetary analysis, the project may be prepared on a middle budget, and there is a potential paradigm regarding the development of revenues in accordance with the subscription, payable and collaboration. Overall, the provided project is expected to provide affordable yet quality innovative intervention that would result in a better access to and outcome of vulnerable babies at the international level. .

Chapter 6

Conclusion

This chapter introduces the summary limitation and future work that provide a brief overview of the study and future work.

6.1 Summary

This research applied to create an automated deep learning based system to detect Retinopathy of Prematurity (ROP) in fundus images as an early disease. The aim of the research was to answer the urgent question of early ROP diagnosis, which is critical in preventing irreversible vision loss in a premature infant. The traditional diagnostic methods, which rely on manual screening performed by ophthalmologists, are time-consuming and vulnerable to human error especially when it comes to detecting cases of ROP at an early stage. Thus, to automatize the process of detection, the proposed system considers the latest deep learning frameworks, such as CNNs (ResNet50, ResNet101, MobileNetV2, EfficientNetV2) and transformers (ViT-B, DeiT).

There was some logical scheme that presupposes data collection, data cleaning, and data enriching. To optimize the models, two-stage fine-tuning, and focal loss were used to train and fine-tune the models. A hybrid model consisting of ResNet101 and DeiT were created, and it showed better results than single models, reaching the highest accuracy of 94.69%.

The models were evaluated on a held-out test set and it was observed that the ensemble model performed better than all other models in terms of accuracy, precision, recall, F1-score, and AUC indicating the prospect of deep learning in automated ROP detection. Despite being a highly promising system, problems such as fluctuating image quality and inconsistency of the device still exist. The next business action plan is to test the model with an expanded range of data to strengthen it and make it externalizable.

In conclusion, the proposed system is a good and reliable way to identify early ROP that may, in its turn, lead to some useful help to the professionals in the setting with limited resources. The system has the potential to improve patient outcomes and lessen the global disease burden associated with ROP-related blindness because it can shorten the time required to diagnose a patient, and more accurately detect them.

6.2 Limitation

Although the deep learning-based Retinopathy of Prematurity (ROP) detection system, developed in this paper, achieves encouraging results, it has a number of limitations that are to be taken into consideration. One, the training and evaluation dataset is not that large and is imbalanced, and it can be assumed that the model would not be able to generalize to various clinical data in real-life settings. Despite the data augmentation methods that enabled the expansion of the dataset, the small sample of the labeled ROP images, particularly those related to early stages, remains a problem. In addition, despite the imbalance between the classes being corrected through the use of techniques like focal loss, the unequal representation of the ROP stages within the dataset may affect the sensitivity of the model to rare cases, particularly the ability to detect the late ROP stages. The other weakness is that the quality of fundus images is inconsistent across devices and imaging conditions, and this inconsistency may affect the performance of the model. Though the models have been trained to deal with some amount of this variability, additional preprocessing and fine tuning might be required to deal with generalization among the various acquisition devices. It is also still unclear whether the model can be generalized to a variety of clinical settings, and external validation of the model using data in different clinical settings is needed. Secondly, this system has not been used in a real clinical environment, and its operation in real-life settings, including device settings and patient demographic variability, is yet to be evaluated. Finally, the deep learning models, especially the ensemble approach, are not interpretable, which can be a barrier to trust in the system by healthcare professionals. It would be important to include explainable AI methods to increase the transparency and trust in the decision making process of the model. However, the provided system will also be a potentially beneficial tool which will be utilized during the process of ROP detection, and the research in the sphere is needed to improve the system and which will become a more powerful and useful tool which can be applied in practice.

6.3 Future Work

Although the deep learning-based system of Retinopathy of Prematurity (ROP) detection developed in this study has demonstrated encouraging results, there are various aspects, which may be addressed in future studies to improve the performance, robustness, and applicability to the real world. Increasing the data is one of the main focus areas. Although data augmentation methods can be used to address the issue of class imbalance, a bigger and more diverse dataset, and in particular, more examples of ROP at the earliest stages, would improve generalization and model precision as well. Furthermore, by sampling data across various clinical environments using varying imaging tools, we will be able to train the model to cope with the variability observed in real-world data.

Further studies may be conducted to enhance the model capabilities to generalize, i.e., to deal better with low-quality images, through advanced preprocessing methods, e.g. removing noise, improving image quality, etc. Specific device fine-tuning models may also be used to maximize the performance of the devices in different acquisition conditions. Less obvious methodological solutions to class imbalance include artificial data generation or oversampling of small classes; as well as may be viewed as more sensitive in models, particularly during finding the rare and severe stages of ROP.

The second domain, which needs to be addressed in a future working plan, is the application of the system to real world clinical settings. The model has demonstrated good performance on the test case but its performance in a real clinical scenario is yet to be tested. Further research is yet to be done on this model to apply it to clinical trials to observe the effectiveness of this model on real patient data, whether this type of research can be successfully applied to the hospitals, and whether implementing this model is not easy to do as part of the health care working process. Live testing would also provide feedback regarding the latency and efficiency of the system in question where determining whether or not the tool could be applied to clinical practice or not.

Moreover, a model interpretability is an essential feature that should be adopted in clinical practice. The next generation work can focus on the use of explainable AI techniques to provide insight into how the model works. This would assist the healthcare workers know how the model comes to its conclusions and build confidence in the system. It can be considered whether methods such as Grad-CAM or SHAP values can be used to display the areas of the picture that the model itself deems important when making predictions, especially in early-stage ROP detection.

Finally, investigating how to transfer learning to other medical imaging data sets or to use the system in other diseases that have similar retinal attributes, such as diabetic retinopathy or glaucoma, may broaden the range of applications of the model. These would allow the model to be expanded to general applicability in ophthalmology, and potentially other effect on global healthcare.

In summary, the future of this research is to improve the system in terms of its accuracy, scalability and interpretability and test its relevance in various clinical environments. Overcoming the limitations recognized in this study and adding feedback obtained in the real-world implementations will be essential in transforming the system into a useful tool in detection and diagnosis of ROP.

References

- [1] X. Zhao *et al.*, “A fundus image dataset for intelligent retinopathy of prematurity system,” *Scientific Data*, vol. 11, no. 1, May 2024, doi: 10.1038/s41597-024-03362-5.
- [2] National Eye Institute, “Retinopathy of Prematurity | National Eye Institute,” Mar. 07, 2024. [https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/retinopathy-prematurity#:~:text=Retinopathy%20of%20prematurity%20\(ROP\)%20is,the%20back%20of%20your%20eye\).](https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/retinopathy-prematurity#:~:text=Retinopathy%20of%20prematurity%20(ROP)%20is,the%20back%20of%20your%20eye).)
- [3] Y.-P. Huang *et al.*, “Automated detection of early-stage ROP using a deep convolutional neural network,” *British Journal of Ophthalmology*, vol. 105, no. 8, pp. 1099–1103, Aug. 2020, doi: 10.1136/bjophthalmol-2020-316526.
- [4] O. Attallah, “DIAROP: Automated Deep Learning-Based Diagnostic Tool for Retinopathy of Prematurity,” *Diagnostics*, vol. 11, no. 11, p. 2034, Nov. 2021, doi: 10.3390/diagnostics11112034.
- [5] J. Zhang, Y. Liu, T. Mitsuhashi, and T. Matsuo, “Accuracy of Deep Learning Algorithms for the diagnosis of Retinopathy of Prematurity by Fundus Images: A Systematic Review and Meta-Analysis,” *Journal of Ophthalmology*, vol. 2021, pp. 1–11, Aug. 2021, doi: 10.1155/2021/8883946.
- [6] A. Bai *et al.*, “Multicenter validation of deep learning algorithm ROP.AI for the automated diagnosis of PLUS disease in ROP,” *Translational Vision Science & Technology*, vol. 12, no. 8, p. 13, Aug. 2023, doi: 10.1167/tvst.12.8.13.
- [7] Y. Tong, W. Lu, Q.-Q. Deng, C. Chen, and Y. Shen, “Automated identification of retinopathy of prematurity by image-based deep learning,” *Eye And Vision*, vol. 7, no. 1, Aug. 2020, doi: 10.1186/s40662-020-00206-2.
- [8] J. Wang *et al.*, “Automated explainable multidimensional deep learning platform of retinal images for retinopathy of prematurity screening,” *JAMA Network Open*, vol. 4, no. 5, p. e218758, May 2021, doi: 10.1001/jamanetworkopen.2021.8758.
- [9] E. K. Yenice, C. Kara, and Ç. B. Erdaş, “Automated detection of type 1 ROP, type 2 ROP and A-ROP based on deep learning,” *Eye*, vol. 38, no. 13, pp. 2644–2648, Jun. 2024, doi: 10.1038/s41433-024-03184-0.
- [10] Z. Tan, S. Simkin, C. Lai, and S. Dai, “Deep learning algorithm for Automated diagnosis of retinopathy of Prematurity Plus Disease,” *Translational Vision Science & Technology*, vol. 8, no. 6, p. 23, Dec. 2019, doi: 10.1167/tvst.8.6.23.
- [11] D. Theckedath and R. R. Sedamkar, “Detecting affect states using VGG16, RESNet50 and SE-RESNet50 networks,” *SN Computer Science*, vol. 1, no. 2, Mar. 2020, doi: 10.1007/s42979-020-0114-9.
- [12] Q. Zhang, “A novel ResNet101 model based on dense dilated convolution for image classification,” *SN Applied Sciences*, vol. 4, no. 1, Dec. 2021, doi: 10.1007/s42452-021-04897-7.
- [13] H. Qassim, A. Verma, and D. Feinzimer, “Compressed residual-VGG16 CNN model for big data places image recognition,” *2022 IEEE 12th Annual Computing and*

- Communication Workshop and Conference (CCWC)*, Jan. 2018, doi: 10.1109/ccwc.2018.8301729.
- [14] A. Howard, A. Zhmoginov, L.-C. Chen, M. Sandler, and M. Zhu, “Inverted residuals and linear bottlenecks: mobile networks for classification, detection and segmentation,” *arXiv (Cornell University)*, Jan. 2018, [Online]. Available: <https://arxiv.org/pdf/1801.04381>
- [15] L. Yao, C. Mao, and Y. Luo, “Graph convolutional networks for text classification,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 7370–7377, Jul. 2019, doi: 10.1609/aaai.v33i01.33017370.
- [16] “Retinopathy of Prematurity (ROP) in Premature Infants,” *Fort Worth Eye Associates*. <https://www.ranelle.com/pediatric-ophthalmology/322-2/>
- [17] Salih, Nazar, et al. “Prediction of ROP Zones Using Deep Learning Algorithms and Voting Classifier Technique.” *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, 17 May 2023, <https://doi.org/10.1007/s44196-023-00268-9>. Accessed 6 Sept. 2025.
- [18] Redd, Travis K, et al. “Evaluation of a Deep Learning Image Assessment System for Detecting Severe Retinopathy of Prematurity.” *British Journal of Ophthalmology*, vol. 103, no. 5, 23 Nov. 2018, pp. 580–584, <https://doi.org/10.1136/bjophthalmol-2018-313156>. Accessed 6 Sept. 2025.
- [19] Bernardes, Rui, et al. “Digital Ocular Fundus Imaging: A Review.” *Ophthalmologica*, vol. 226, no. 4, 2011, pp. 161–181, <https://doi.org/10.1159/000329597>.
- [20] Chen, Jing, and Lois E. H. Smith. “Retinopathy of Prematurity.” *Angiogenesis*, vol. 10, no. 2, 27 Feb. 2007, pp. 133–140, <https://doi.org/10.1007/s10456-007-9066-0>.
- [21] Dogra, Mangat Ram, et al. “An Update on Retinopathy of Prematurity (ROP).” *The Indian Journal of Pediatrics*, vol. 84, no. 12, 4 July 2017, pp. 930–936, <https://doi.org/10.1007/s12098-017-2404-3>.
- [22] Dong, Xibin, et al. “A Survey on Ensemble Learning.” *Frontiers of Computer Science*, vol. 14, no. 2, 30 Aug. 2019, pp. 241–258, <https://doi.org/10.1007/s11704-019-8208-z>.
- [23] Houston, Samuel K., et al. “Laser Treatment for Retinopathy of Prematurity.” *Lasers in Medical Science*, vol. 28, no. 2, 2 Dec. 2011, pp. 683–692, <https://doi.org/10.1007/s10103-011-1021-z>.
- [24] Sagi, Omer, and Lior Rokach. “Ensemble Learning: A Survey.” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, 27 Feb. 2018, <https://doi.org/10.1002/widm.1249>.
- [25] Mueller, B, et al. “Treatment of Type I ROP with Intravitreal Bevacizumab or Laser Photocoagulation according to Retinal Zone.” *British Journal of Ophthalmology*, 14 June 2016, p. [bjophthalmol-308375](https://doi.org/10.1136/bjophthalmol-2016-308375), <https://doi.org/10.1136/bjophthalmol-2016-308375>.
- [26] Rajan, Renu P, et al. “Treatment of Retinopathy of Prematurity (ROP) Outside International Classification of ROP (ICROP) Guidelines.” *Graefes S Archive for Clinical and Experimental Ophthalmology*, vol. 258, no. 6, 22 Apr. 2020, pp. 1205–1210, <https://doi.org/10.1007/s00417-020-04706-8>. Accessed 6 Sept. 2025.

Shayla
14.9.25

213-15-4305

ORIGINALITY REPORT

22% SIMILARITY INDEX	17% INTERNET SOURCES	14% PUBLICATIONS	12% STUDENT PAPERS
--------------------------------	--------------------------------	----------------------------	------------------------------

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	4%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
3	arxiv.org Internet Source	1%
4	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dharendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	1%
5	Submitted to Higher Education Commission Pakistan Student Paper	1%
6	www.mdpi.com Internet Source	1%
7	Submitted to United International University Student Paper	<1%
8	link.springer.com Internet Source	<1%
9	Bui Thanh Hung, M. Sekar, Ayhan Esi, R. Senthil Kumar. "Applications of Mathematics in Science and Technology - International Conference on Mathematical Applications in Science and Technology", CRC Press, 2025 Publication	<1%



21% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups

-  **31 AI-generated only 21%**
Likely AI-generated text from a large-language model.
-  **0 AI-generated text that was AI-paraphrased 0%**
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

