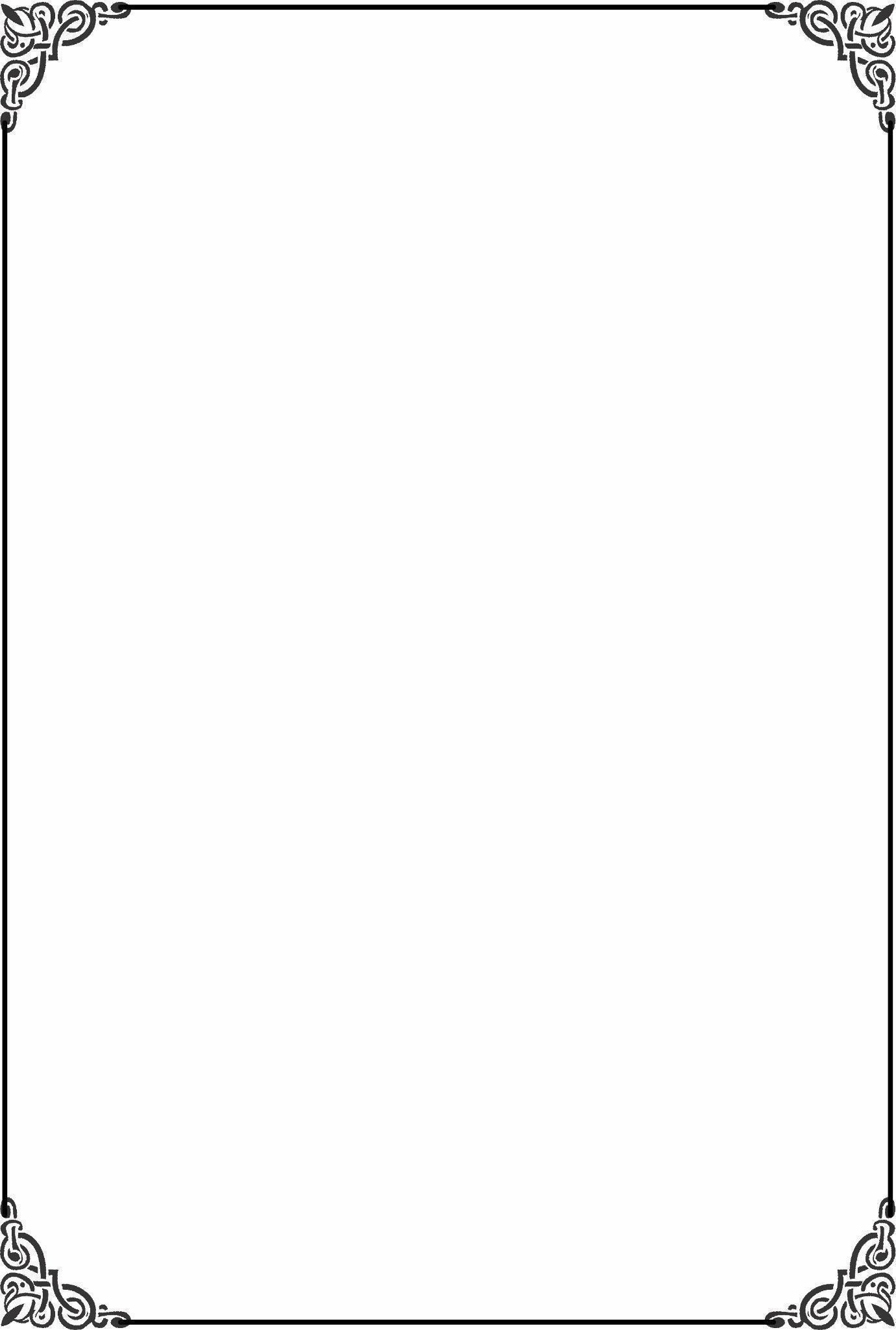
z



# A Dual-Branch CNN Network for Multi-Class Classification Model for Eye Disease Detection

## A PROJECT REPORT

***Submitted in partial fulfilment of the requirements for the award of the degree of***

## BACHELOR OF TECHNOLOGY IN

**ELECTRONICS & COMMUNICATION ENGINEERING**

***Submitted by***

## CH.NAGA GOPAL 21NG1A04F3 CH.SAI MANIKANTA 22NG5A0420 S.SRI CHANDRA 21NG1A04E2 Y.SAI RAVI CHANDRA 21NG1A04F5

**Under the Supervision of**

**Dr. CH. SANTHI RANI,** M.TECH, Ph.D, MISTE, MIE

**Professor & Dean**

# Usha Rama College of Engineering and Technology

**An Autonomous Institution, Approved by AICTE, NEW DELHI, Affiliated to JNTUK Kakinada Accredited by NAAC with A Grade and ISO 9001:2015 Certified**

**NH-16, Telaprolu, Ungutur Mandalam, Near Gannavaram, Krishna District, Andhra Pradesh, India, 521109**

**April, 2025**

# A Dual-Branch CNN Network for Multi-Class Classification Model for Eye Disease Detection

## A PROJECT REPORT

***Submitted in partial fulfilment of the requirements for the award of the degree of***

## BACHELOR OF TECHNOLOGY IN

**ELECTRONICS & COMMUNICATION ENGINEERING**

***Submitted by***

## CH.NAGA GOPAL 21NG1A04F3 CH.SAI MANIKANTA 22NG5A0420 S.SRI CHANDRA 21NG1A04E2 Y.SAI RAVI CHANDRA 21NG1A04F5

**Under the Supervision of**

**Dr. CH. SANTHI RANI,** M.TECH, Ph.D, MISTE, MIE

**Professor & Dean**

****

# Usha Rama College of Engineering and Technology

**An Autonomous Institution, Approved by AICTE, NEW DELHI, Affiliated to JNTUK Kakinada Accredited by NAAC with A Grade and ISO 9001:2015 Certified**

**NH-16, Telaprolu, Ungutur Mandalam, Near Gannavaram, Krishna District, Andhra Pradesh, India, 521109**

**April, 2025**



**BONAFIDE CERTIFICATE**

This is to certify that the Project report entitled “**A Dual-Branch CNN Network for Multi-Class Classification Model for Eye Disease Detection”,** is the Bonafide record of the original work carried out by **CH.NAGAGOPAL(21NG1A04F3),CH.SAI MANIKANTA(22NG5A0420), S.SRI CHANDRA(21NG1A04E2),Y.SAI RAVI CHANDRA (21NG1A04F5),** under our supervision during the academic year 2024-2025.

**Signature of the Supervisor Signature of the Head of the Department (Dr.Ch.Santhi Rani,**M.Tech,Ph.D,MISTE,ME**) (DrB.Nancharaiah)**

**Professor &Dean Professor and Head, ECE**

**External Viva Voce Conducted on**

**DECLARATION**

We, the undersigned here by declare that the project report titled “**A Dual-Branch CNN Network for Multi-Class Classification Model for Eye Disease Detection**” submitted in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Electronics and Communication Engineering, USHA RAMA COLLEGE OF ENGINEERING AND TECHNOLOGY, is an original of work carried out by us under the guidance and supervision of **Mrs.Dr.Ch.Santhi Rani**.

We further declare that this project has not been submitted to any other Institution/University for the award of any degree, diploma.

We affirm that the data, results, and information presented in this report are authentic to the best of our knowledge and belief. Any materials derived from other sources have been duly acknowledged.

Student Names and Signatures:

1. **CH.NAGA GOPAL Roll No: 21NG1A04F3)** - Signature:
2. **CH.SAI MANIKANTA(RollNo:22NG5A0420)**-Signature:
3. **S.SRI CHANDRA (Roll No: 21NG1A04E2)** - Signature:
4. **Y.SAI RAVI CHANDRA (Roll No: 21NG1A04F5)** - Signature:

**Date:**

**Place: TELAPROLU**

**ACKNOWLEDGEMENT**

We express our heartfelt gratitude and sincere thanks to all those who have guided, supported, and encouraged us during the course of this project.

we are deeply indebted to our Project Supervisor, **Mrs.Dr.Ch.Santhi Rani,**M.Tech,Ph.D,MISTE,ME, for his expert guidance, mentorship, and unwavering support. Their insightful suggestions and encouragement have been instrumental in the successful completion of this project.

We would also like to express our sincere appreciation to **Dr.B.Nancharaiah**, Head of the Department of **Electronics and Communication Engineering**, for his invaluable guidance and for facilitating access to tools and resources required for this work.

We are profoundly grateful to **Dr. G V K S V Prasad**, Principal of **Usha Rama College of Engineering and Technology**, for his constant encouragement and support throughout the duration of this project.

Our special thanks go to **Dr.SUNEEL KUMAR ASILETI**, Project Coordinator, for his assistance in planning and organizing the project, ensuring that we stayed on track, and providing us with constructive feedback.

We also extend our gratitude to our institution, faculty members, and peers for their direct and indirect contributions, which have greatly enriched this endeavour.

**Project Associates**

## CH.NAGA GOPAL 21NG1A04F3 CH.SAI MANIKANTA 22NG5A0420 S.SRI CHANDRA 21NG1A04E2 Y.SAI RAVI CHANDRA 21NG1A04F5



**VISION :**

To be a pioneer in Electronics and Communication Engineering and research, promoting entrepreneurship and delivering innovative solutions to societal needs

## MISSION:

1. To provide a strong foundation in Electronics and Communication Engineering, preparing students to tackle emerging technological challenges.
2. To drive research in Electronics and Communication Engineering that delivers innovative solutions to societal needs.
3. To promote lifelong learning, empowering students to adapt to the evolving technological advancements.

**PROGRAM EDUCATIONAL OBJECTIVES (PEOs):**

1. Exhibit continuous growth in technical expertise and leadership within the engineering field, while upholding professional ethics.
2. Communicate effectively and manage resources skillfully as members and leaders of the profession
3. Commit to continuous learning and adapt to emerging technologies to meet the evolving needs of society.

**PROGRAM SPECIFIC OUTCOMES (PSOs):**

**Electronics and Communication Engineering graduates will be able to:**

* 1. Develop electronics and communication systems in VLSI, embedded systems, signal processing, and RF communications using advanced tools.
  2. Apply ECE knowledge to design, develop, and test systems, considering societal, environmental, ethical, and economic factors.

**PROGRAM OUTCOMES (POs):**

**PO1: Engineering Knowledge:** Apply knowledge of mathematics, natural science, computing, engineering fundamentals and an engineering specialization as specified in WK1 to WK4 respectively to develop to the solution of complex engineering problems.

**PO2: Problem Analysis:** Identify, formulate, review research literature and analyze complex engineering problems reaching substantiated conclusions with consideration for sustainable development. (WK1 to WK4)

**PO3: Design/Development of Solutions:** Design creative solutions for complex engineering problems and design/develop systems/components/processes to meet identified needs with consideration for the public health and safety, whole-life cost, net zero carbon, culture, society and environment as required. (WK5)

**PO4: Conduct Investigations of Complex Problems:** Conduct investigations of complex engineering problems using research-based knowledge including design of experiments, modelling, analysis & interpretation of data to provide valid conclusions. (WK8).

**PO5: Engineering Tool Usage:** Create, select and apply appropriate techniques, resources and modern engineering & IT tools, including prediction and modelling recognizing their limitations to solve complex engineering problems. (WK2 and WK6)

**PO6: The Engineer and The World:** Analyze and evaluate societal and environmental aspects while solving complex engineering problems for its impact on sustainability with reference to economy, health, safety, legal framework, culture and environment. (WK1, WK5, and WK7).

**PO7: Ethics:** Apply ethical principles and commit to professional ethics, human values, diversity and inclusion; adhere to national & international laws. (WK9)

**PO8: Individual and Collaborative Team work:** Function effectively Formatting Guidelines

**PO9: Communication:** Communicate effectively and inclusively within the engineering community and society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations considering cultural, language, and learning differences

**PO10: Project Management and Finance:** Apply knowledge and understanding of engineering management principles and economic decision-making and apply these to one’s own work, as a member and leader in a team, and to manage projects and in multidisciplinary environments.

**PO11: Life-Long Learning:** Recognize the need for, and have the preparation and ability for i) independent and life-long learning ii) adaptability to new and emerging technologies and iii) critical thinking in the broadest context of technological change. (WK8)

**LIST OF CONTENTS**

|  |  |
| --- | --- |
| **Description** | **Page No** |
| **LIST OF CONTENTS** | vi |
| **LIST OF FIGURES** | viii |
| **LIST OF TABLES** | x |
| **LIST OF ABBREVIATIONS** | xi |
| **ABSTRACT** | xii |
| **CHAPTER 1:INTRODUCTION** | 1-5 |
| 1.1 Introduction | 2 |
| **CHAPTER 2: LITERATURE REVIEW** | 7-13 |
| 2.1 Literature Survey on Fundus Image Processing and Ophthalmic Diagnosis | 7 |
| 2.2 Fundus Image Registration and Processing  Techniques | 7 |
| 2.3 Deep Learning in Fundus Image Analysis | 7 |
| 2.4 Epidemiology and Disease Detection | 8 |
| 2.5 Classification of Fundus Images Using Deep  Learning | 8 |
| **CHAPTER 3: EXISTING METHOD** | 14-21 |
| 3.1 Existing Method | 15 |
| 3.2 Block Diagram | 15 |
| **CHAPTER 4: PROPOSED METHOD** | 22-34 |
| 4.1 Materials and methods | 23 |
| 4.2 VIT branch | 28 |

|  |  |
| --- | --- |
| 4.3 Feature fusion | 29 |
| **CHAPTER 5: RESULTS** | 35-48 |
| 5.1 Evaluation indicators | 36 |
| **CHAPTER 6: CONCLUSION** | 49-51 |
| 6.1 Conclusion | 50 |
| 6.2 Advantages | 50 |
| 6.3 Applications | 51 |
| 6.4 Future Scope | 51 |
| **REFERENCES** | 52-55 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Fig No** | **Figure Name** | **Page No** |
| 3.1 | Deep learning approach for detection and  classification of cataract, glaucoma and retinal disease. | 15 |
| 3.2 | Agumented Images | 17 |
| 3.3 | Pre-processed Images | 18 |
| 3.4 | Block diagram of Inception v4 Model | 19 |
| 3.5 | Samples Having False Positive(FP) and False  Negative(FN). | 20 |
| 3.6 | Accuracy Graph | 20 |
| 3.7 | Model predictions | 21 |
| 4.1 | process of removing black areas | 23 |
| 4.2 | The overall architecture of the CNN-Trans model. | 24 |
| 4.3 | Architecture | 26 |
| 4.4 | The Internal Structure of LSTM | 27 |
| 4.5 | The structure of coordinated attention | 30 |
| 4.6 | The architecture of VIT | 31 |
| 4.7 | Two of the most commonly used feature fusion  methods | 33 |
| 5.1 | Samples Diagram of seven types of eye diseases in  D1 | 36 |
| 5.2 | Results saved to classification\_results.mat | 38 |
| 5.3 | Confusion Matrix for Validation data | 40 |
| 5.4 | Samples Diagram of seven types of eye diseases in  D1 | 41 |
| 5.5 | Results | 45 |
| 5.6 | Confusion Matrix | 45 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table No** | **Table Name** | **Page No** |
| 3.1 | Dataset statistics | 17 |
| 3.2 | Accuracy table with different learning rates along without and with augmentation | 20 |
| 5.1 | Input data normalization | 37 |

**LIST OF ABBREVIATIONS**

|  |  |
| --- | --- |
| **Abbreviation** | **Description** |
| LSTM | Long Short-Term Memory |
| ViT | Vision Transformer |
| ROC | Receiver operating characteristics |
| CNN | Convolution Neural Network |
| OCTA | Optical Coherence Tomography Angiography |
| FAZ | Foveal Avascular Zone |
| DFI | Digital Fundus Imaging |
| CRC | Colorectal Cancer |
| AMD | Age-Realted Macular Degeneration |

**ABSTRACT**

The existence of fundus diseases not only endangers people’s vision, but also brings serious economic burden to the society. Fundus images are an objective and standard basis for the diagnosis of fundus diseases. With the continuous advancement of computer science, deep learning methods dominated by convolutional neural networks (CNN) have been widely used in fundus image classification. However, the current CNN-based fundus image classification research still has a lot of room for improvement: CNN cannot effectively avoid the interference of repeated background information and has limited ability to model the whole world.

In response to the above findings, this proposes the CNN-Trans model. The CNN-Trans model is a parallel dual-branch network, which is the two branches of CNN-LSTM and Vision Transform (ViT). The CNN-LSTM branch uses Xception after transfer learning. As the original feature extractor, LSTM is responsible for dealing with the gradient disappearance problem in neural network iterations before the classification head, and then introduces a new type of lightweight attention mechanism between Xception and LSTM: Coordinate Attention, so as to emphasize the key information related to classification and suppress the less useful repeated background information; while the self-attention mechanism in the ViT branch is not limited by local interactions, it can establish long-distance dependence on the target and extract global features.

Finally, the concatenation (Concat) operation is used to fuse the features of the two branches. The local features extracted by the CNN-LSTM branch and the global features extracted by the ViT branch form complementary advantages. After feature fusion, more comprehensive image feature information is sent to the to the classification layer. Finally, after a large number of experimental tests and comparisons, the results show that: the CNN-Trans model achieved an accuracy of 80.68% on the fundus image classification task, and the CNN-Trans model has a classification that is comparable to the state-of-the-art methods. performance.

**Keywords:** *Fundus image classification Parallel dual branch Attention mechanism Feature fusio CNN.*

## CHAPTER 1 INTODUCTION

* 1. **INTRODUCTION**

With the continuous development of modern lifestyle, computers, mobile phones and other digital screens have become more and more popular. However, staring at electronic screens for a long time can cause eye fatigue, dryness and discomfort. This phenomenon is very common. The main reason is that we When using electronic devices, the frequency of blinking decreases and the eyes are not adequately lubricated. In addition, factors such as light radiation, unhygienic eye use, age decline, etc. can also cause eye damage.

For individuals, eye health problems can lead to less efficient learning, as poor vision can affect work performance and academic performance. For society, people with vision problems may need more frequent eye treatments and glasses, which increases the demand for medical resources and is a significant economic burden. There are many kinds of fundus diseases, which seriously affect human visual health and are one of the main causes of blindness worldwide. Cataracts, diabetic retinopathy, age-related macular degeneration and high myopia are all common blinding diseases. Although there are many types of fundus diseases, many fundus diseases will produce certain signs in the fundus.

However, since most patients have low awareness of these signs, they may ignore eye changes and delay treatment. Therefore, early detection of fundus diseases is of great significance for subsequent treatment. Fundus examination uses a professional fundus camera or scanning instrument to capture fundus images for analysis. These images show the structures inside the fundus of the eye, including the retina, optic nerve, blood vessels, and other important tissues. By observing fundus images, doctors can detect whether the fundus blood vessels are normal and whether there are abnormal expansions, lesions, or bleeding. Fundus imaging has many advantages and is a non-invasive method that does not require any surgery or irritating procedures. Secondly, fundus images provide objective visual information, reducing the impact of subjective factors on diagnosis.

Doctors can assess the status of systemic blood circulation, thereby early detecting and managing systemic health-related problems. Therefore, fundus examination can not only be used to detect eye diseases, such as vitreous, retina, choroid and optic nerve diseases, but also a monitoring window for many systemic diseases [1]. However, clinical diagnosis of fundus diseases relies on professional ophthalmologists. As the number of patients continues to increase, fundus image data is also becoming larger and larger. The diagnosis of fundus data will consume a lot of energy and time of ophthalmologists. And limited professional medical resources should be used in more valuable directions. In order to assist ophthalmologists to make accurate diagnosis based on fundus images, it is necessary to develop a computer-aided diagnosis system (CAD for short) for fundus image classification as soon as possible. At present, in the field of computer vision, the challenges of fundus image classification are mainly reflected in the following aspects [2,3]:

* + 1. Publicly available fundus image datasets are generally small in size. There are two main reasons: a. Fundus images contain sensitive personal information, such as retinal structure, etc. When processing and sharing fundus image data, appropriate privacy protection measures must be taken to ensure that patient privacy and data security are not violated. To protect patient privacy, fundus image data collected by hospitals are usually kept confidential. b. In order to accurately classify fundus images, manual annotation and verification of the data set is the first step, but this requires highly specialized knowledge and experience to ensure the accuracy of diagnostic results, therefore, mobilizing a large number of professional ophthalmologists Participating in the creation and annotation process of data sets is a difficult problem for researchers.
    2. Most fundus image datasets suffer from severe class imbalance problem, which becomes more prominent when the number of classes increases. The direct reason why category imbalance occurs is that the probability of developing fundus diseases is quite different. However, there are various factors that affect the probability of developing fundus diseases[4–6], such as: a.

Geographical factors: The incidence of trachoma is usually higher in warm, dry and poor areas, which are more likely to become hotbeds for the spread of infectious diseases. For example, sub-Saharan Africa, parts of Asia, and some countries in Oceania have long been highincidence areas for trachoma. b. Hygiene factors. In areas with poor sanitary conditions, such as places lacking clean drinking water and sanitation facilities, people are more likely to be exposed to sources of infection, thereby increasing the risk of infection from bacterial or viral fundus diseases. c. Genetic factors, there are racial differences in the occurrence and development of myopia. Studies on the incidence of myopia in different races have found that in rapidly developing economies in East Asia such as China, its prevalence is growing at an alarming rate. d. Age factor. Cataracts are common in middle-aged and elderly people over 50 years old. This is because the organs in the body gradually age and the lens will also undergo degenerative changes, such as turbidity.

Different models of equipment used to obtain fundus images in hospitals may have differences in image resolution, lighting conditions, imaging quality, etc.. For example, some images may be produced by old equipment that has not been updated in time, and there are cases of low pixels and resolutions. It may also be taken in a dimly lit environment, resulting in uneven brightness and darkness in various areas of the fundus image. The low-quality images caused by these factors greatly increase the difficulty of classification.

In recent years, with the rapid development of deep learning technology in the field of computer vision, more and more researchers have begun to use deep learning algorithms to analyze and process fundus images. Deep learning is a machine learning method based on multilayer neural networks that can automatically learn and extract features from large amounts of data to achieve highly accurate image analysis and recognition. In fundus image analysis, deep learning technology can extract features in fundus images by training a large amount of fundus image data, and perform tasks such as classification, positioning and segmentation. In 2019, Imran Qureshi et al. [7] compared related CAD

systems based on statistical parameters for quantitative evaluation. The comparison results showed that accurate development of CAD systems is still needed to assist clinical diagnosis of diabetic retinopathy.

In 2021, Chea et al. [8] found that existing computer-aided systems cannot simultaneously detect multiple major eye diseases. To better understand the multi- category classification of fundus images, they used an optimal residual deep neural processing technology. In 2023, Wen Jingyi et al. [9] believed that the symptoms of chronic kidney disease, hypertension, type 2 diabetes and other diseases could be found based on the specific performance of retinal images, and pointed out that the latest development of AI technology is the use of retinal images to diagnose kidney diseases.

Rapid large-scale screening and prognosis prediction bring great potential. These research results will help establish an accurate computer-aided diagnosis (CAD) system to provide doctors with more accurate diagnosis results and treatment suggestions, thereby improving patients’ diagnosis and treatment experience and treatment effects. However, in the face of the urgent needs of clinical fundus disease screening, diagnosis and other auxiliary medical care, Existing CAD systems generally can only detect a few major eye diseases. There are few related studies on detecting multiple eye diseases and their performance is average. And most studies directly transfer the features learned by the last convolutional layer of CNN to the classification layer.

A problem: (1) It is impossible to effectively avoid the interference of repeated fundus image background information, and ignores the subtle changes in the fundus image that may represent the diseased tissue; (2) The core operation of CNN is the convolution kernel, which has local sensitivity. However, in the fundus image The lesion area may be discontinuous, so it is difficult for CNN to grasp the global features and extract them.

This study contributes a new solution to the two limitations of the above- mentioned fundus image classification research: through a series of algorithm improvements to the convolutional neural network (CNN).

## CHAPTER 2 LITERATURE REVIEW

* 1. **Literature Survey on Fundus Image Processing and Ophthalmic Diagnosis**

Fundus image processing plays a crucial role in ophthalmic diagnosis, particularly in detecting and predicting eye diseases such as diabetic retinopathy, cataracts, and myopia. The following literature survey provides insights into the advancements in fundus image processing, machine learning applications, and deep learning techniques for ophthalmic diagnosis based on the given references.

* 1. **Fundus Image Registration and Processing Techniques**

Yu et al. [1] discuss the progress in fundus image registration technology, which is essential for aligning images taken at different times or from different perspectives. Their study focuses on different registration methods, including intensity-based and feature-based techniques, to enhance the accuracy of image alignment, thus facilitating disease diagnosis.

The book by CRC Press [2] explores signal processing and machine learning applications in biomedical big data, including fundus imaging. It highlights how advanced computational techniques help in image enhancement, segmentation, and feature extraction, leading to improved diagnostic accuracy.

* 1. **Deep Learning in Fundus Image Analysis**

Sengupta et al. [3] provide a comprehensive review of deep learning applications in fundus image processing for ophthalmic diagnosis. Their study covers convolutional neural networks (CNNs) and transfer learning approaches used in detecting various retinal diseases. The authors emphasize the importance of large annotated datasets and robust model architectures in achieving high diagnostic accuracy.

Socia et al. [4] propose machine learning-based detection of trachoma using fundus images. Their study employs convolutional and traditional machine learning algorithms to classify images with high sensitivity and specificity. This research demonstrates the potential of artificial intelligence in automating disease detection in ophthalmology.

* 1. **Epidemiology and Disease Detection**

Foster and Jiang [5] discuss the epidemiology of myopia, analyzing various factors contributing to its prevalence. Their research focuses on population-based studies and genetic predispositions that influence the development of myopia.

Faizal et al. [6] present an automated cataract detection system using adaptive thresholding and a fine-tuned InceptionV3 model. Their method significantly improves early cataract diagnosis by leveraging deep learning-based feature extraction and classification techniques applied to anterior segment eye images.

Qureshi et al. [7] review recent developments in diabetic retinopathy detection methods. The discusses feature-based approaches and deep learning models for early detection and grading of diabetic retinopathy, highlighting the advantages of automated diagnostic systems over traditional manual assessments.

* 1. **Classification of Fundus Images Using Deep Learning**

Chea and Nam [8] propose a deep learning-based classification system for fundus images to detect various eye diseases. Their study evaluates CNN architectures trained on large datasets, demonstrating improved diagnostic accuracy and reliability compared to conventional methods.

Wen et al. [9] explore the use of retinal images for artificial intelligence-driven kidney disease detection. The study highlights how retinal biomarkers can serveas indicators of systemic diseases, opening new avenues for AI-assisted diagnosis beyond ophthalmology.

Choi et al. [10] develop a multi-categorical deep learning neural network for classifying retinal images. Their pilot study shows that deep learning models can effectively differentiate multiple eye conditions, even with small datasets, suggesting the potential for future improvements with larger, more diverse datasets.

Diaz-Pinto et al. [11] extensively validate CNNs for automatic glaucoma assessment using fundus images. Their study highlights the effectiveness of deep learning in diagnosing glaucoma and achieving high accuracy in automated detection.

Balaji et al. [12] compare the foveal avascular zone in diabetic retinopathy, high myopia, and normal fundus images, providing insights into the structural variations among these conditions through ophthalmic imaging analysis.

Junayed et al. [13] introduce CataractNet, an automated cataract detection system using deep learning for fundus images, demonstrating high performance in identifying cataracts with minimal manual intervention.

Butt et al. [14] employ hybrid deep learning features to detect diabetic retinopathy from fundus images, demonstrating how feature fusion techniques enhance the precision and recall of automated diagnostic systems.

Shyamalee and Meedeniya [15] develop a CNN-based fundus image classification system for glaucoma identification, showcasing the potential of deep learning in detecting early-stage glaucoma.

Ilesanmi et al. [16] conduct a systematic review of retinal fundus image segmentation and classification methods using CNNs, summarizing recent advancements and challenges in the field.

Lu et al. [17] propose a multilabel classification method for multiple fundus diseases using a CNN with squeeze-and-excitation attention, improving the interpretability of deep learning models in ophthalmic diagnosis.

Wang et al. [18] introduce COVIDX-LwNet, a lightweight network ensemble model for COVID-19 detection using chest X-ray images, illustrating the adaptability of deep learning models across different medical imaging domains.

Tartaglione et al. [19] explore the challenges of detecting COVID-19 from chest X-rays with deep learning, emphasizing the difficulties posed by small datasets and class imbalances.

Chollet [20] presents exception, a deep learning model utilizing depth wise separable convolutions, which has been widely adopted in medical image analysis, including ophthalmic imaging.

Greff et al. [21] conducted an in-depth analysis of the Long Short-Term Memory (LSTM) architecture, investigating the search space and optimization challenges in training LSTMs for sequential tasks. Their study provided insights into the advantages of LSTMs in capturing long-range dependencies, leading to improvements in various neural network applications.

Hou et al. [22] introduced Coordinate Attention, an efficient attention mechanism designed to enhance feature representation in mobile networks. Their work demonstrated how coordinate attention could improve the performance of deep networks, particularly in mobile-friendly architectures, by capturing long-range dependencies while preserving spatial details.

Dosovitskiy et al. [23] proposed the Vision Transformer (ViT), a groundbreaking deep learning model that treats images as sequences of patches, similar to the token-based approach in NLP. Their findings indicated that transformers could outperform conventional CNNs when trained on large-scale datasets, revolutionizing image classification tasks.

Carion et al. [24] presented an end-to-end object detection model using transformers, termed DETR. This work introduced a novel way of formulating object detection as a direct set prediction problem, leading to improved accuracy and robustness against variations in object scale and occlusion.

Joshi and Masilamani [25] proposed a transfer learning-based approach for detecting abnormal fundus images. Their model leveraged pre-trained deep learning architectures to achieve high classification accuracy, making it a valuable tool for early diagnosis of retinal disorders.

Raza et al. [26] employed the Inception-V4 deep learning model for classifying eye diseases and detecting cataracts using digital fundus imaging. Their study emphasized the importance of automated diagnostic tools in ophthalmology, highlighting the potential of deep learning in medical image analysis.

Lai et al. [27] investigated the application of CNNs in cataract detection using digital camera images. Their work showcased the feasibility of leveraging deep learning for automated cataract screening, improving accessibility to early diagnosis and treatment.

Smitha and Jidesh [28] utilized a semi-supervised Generative Adversarial Network (GAN) for classifying multiple retinal disorders from enhanced fundus images. Their approach demonstrated the effectiveness of GANs in improving classification accuracy in medical imaging tasks.

Pan et al. [29] developed a classification model using Inception V3 and ResNet-50 for early diagnostics of fundus diseases. Their research emphasized the potential of combining multiple deep learning architectures to achieve superior classification performance.

Shamsan et al. [30] proposed an automatic classification system for color fundus images to predict various eye disease types based on hybrid feature extraction methods. Their approach demonstrated improved accuracy and robustness in disease classification.

Ali et al. [31] introduced AMDNet23, a hybrid CNN-LSTM deep learning approach with enhanced preprocessing techniques for detecting Age-Related Macular Degeneration (AMD). Their work highlighted the effectiveness of combining CNNs with LSTMs for sequential feature learning in medical image analysis.

Selvaraju et al. [32] proposed Grad-CAM, a gradient-based visualization xtechnique for deep learning models. Their work enabled interpretability in CNN- based models, providing valuable insights into decision-making processes in deep networks.

Carion et al. [33] further refined object detection methodologies using transformers, demonstrating their efficiency in end-to-end detection tasks. Their study paved the way for more advanced transformer-based vision models.

Chen et al. [34] explored generative pretraining from pixels, highlighting the effectiveness of self-supervised learning techniques in training large-scale deep learning models for image analysis tasks.

Dosovitskiy et al. [35] expanded on the capabilities of Vision Transformers, reinforcing their dominance in image classification tasks and emphasizing their scalability and performance improvements over CNNs.

Parikh et al. [36] introduced a decomposable attention model for natural language inference, demonstrating its effectiveness in text classification and interpretation tasks. Their approach contributed significantly to the development of attention mechanisms in deep learning.

Cheng et al. [37] examined LSTM-based networks for machine reading, emphasizing their ability to understand complex textual structures and dependencies. Their work showcased the adaptability of LSTMs in various NLP applications.

Gal and Ghahramani [38] proposed using dropout as a Bayesian approximation to represent model uncertainty in deep learning. Their study contributed to the understanding of uncertainty quantification in neural networks, improving robustness and generalization capabilities.

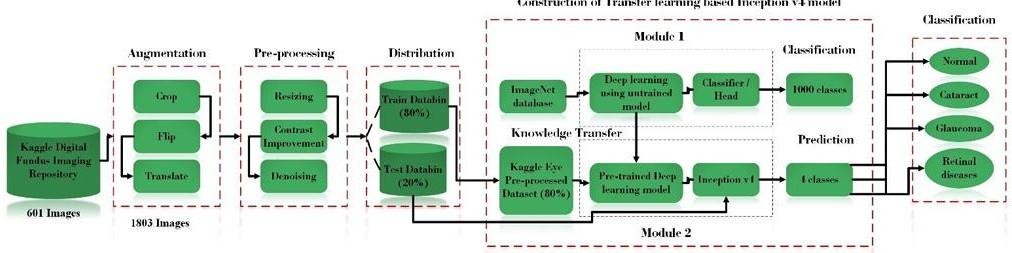
Shamsan et al. [39] proposed an automatic classification system for color fundus images to predict various eye disease types based on hybrid feature extraction methods. Their approach demonstrated improved accuracy and robustness in disease classification.

Selvaraju et al. [40] proposed Grad-CAM, a gradient-based visualization technique for deep learning models. Their work enabled interpretability in CNN- based models, providing valuable insights into decision-making processes in deep networks.

The reviewed literature highlights significant advancements in fundus image processing, with a strong focus on deep learning and machine learning applications. The studies demonstrate the increasing role of AI-driven techniques in enhancing diagnostic accuracy and automating disease detection in ophthalmology. Future research should aim to improve model generalization, address dataset limitations, and integrate multimodal medical data.

## CHAPTER 3 EXISTING METHOD

* 1. **EXISTING METHOD**

****

**Fig.3.1 Deep learning approach for detection and classification of cataract, glaucoma and retinal disease.**

As Shown in the Figure 3.1 Eye is a sensitive part of human body so we need to detect the seriousness of eye disease and their types. Block diagram shown in Fig.3.1 shows a deep learning approach for detection and classification of cataract, glaucoma and retinal disease. This system also successfully detects a normal eye when no disease is found. We trained and tested the inceptionV4 deep learning model with pre-trained weights and different learning rates. Overall best accuracy of 96.6% is achieved by augmented dataset.

# Block Diagram

1. Dataset Acquisition
   1. Digital Fundus Imaging: The fundus of the eye consists of the inner surface of the eye towards the lens and comprises the retina, optical disc, macula, fovea and back pole. The funds can be tested using a picture of the ophthalmoscopy. Fundus photos are ocular documentation that shows a patient's retina appearance. For the monitoring of the course of certain eye illnesses, optometrists, ophthalmologists, orthoptists and other qualified medical practitioners utilize fundus photography. Fundus pictures are also utilized for documenting illness abnormalities affecting the eye and/or to track disease development. Fundus photos capture the retina, the tissue in our eyes that converts the optical pictures that are seen to our brains in the electrical impulses. The retina

may be directly imaged as the pupil is employed as an entry and exit to the light and imaging rays of the fundus camera.

2) Dataset: The eye diseases dataset has been obtained from publicly available source Kaggle where normal eye images along with the diseased fundus images of cataract, glaucoma and retina are provided. There is total 601 images available where 300 images are normal and the rest are diseased images. 100 images belong to cataract, 101 images to glaucoma and 100 images to retinal diseases.

1. Dataset Augmentation

Data Augmentation is a technique for artificially increasing the size of a training set by generating changed data from existing data.

Here the dataset was augmented by cropping, flipping, and translation of all the raw images as shown in Fig.3.2 and the detail is listed below,

* 1. Cropping is the process of removing undesirable portions from an image

or illustration.

* 1. A flipped image is a static or moving image created by mirroring an

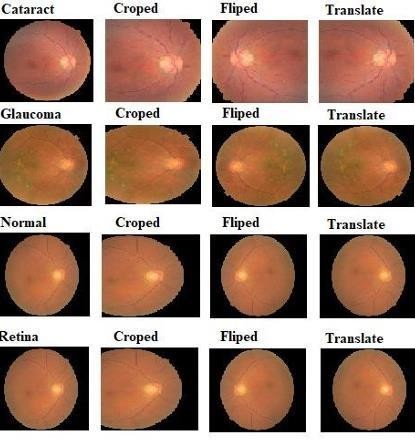
original across a horizontal axis.

* 1. During image translation, each pixel of the item must be adjusted in the

identical orientation and similar length.

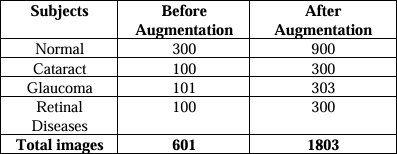
* 1. The initial object is known as the pre- image, and the item after

translation is known as the image.



**Fig.3.2 Agumented Images**

Table 3.1 shows dataset statistics before and after the augmentation

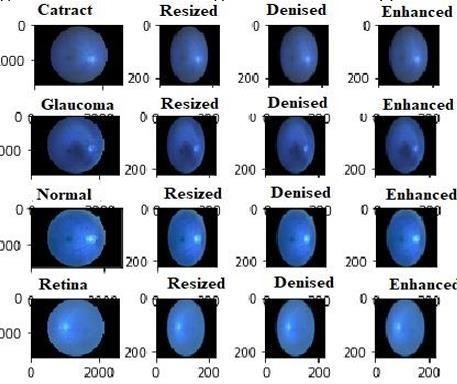


C. Pre-processing 1803 The augmented dataset is pre-processed.

Following steps are applied which is also shown in Fig.3.2

1. Each image has dimension of 2592\*1728 which has been resized to a fixed dimension of 244\*244.
2. After resizing, contrast enhancement has been applied to each of the image where the quality of the is enhanced so that understanding of each image by DLM could be easy.
3. The last step of the pre-processing is the removal of the noise from every image. The averaging filter has been applied to each image before input into deep learning algorithm.

Dataset Distribution We randomly separated the dataset into 80% and 20% for training and testing purpose respectively for both augmentation and without augmentation dataset approach.



**Fig.3.3 Pre-processed Images**

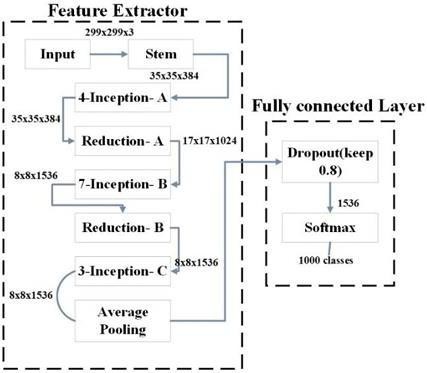
1. InceptionV4 Deep Learning Model Up to 2014, the standard Convolutional neural network (CNN) structure was stacked with convolutional layers, max pooling, followed by one or more fully connected layers (FCL).This has limitations.
2. Memory footprint is rather large.
3. High computational demand.
4. Overfitting is a risk.
5. Gradients that fade and erupt.

Here we used a variant of CNN named as InceptionV4 deep learning model with its pre-trained wights on ImageNet dataset (1000 classes) for the detection and classification of eye diseases. The complete architecture of InceptionV4 is shown in Fig.3.4 Inception v4 is a deep convolutional network design that has been demonstrated to attain excellent performance at a minimal computation complexity. It is a pure Inception version with residual connections that performs better than Inception-ResNet-v2. Its functionality is comparable to

that of the most recent generation Inception v3 network and its main contribution are as follows:

1. Concatenate all features after filtering the same region using various kernels.
2. Reduce the computation by introducing a bottleneck as a dimension reduction.
3. Implement Batch Normalization.
4. Use a tiny kernel and an asymmetric kernel to make your network more efficient.
5. Smoothing of labels.
6. Substituted filter concatenation.
7. Reduced training time.

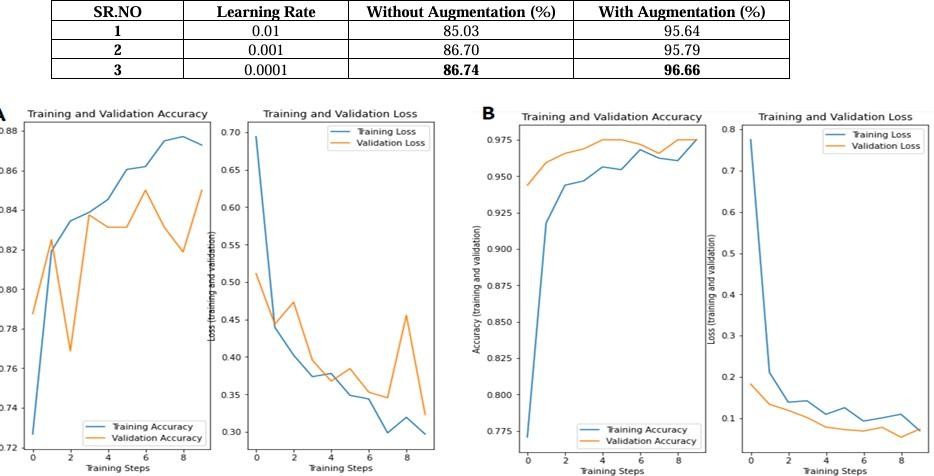
viii. Potential network "death"—training instabilities



**Fig 3.4 Block diagram of Inception v4 Model**

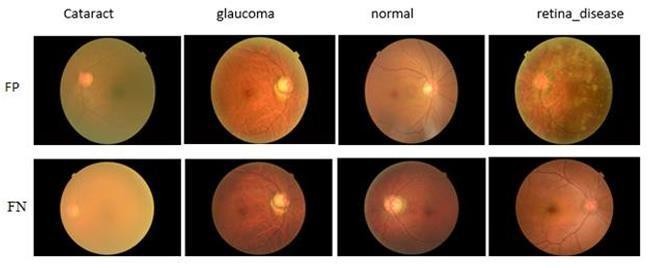
1. Classification Here, the transfer-learning approach is applied to classify four types of digital images (cataract, glaucoma, and retinal) of the digital fundus, i.e., normal and disordered eye images as shown in figure 3.4 and scientifically examine the effect of DA on DLM and try to address the issue of whether we should increase the data in order for melanoma classification to be more effective. Inception v4 trained using the same features retrieved by its first layers from the augmented and original data sets.

**Table 3.2 Accuracy table with different learning rates along without and with augmentation**

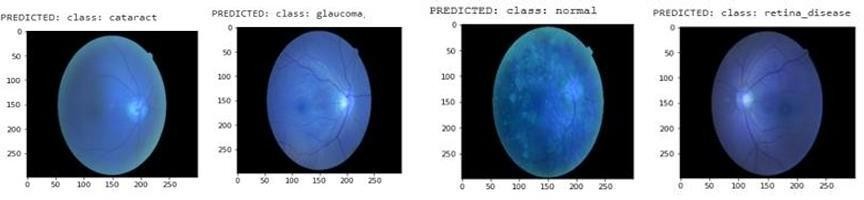
****

**Fig. 3.6 Accuracy Graph (A) Without Augmentation (B) With Augmentation at**

**learning rate of 0.0001.**

****

**Fig. 3.5 Samples Having False Positive(FP) and False Negative(FN).**



**Fig. 3.7 Model predictions.**

Fig. 3.5 shows the training and validation accuracy graph by keeping the learning rate 0.0001. Fig. 3.5(A) shows that intially 86.74% accuracy was achieved which was initially low and gradually increased.

Whereas validation accuracy was initially low and improves as the epochs increased. Similarly, Fig.3.5(B) shows the training and validation accuracy with augmented dataset.

It can be seen that model has achieved the significant accuracy from the beginning and continuously improved it until last epoch. Fig. 3.6 shows the results of each class where our model has false prediction.

The first row shows the results where model has false positive predictions and the second row shows the examples of false negative predicted by the model.

A false positive is an outcome in which the model forecasts the positive class erroneously. A false negative is an outcome in which the model forecasts the negative class inaccurately.

Fig. 3.7 shows model predictions, when randomly test images from each class are given to the model.

## CHAPTER 4 PROPOSED METHOD

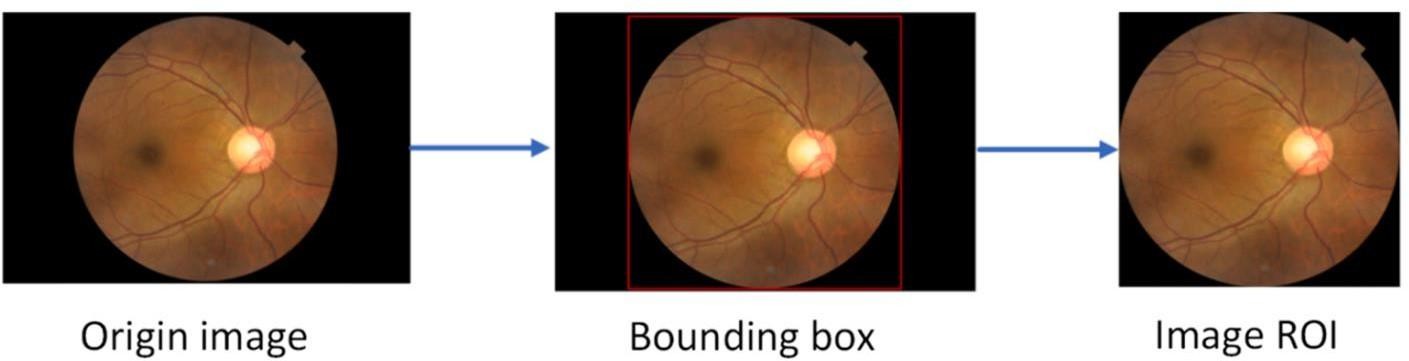
* 1. **Materials and methods**

In this chapter, the datasets used for training and testing and the deep learning model used in this study will be discussed. Datasets are discussed further in Section 5.1. Likewise, Section 5.2 discusses the proposed classification method for fundus images.

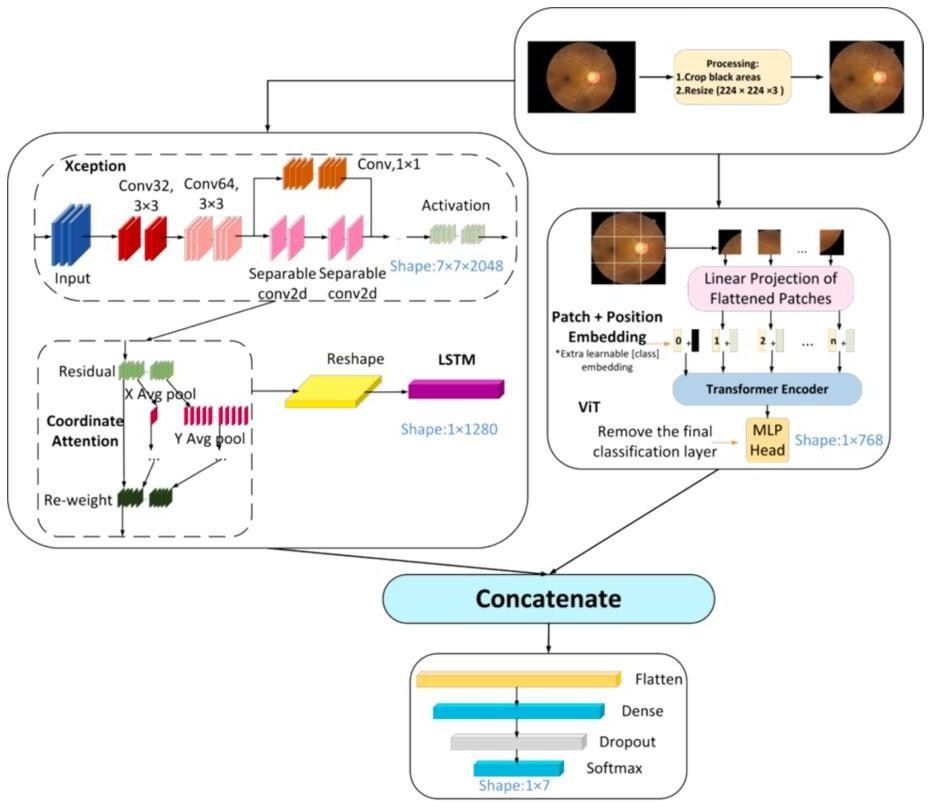
**Proposed method**

The proposed fundus image classification model based on attention mechanism and feature fusion which is named CNN-Trans model. Its architecture is shown in Fig 4.1 The model is a parallel dualbranch network. Firstly, the patient’s fundus image is cropped to remove the black area, and the size is reset to 224 × 224 × 3, and then normalized.

After the above preprocessing steps are completed, it is ready to be sent to the next model. Then, on the one hand, the preprocessed pictures enter the CNNLSTM model branch. Firstly, a feature map with a size of 7 × 7 × 2048 is extracted through pre-trained Xception, and then the coordinated attention module decomposes channel attention into two onedimensional feature encoding processes, and aggregates features along two spatial directions to obtain a feature map. Then, the feature maps are encoded as a pair of orientation-aware pairs and a location-sensitive attention map to augment the representation of objects of interest, these steps can effectively improve the processing efficiency of the input feature maps. After this, the feature map of the coordinated attention



**Fig.4.1 process of removing black areas**.



**Fig.4.2 The overall architecture of the CNN-Trans model**.

module is input to the reshape layer, which changes the shape from (h, w, c) to (h

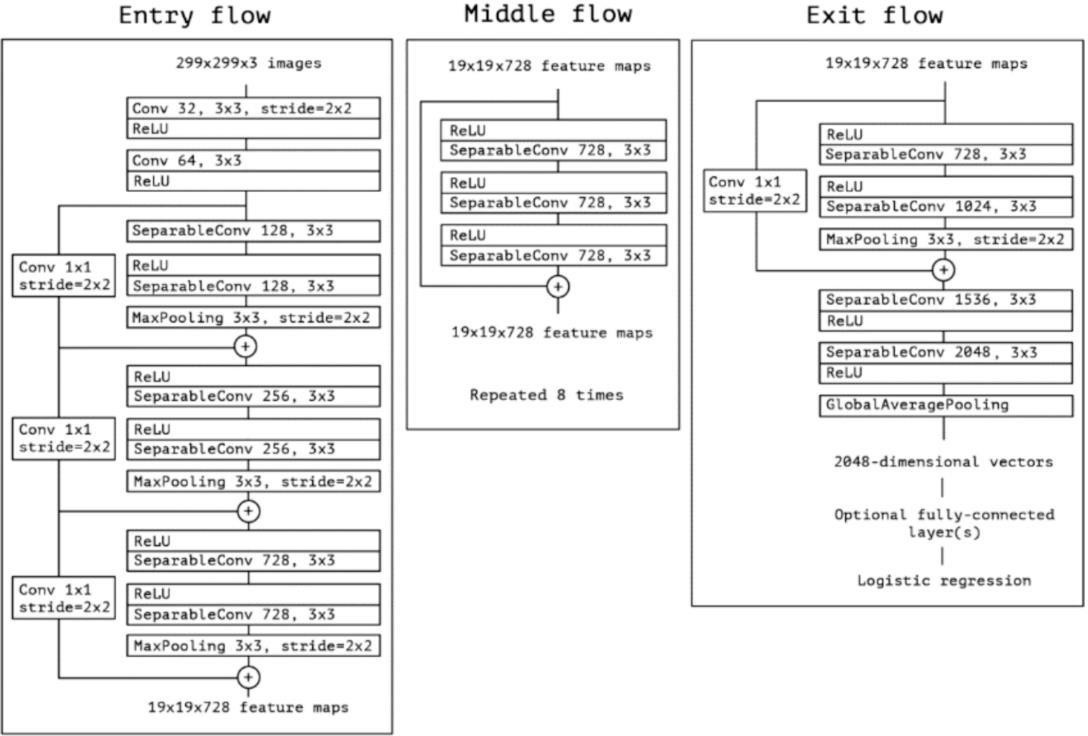
\* w, c) input to LSTM. On the other hand, the preprocessed image enters the ViT branch. First, the input image is divided into image blocks and serialized, position encoding is added, then adds learnable embedding vectors, and inputs them into the encoder for encoding, and finally outputs Learnable embedding vectors are used for subsequent classification. Finally, the features extracted by the Xception branch and the features extracted by the ViT branch are concat spliced and input to the reconstructed classification head, that is, a Flatten layer, a Dense layer with an L2 regularization parameter of 0.01 and 512 output units, and a Dropout layer

with a dropout rate of 0.5, a Dense layer with a SoftMax function and an output unit of 7. 3.2.1. CNN-LSTM branch The CNN-LSTM branch mainly consists of three parts: Xception, LSTM and coordinated attention mechanism, which is also part of our previous research work [18]. (1) Xception The limited available data discourages the use of large CNN models, and very large models such as DenseNet-121 may negatively impact generalization ability [19].

Therefore, when we choose the model, we consider the lightweight and simple model as much as possible, and use the model with better initial classification effect as the feature extractor through simple experiments. In the end, we chose Xception as the original feature extractor: Xception [20] is an improvement to Inception-v3 proposed by Google after Inception in 2016. The structure is shown in Fig. 4.2 Its full name is “Extreme Inception”, that is, Limit Inception. Xception uses depthwise separable convolution (Depthwise Separable Convolution) instead of conventional convolution operations, which is different from Inception. Depth-separable convolution divides the standard convolution operation into two steps: depthwise convolution and pointwise convolution.

The former focuses on the spatial features within the input channel, while the latter focuses on the relationship between different channels. In this way, the number of parameters and calculations in the model can be greatly reduced, the computational efficiency of the model can be improved and the risk of overfitting can be reduced. In addition, Xception also introduces a special extreme Inception block, which replaces the standard convolution operation in the branch convolutional layer of the Inception block with a depthwise separable convolution. The role of this extreme Inception block is to further reduce the number of parameters and calculations in the model and improve the accuracy of the model.

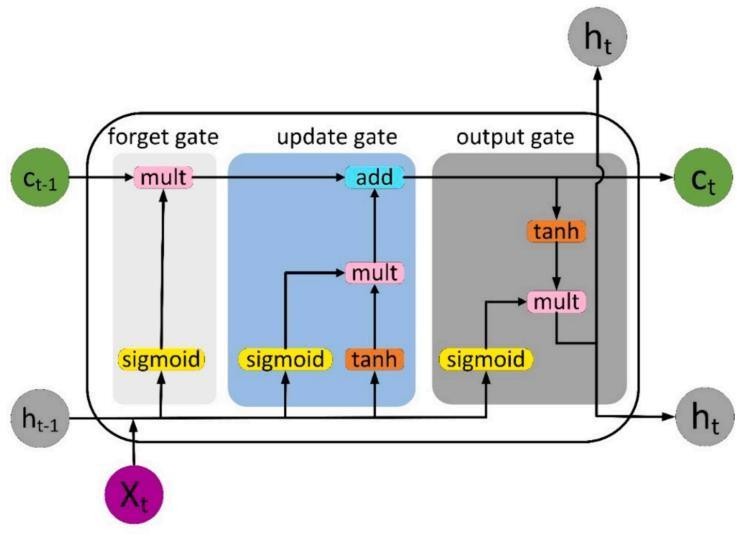
Xception has achieved very good performance in many computer vision fields, such as image classification, target detection, semantic segmentation and other tasks. In the ImageNet image classification challenge, Xception achieved an accuracy of 0.790 and 0.945 on the Top-1 and Top-5 error rates, respectively,

surpassing the most advanced ResNet model at the time. (2) LSTM. Long Short- Term Memory (LSTM) is a common Recurrent Neural Network (RNN) architecture. The emergence of LSTM is mainly due to the problem of gradient disappearance or explosion in traditional RNN.

**Fig.4.3 Architecture**

As shown in above Figure 4.3 it is a RNN based on a gating mechanism. The structure of LSTM consists of a group of special neuronal units that can selectively forget, retain or add information to achieve long-term memory.

The basic units of LSTM include forget gate, input gate and output gate, which fuse the input data and the information of the previous state through nonlinear transformation, and decide whether to retain.



**Fig.4.4 The Internal Structure of LSTM.**

forget certain information through the gating mechanism. The forget gate controls the forgetting of the past state, the input gate controls the input of new information, and the output gate controls the output of the current state. During the training process of LSTM, these gates are trained to automatically adapt to the pattern of input data and the corresponding context information. The value range of the gate is (0,1), controlled by the Sigmoid function, where Xt refers to the current input, ct and ct-1 denote the new and previous cell states, respectively, ht and ht1 are the current and previous outputs, respectively.

The internal structure of LSTM is shown in Fig.4.4 LSTM adds a cell state to save the long-term state, which is its main difference from RNN, which can remember the previous information and connect it to the currently obtained data. LSTM is mainly used to process sequence data, which can capture the long-term

dependence in the sequence, while CNN is mainly used to process image data, which can capture the spatial local features in the image. Combining the two and making full use of their respective characteristics can effectively improve the training speed and efficiency of the model, and maintain the state information of the features encountered in the previous generation of image classification. (3) Coordinate Attention Coordinate Attention (CoordAtt for short) is a plug-in proposed by Qibin Hou et al. [22] inspired by the Squeeze-Excitation (SE) attention module in 2021.

A ready-to-use mobile network attention mechanism, coordinated attention is different from channel attention, it uses two one-dimensional feature encoding processes to aggregate features in different spatial directions to capture long-range dependencies and precise location information at the same time, and ultimately achieve enhanced object of interest expressed purpose. Qibin Hou et al. found that compared with the most popular SE attention module in mobile networks, the coordinated attention mechanism is simple, can be flexibly inserted into the network, and has almost no computational overhead. While taking advantage of the modularity, it is also able to capture the long-term dependencies of precise position information.

* 1. **VIT branch**

Vision Transformer (ViT for short) is a method of using the Transformer model to process image data [23], proposed by Alexey Dosovitskiy et al. in 2020. The core idea of ViT is to divide the image into several small blocks, then convert these small blocks into vectors, and input these vectors into the Transformer model for processing. The ViT architecture is shown in Fig.4.6, which mainly consists of a module that splits the input image into blocks, a layer that embeds blocks and positions, a Transformer encoder, and an MLP classification head. Specifically, ViT first divides the image into some small blocks of the same size, and then maps each small block into a vector through a fully connected layer.

These vectors are treated as tokens in the sequence and fed into the Transformer encoder for processing. In the Transformer encoder, each token is

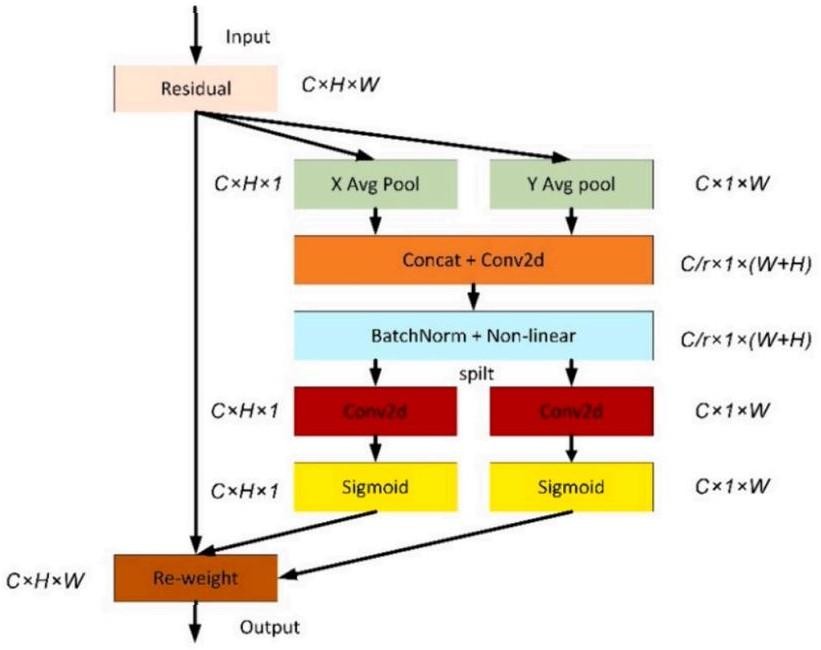
encoded through a multi-layer self-attention mechanism and a feed-forward neural network, and a contextdependent vector representation is obtained. Finally, these vectors are passed through a fully connected layer for classification or regression. In this way, ViT does not need to use traditional convolutional neural networks, and can perform tasks such as classification and detection on images using only Transformer, and has achieved good results on some standard computer vision datasets. When training on large amounts of data in the early stage, the performance of ViT is also comparable to the performance of CNN on small or medium-sized datasets [24].

* 1. **Feature fusion**

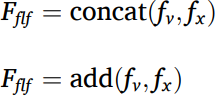
In the feature fusion of image processing, the two commonly used feature fusion (FLF) methods are concatand add. addThe method increases the information content of the image features, but does not increase the dimensionality of the image.

On the other hand, concatthe method merges the number of channels to increase the number of channels describing the image, thereby keeping the relevant information of each feature unchanged. If xthe dimension of the sum of the two input features is p, ythen qthe zdimension of the output feature is p + q. Equations (2) and (3) are related mathematical expressions.

In this study, since the number of channels for extracting features from the two parallel branches of ViT and Xception is different, we use concat to fuse the features extracted by the two. as the picture shows. Fusion( xy ) The number of channels is Feature(x) the sum of the number of channels of and Feature(y).

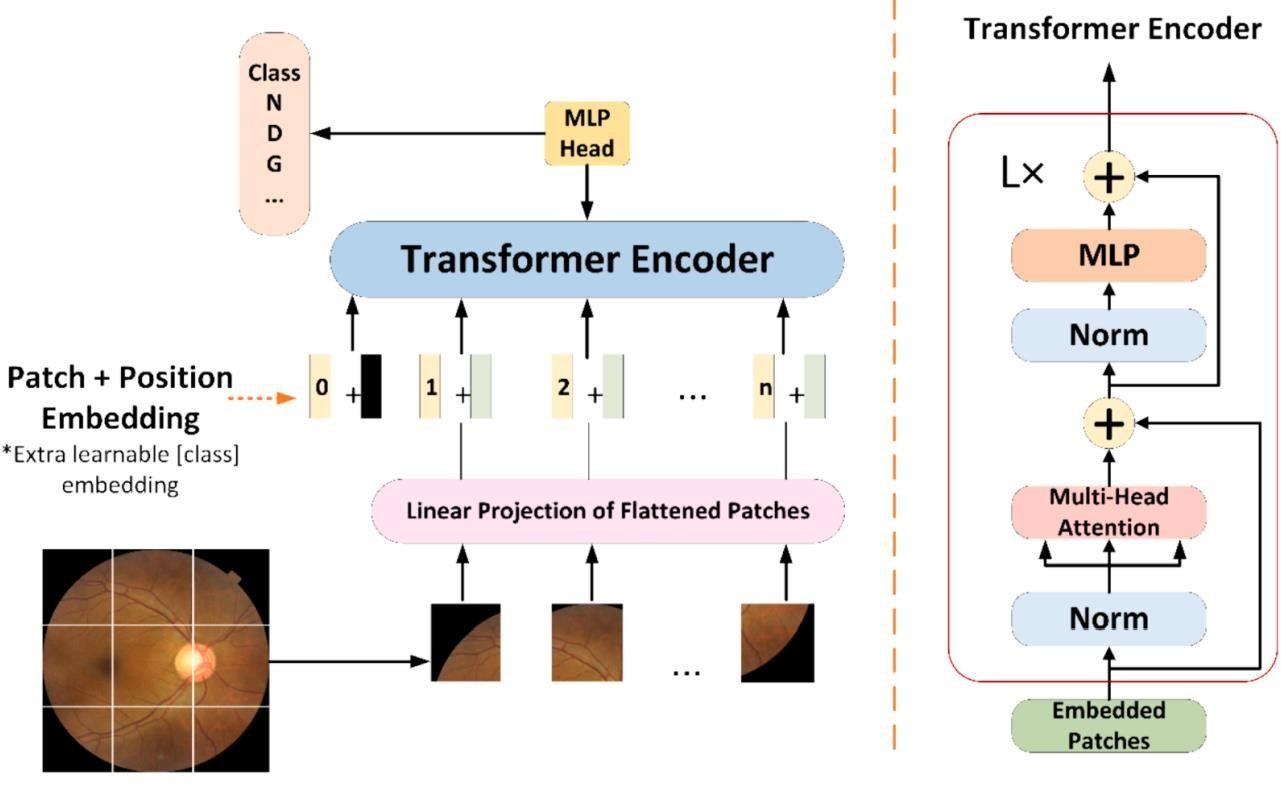


**Fig.4.5 The structure of coordinated attention.**

****

As shown in above Fig 4.5 Among them, fv the features extracted fx by ViT are the features extracted by Fflf Xception, which is the fused feature set. Concat fusion in the CNN-Trans model is: in the branch CNN-LSTM and another branch ViT, they can both perform feature extraction on the input medical image data and output a vector representation. These two vector representations can represent different understandings of input data by CNN-LSTM and ViT respectively.

In order to fuse the different understandings extracted by the two, that is, the features of medical images, and use the fused results for final classification. At this point, this article uses the concat operation to splice these two vectors along a certain dimension, and the resulting new vector will be a tensor of shape (N, c1 + c2), where N represents the number of samples, and c1 and c2 are respectively Represents the vector length of CNN-LSTM and VIT output.

This new fused tensor contains the features of CNN-LSTM and VIT, is a richer vector representation, and is used for the final classification task. It should be noted that the contact fusion method needs to normalize or scale the feature representation output by each model before splicing to avoid large deviations in the spliced results. At the same time, the contact fusion method also has some shortcomings, such as increasing the amount of calculation and storage of the model, so it needs to be selected and optimized according to specific tasks and data conditions.

**Fig.4.6 The architecture of VIT**

As shown in above Fig 4.6 Vision Transformer (ViT) is a deep learning model designed for image classification tasks using the transformer architecture, which was originally developed for natural language processing (NLP). Unlike traditional convolutional neural networks (CNNs), which extract hierarchical spatial features using convolutional filters, ViT treats an image as a sequence of patches and processes them similarly to how transformers handle words in text. This ability to capture long-range dependencies across the entire image makes ViT highly effective for medical image classification, including the analysis of fundus images for detecting eye diseases like diabetic retinopathy and glaucoma.

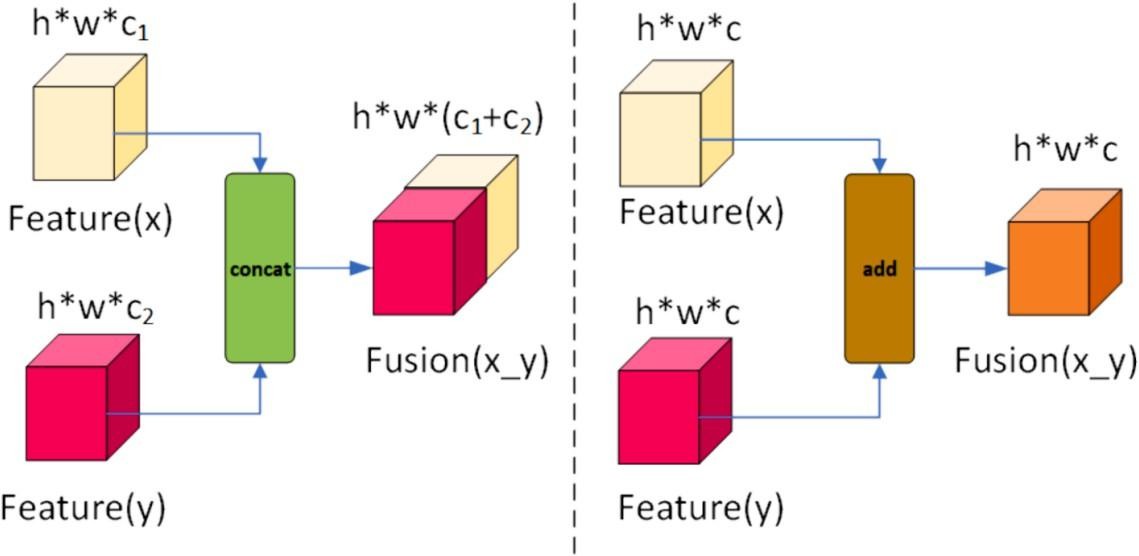
The first step in ViT is to divide the input image into fixed-size, non-overlapping patches. Each of these patches is flattened into a one-dimensional vector and passed through a linear projection layer to transform them into feature representations. Since transformers do not inherently understand spatial relationships, positional embeddings are added to the patch embeddings to retain information about the original structure of the image. Additionally, an extra learnable classification token is introduced, which will later be used for making the final prediction.

Once the patch embeddings are prepared, they are passed through the transformer encoder, which is the core processing unit of ViT. The encoder consists of multiple layers, each containing a multi-head self-attention mechanism and a feed-forward network. The self-attention mechanism allows each patch to interact with every other patch, enabling the model to capture global contextual relationships. This is a major advantage over CNNs, which focus only on local patterns through convolutional kernels. Each encoder layer also includes normalization layers and residual connections, which help in stabilizing training and improving performance.

After passing through multiple transformer encoder layers, the output embeddings are processed by a Multi-Layer Perceptron (MLP) head. The classification token, which has learned information from all patches, is used to predict the final class of the image. In the case of fundus image classification, the model can differentiate between different eye diseases such as normal, diabetic retinopathy, and glaucoma. By leveraging the power of self-attention and global feature extraction, ViT is able to achieve high accuracy in medical image analysis, particularly when trained on large-scale datasets.

One of the key advantages of ViT is its ability to capture global dependencies in an image without being limited by local receptive fields, as seen in CNNs. This makes it especially useful for medical imaging tasks where small and subtle patterns may be distributed across the image. However, ViT typically requires large amounts of training data to perform well. Unlike CNNs, which rely on strong inductive biases to generalize from small datasets, ViT benefits significantly from pretraining on large datasets such as ImageNet before fine-tuning on specific medical imaging tasks.

Despite its high computational cost, ViT has demonstrated superior performance compared to CNNs in various image classification tasks when sufficient training data is available. The flexibility of transformers also allows ViT to be adapted for different medical applications, including disease detection and segmentation. As deep learning continues to evolve, the Vision Transformer represents a promising direction for advancing automated medical image analysis, providing highly accurate and efficient solutions for diagnosing eye diseases using fundus images.

****

**Fig.4.7 Two of the most commonly used feature fusion methods.**

As shown in the above Fig 4.7 given image illustrates two fundamental feature fusion techniques used in deep learning: concatenation-based fusion and addition-based fusion. These methods are widely applied in multi-stream neural networks, where features extracted from different sources (e.g., different layers, modalities, or models) need to be combined to enhance learning performance.

On the left side of the image, the concatenation-based fusion technique is shown. Here, two feature tensors, Feature(x) and Feature(y), with dimensions h×w×c1and h×w×c2 , respectively, are combined along the channel dimension. This results in a fused feature map with dimensions h×w×(c1+c2). The concatenation operation preserves the individual characteristics of both feature maps, allowing the network to retain diverse feature representations. This technique is commonly used in architectures like DenseNet, where feature reuse is crucial for improving network efficiency.

On the right side, the addition-based fusion technique is depicted. In this case, both feature maps Feature(x) and Feature(y) have the same shape (h×w×c), allowing an element-wise addition operation. This results in a final fused feature map of the same dimension, h×w×c. Unlike concatenation, addition-based fusion blends the feature maps by averaging or summing their respective values, effectively integrating the complementary information from both sources. This method is widely used in residual networks (ResNets), where it facilitates gradient flow and helps prevent the vanishing gradient problem.

While concatenation-based fusion retains a richer feature representation by preserving all information from both inputs, it increases the computational complexity due to a higher number of parameters. In contrast, addition-based fusion is computationally efficient and maintains the original feature space dimensions, but it may lose some distinct feature information if the features are not well-aligned. The choice between these two fusion techniques depends on the specific requirements of the deep learning model, such as computational efficiency, feature diversity, and model architecture.

## CHAPTER 5 RESULTS

* 1. **Evaluation indicators**

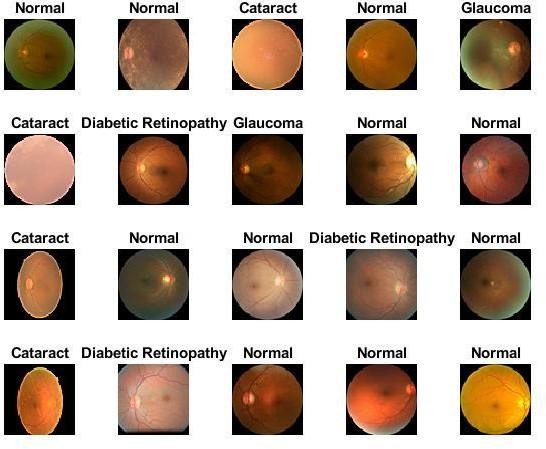
For each image, if the predicted value of a certain label is greater than 0.5, it will be identified as a hard label in the experiment. According to the predicted label and the real label, the confusion matrix of each category can be obtained. Performance indicators based on confusion matrix calculations, such as Precision, Sensitivity (Recall), Specificity, F1, and Accuracy are still essential:

Precision = TP/ (TP + FP)

Sensitivity(Recall) = TP/( TP + FN) Specificity = TN /(TN + FP)

F1 − score = 2 × (Precision × Recall )/(Precision + Recall)

Accuracy = (TP + TN )/(TP + FN + TN + FP) TP, TN, FP, and FN are elements of the confusion matrix.



**Fig.5.1 Samples Diagram of seven types of eye diseases in D1**

As shown in the above Fig 5.1 given diagram presents a collection of fundus images of the human eye, each labeled with its corresponding condition, including Normal, Cataract, Diabetic Retinopathy, and Glaucoma. Fundus images capture the interior surface of the eye, including the retina, optic disc, and blood vessels, and they are widely used in ophthalmology for diagnosing various eye diseases. The labels above each image indicate whether the eye is healthy or affected by a specific condition, providing essential ground truth data for medical image analysis. These labeled images play a crucial role in training machine learning models for automated disease classification and early detection.

Each category in the diagram represents a distinct eye condition with characteristic features. The Normal images depict a healthy retina with no visible abnormalities, serving as a reference for comparison. The Cataract images show clouding of the eye’s lens, which affects vision clarity and can be detected through fundus imaging. Diabetic Retinopathy images display signs of retinal damage caused by diabetes, such as microaneurysms, hemorrhages, and exudates, which can lead to vision loss if untreated. Glaucoma images illustrate potential optic nerve damage, often identified through changes in the cup-to-disc ratio and thinning of the retinal nerve fiber layer. Detecting these conditions early is critical to preventing severe vision impairment.

The primary purpose of this diagram is to highlight the visual differences among these conditions and support the development of AI-driven diagnostic models. By utilizing deep learning techniques such as Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), researchers can extract meaningful features from these images and classify them accurately. These AI-based approaches can significantly assist ophthalmologists by providing rapid and reliable diagnoses, reducing the burden of manual screening. The integration of such models into clinical practice can improve early detection, treatment planning, and patient outcomes, making them valuable tools in modern healthcare.

Training on single GPU.

Initializing input data normalization.

Table 5.1 Input data normalization.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Epoch | Train Loss | Train Accuracy | Validation Loss | Validation Accuracy |
| 1 | 1.6347 | 0.431 | 1.3356 | 0.4996 |
| 2 | 1.2019 | 0.5853 | 1.0809 | 0.6183 |
| 3 | 1.0015 | 0.6581 | 0.937 | 0.6624 |
| 4 | 0.8796 | 0.699 | 0.8233 | 0.7075 |
| 5 | 0.7917 | 0.7322 | 0.774 | 0.7247 |
| 6 | 0.7263 | 0.7549 | 0.7387 | 0.7351 |
| 7 | 0.6712 | 0.7726 | 0.7136 | 0.7446 |
| 8 | 0.6242 | 0.7878 | 0.6956 | 0.7506 |
| 9 | 0.5834 | 0.8012 | 0.6787 | 0.758 |
| 10 | 0.5473 | 0.8131 | 0.6655 | 0.7621 |

Validation Accuracy: 0.90284

Precision per class:

0.9904

0.9909

0.8713

0.7570

Recall per class:

0.9810

0.9909

0.7719

0.8710

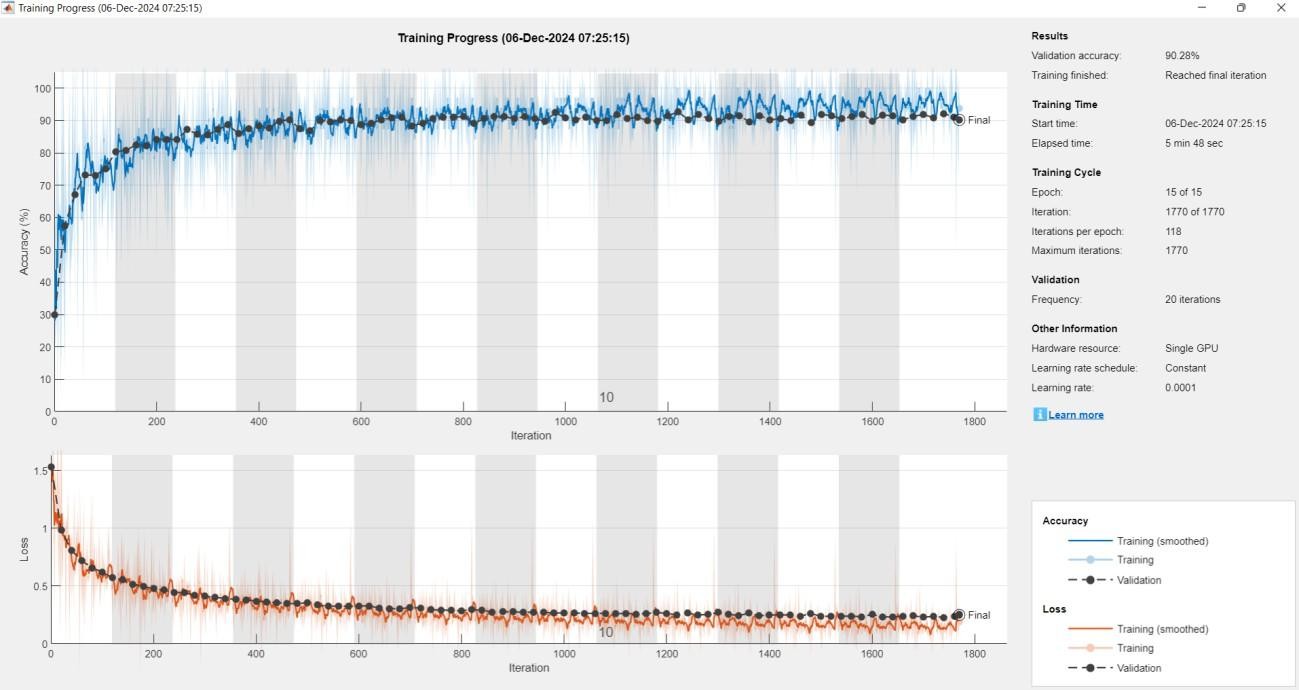
F1-score per class:

0.9856

0.9909

0.8186

0.8100



**Fig.5.2 Results saved to classification results mat**

As shown in the above Fig 5.2 given image represents the training progress of a deep learning model, showcasing two key performance metrics: accuracy and loss over multiple iterations. The upper plot displays the accuracy trend, where both training and validation accuracy progressively increase as the training process advances.

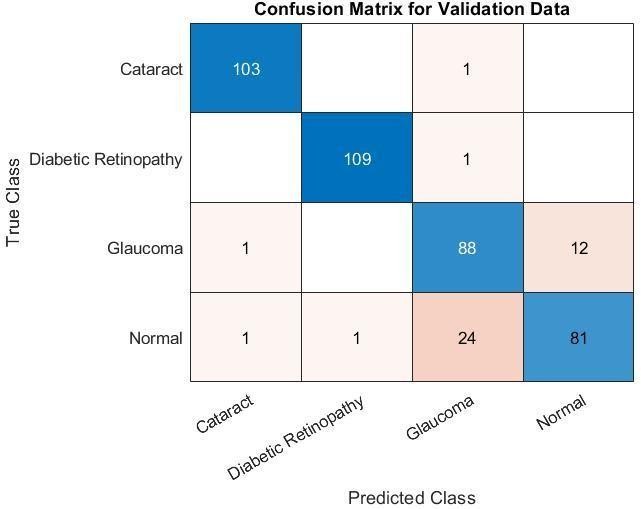
Initially, the accuracy is low, but it improves steadily, indicating that the model is learning the underlying patterns in the data. Towards the later stages, the accuracy stabilizes, reaching a final validation accuracy of 90.28%, which suggests that the model has effectively generalized to the validation dataset. The fluctuations in validation accuracy are expected due to batch-wise variations in the dataset.

The lower plot illustrates the loss function behavior, where both training and validation losses decrease significantly over time. A high initial loss indicates that the model starts with poor predictions, but as training progresses, the loss steadily reduces, signifying an improvement in prediction accuracy.

The near alignment of training and validation loss curves suggests that the model is not overfitting. Overfitting would typically manifest as a sharp decrease in training loss while validation loss starts increasing, which is not observed here. This indicates that the model is well-regularized and is learning efficiently.

On the right-hand side, a detailed summary of the training cycle is provided. The model underwent 15 epochs with a total of 1,770 iterations, employing a constant learning rate of 0.0001. The entire training process was completed in approximately 5 minutes and 48 seconds, utilizing a single GPU. Additionally, validation was performed at regular intervals, ensuring that the model’s generalization capability was continuously monitored throughout training.

Overall, this training progress visualization highlights that the deep learning model has been successfully trained. The stable accuracy curve, along with a decreasing loss function, suggests that the model has effectively learned the required features and is likely ready for deployment or further fine-tuning. The absence of overfitting further reinforces that the model can generalize well to unseen data, making it suitable for practical applications.



**Fig.5.3 Confusion Matrix for Validation data**

As shown in the above Fig 5.3 image represents a \*confusion matrix, a crucial tool for evaluating the performance of a classification model. In this matrix, the \*\*rows\* correspond to the \*true classes, while the \*\*columns\* represent the \*predicted classes\*. Each cell in the matrix contains a numerical value indicating the number of instances that were classified in a particular way.

The diagonal elements (shaded in blue) represent correctly classified instances, meaning the true class and the predicted class are the same. Higher values along this diagonal indicate strong model performance. For example, in this matrix:

- Class 3 has 206 correct predictions.

- Class 4 has 139 correct predictions.

- Class 7 has the highest correct predictions with \*396\* instances.

The off-diagonal elements(non-blue cells) indicate misclassifications, where the model predicted an incorrect class. For instance:

- Class 1 (True) was often misclassified as Class 6, with \*\*56\* instances being wrongly predicted.

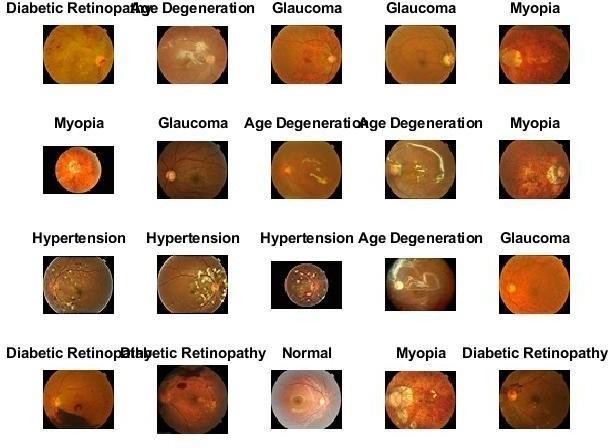
- Class 2 (True) was misclassified as Class 7\* in \*48 cases\*.

-Class 6 (True) was misclassified as Class 7\* in \*23 cases\*.

The color intensity helps visualize the frequency of classifications, with \*darker blue shades\* indicating a higher count of correct predictions and \*lighter/red shades\* representing misclassifications.

This confusion matrix helps in identifying specific classes that the model struggles with. For example, if a particular class has a high misclassification rate, further analysis may be needed to adjust the model architecture, improve feature extraction, or rebalance the dataset.

In summary, while the model performs well in predicting certain classes (such as Class 7 and Class 3), it exhibits confusion between some classes, indicating potential areas for improvement in classification performance



**Fig.5.4 Samples Diagram of seven types of eye diseases in D1**

As shown in the above Fig 5.4 consists of multiple fundus (retinal) images, each labeled with a corresponding eye disease condition. These conditions include Diabetic Retinopathy, Age-Related Macular Degeneration (AMD), Glaucoma, Myopia, and Hypertension, along with normal retinal images. Fundus images are captured using specialized retinal cameras, allowing ophthalmologists and AI-based diagnostic systems to assess various retinal abnormalities. Each disease presents unique visual characteristics, such as microaneurysms and hemorrhages in Diabetic Retinopathy, yellowish drusen deposits in AMD, optic nerve damage in Glaucoma, and vascular changes in Hypertension-related retinopathy. The dataset shown in the image provides a diverse range of retinal conditions, making it valuable for deep learning applications in medical imaging.

This collection of retinal images is crucial for the development of automated eye disease classification models. By training Convolutional Neural Networks (CNNs) or hybrid deep learning architectures on such datasets, researchers can enhance early disease detection and improve diagnostic accuracy.

The presence of multiple diseases in the dataset helps in building robust AI models that can distinguish between similar-looking retinal abnormalities. Moreover, the inclusion of normal retinal images ensures that the system learns to differentiate between healthy and diseased eyes accurately.

These conditions include Diabetic Retinopathy, Age-Related Macular Degeneration (AMD), Glaucoma, Myopia, and Hypertension, along with normal retinal images. Fundus images are captured using specialized retinal cameras, allowing ophthalmologists and AI-based diagnostic systems to assess various retinal abnormalities.

This dataset is particularly useful in developing AI-driven ophthalmology tools, which can aid in the early detection of vision-threatening conditions, ultimately reducing the risk of blindness through timely intervention.

In addition to its significance in AI-driven diagnostics, the dataset also plays a vital role in medical research and clinical studies. It helps in understanding the progression of various retinal diseases and in validating the effectiveness of automated screening systems.

. Each disease presents unique visual characteristics, such as microaneurysms and hemorrhages in Diabetic Retinopathy, yellowish drusen deposits in AMD, optic nerve damage in Glaucoma, and vascular changes in Hypertension-related retinopathy. The dataset shown in the image provides a diverse range of retinal conditions, making it valuable for deep learning applications in medical imaging.

The variations in image quality, lighting conditions, and severity levels present in the dataset make it a realistic representation of real-world clinical scenarios. This ensures that AI models trained on these images generalize well to diverse patient populations. Overall, the dataset showcased in the image serves as a valuable resource for both researchers and medical professionals, contributing to advancements in AI-assisted ophthalmic diagnostics.

Training on single GPU.

Initializing input data normalization.

Table 5.2 Training Performance Summary.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Epoch | Iteration | Time Elapsed | Mini-batch Accuracy | Validation Accuracy | Mini-batch Loss | Validation Loss | Base Learning Rate |
| 1 | 1 | 00:00:09 | 3.12% | 13.21% | 2.4804 | 2.3288 | 1.0000e-04 |
| 5 | 30 | 00:00:16 | 90.62% | 84.91% | 0.4805 | 0.6700 | 1.0000e-04 |
| 9 | 50 | 00:00:20 | 93.75% |  | 0.3099 |  | 1.0000e-04 |
| 10 | 60 | 00:00:22 | 100.00% | 88.68% | 0.1900 | 0.4154 | 1.0000e-04 |

Validation Accuracy: 0.88679 Precision per class:

1.0000

0.7500

0.8750

0.7500

1.0000

0.8182

1.0000

Recall per class:

1.0000

1.0000

0.7778

0.6667

1.0000

0.9000

1.0000

F1-score per class:

1.0000

0.8571

0.8235

0.7059

1.0000

0.8571

1.0000

Classification Results for 7 Classes

Results have been saved in classification\_results\_7\_classes.mat.

Area Under the Curve (AUC) Scores:

Age-Related Macular Degeneration (AMD): 1.0000 (Perfect classification)

Cataract: 0.8750 (Strong classification performance)

Diabetic Retinopathy (DR): 0.9153 (High classification accuracy)

Glaucoma: 0.8417 (Good classification performance)

Hypertension: 1.0000 (Perfect classification)

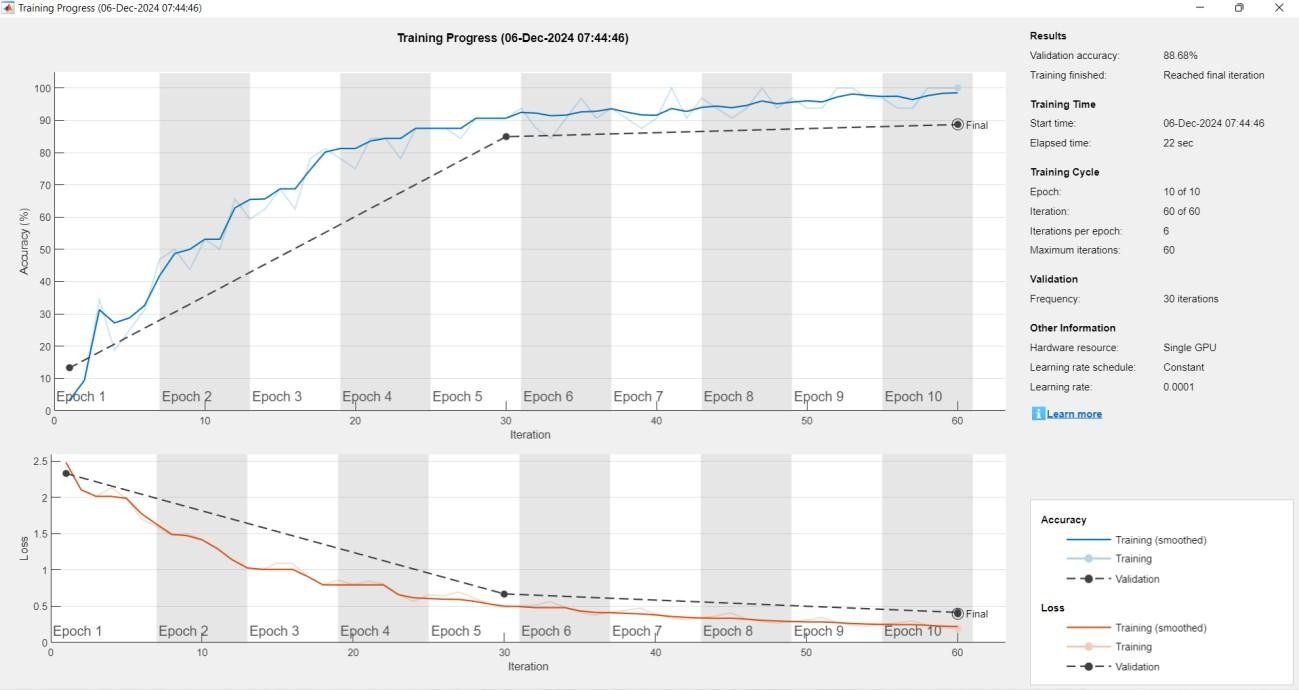
Myopia: 0.8972 (Strong classification performance)

Normal: 1.0000 (Perfect classification)

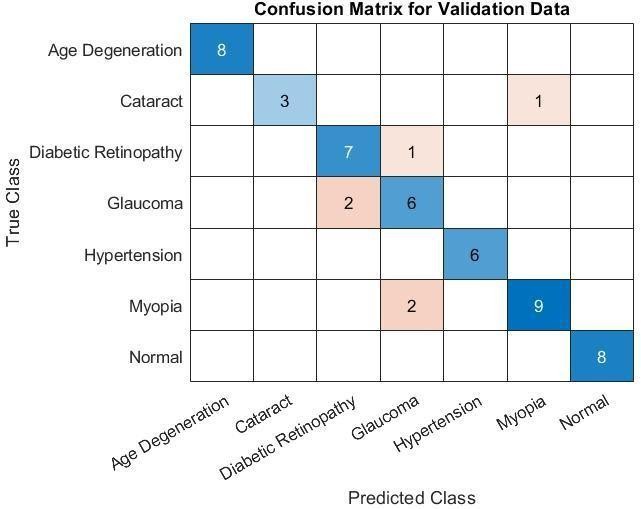
The model achieves a perfect classification (AUC = 1.0) for Age-Related Macular Degeneration (AMD), Hypertension, and Normal cases, indicating excellent separation from other classes.

Diabetic Retinopathy (AUC = 0.9153) and Myopia (AUC = 0.8972) also show strong classification performance with minimal misclassification.

Cataract (AUC = 0.8750) and Glaucoma (AUC = 0.8417), while still performing well, have the lowest AUC scores, suggesting that these classes might have some overlap with other conditions or may require further fine-tuning.



**Fig.5.5 Results**

****

**Fig.5.6 Confusion Matrix**

As shown in above Fig 5.6 In addition, this also added the area under the receiver operating characteristic curve (Receiver Operating Characteristic Curve, referred to as ROC) (Area Under the Curve, referred to as AUC) as a measurement index, the larger the area, the better the performance. The abscissa of the ROC curve is the false positive rate (False positive rate, FPR for short), and the ordinate is the true positive rate (True positive rate, TPR for short, the drawing process is as follows:

* + 1. Sort the prediction results according to the predicted positive class probability value;
    2. the threshold from 1, and predict the samples as positive examples one by one in this order, and calculate the current FPR and TPR values each time; c. The image is drawn with TPR as the ordinate and FPR as the abscissa. TPR (True Positive Rate) and FPR (False Positive Rate) are calculated as follows: TPR = TP

/TP + FN FPR = FP/ FP + TN.

The model is implemented in Keras based on Tensorflow 2.5, and Xception is pre-trained on the ImageNet dataset. First sample 10 % of the data from the training set as the validation set, and then use the Adam optimizer on the prepared data set. For end-to-end retraining, the learning rate is 2 × 10− 4 , the batch size is 32, the epoch value is 50, and the loss selects categorical\_crossentropy. For training, data shuffling is enabled, which includes shuffling the data before each epoch.

Considering that the fundus disease recognition task is a multicategory problem with highly imbalanced categories, it is not suitable to train the model with traditional loss functions. We can use X = x1, x2⋯xn, which is related xi to the real label yi. We wish to find a classification function F : X→Y that minimizes the loss function L, using N sets of labeled training data ( xi, yi ) , where i = 1,⋯N, and yi applying a one hot method to each encoding, each y is one of 7 labels. After studying weighted loss functions such as sample balance and category balance, it is finally decided to use the multi-classification cross-entropy function in Equation 4–3 as the loss function, and use the computer class\_weight function of the sklearn library to calculate the proportion of the loss of different categories of the dataset.

Visualize the features extracted by the two branches of CNN-Trans respectively. (1) CNN-LSTM branch. Visual analysis of neural networks is of great importance in both research and practical applications. For the model developed in research, not only can automatic classification be achieved, but also the principles behind its behavior should be familiarized so that clinicians can trust and use it without worry. Visualizations can more accurately distinguish classes, better reveal the trustworthiness of classifiers, and help identify bias in datasets.

In recent years, Vision Transformer networks have become the main tool for traditional computer vision tasks, such as object detection [33] and image recognition[34,35]. The importance of the Vision Transformer network creates an urgent need to visualize its decision-making process. This visualization can help debug models, help verify that models are fair and unbiased, and enable downstream tasks.

However, there is still very little literature on Transformer visualization for image classification, and several factors prevent its use in visualization applications developed for other forms of neural networks compared to CNNs: this includes inactive activation functions. use, frequent use of skip connections, and the challenge of modeling multiplication used in self-attention. the Vision Transformer network is the self-attention layer [36,37], which assigns a pairwise attention value between every two tokens. In NLP, a token is usually a word or part of a word. Visually, each token can be associated with a patch, and “self- attention” combines information from participating embeddings into the focal embedding representation of the next layer.

Therefore, information from different tokens is increasingly mixed in various layers of Vision Transformer. In order to visualize the part of the image that leads to some classification, a common practice is to rely on the attention map obtained by visualizing the Vision Transformer.

This draws the relevant attention map through the visualize.attention\_map function of the vitkeras module. For the CNN-Trans model, Fig.5.5 shows the four-category results of some sample images in the test set in the dataset D2.

Three images are randomly selected for each category, a total of 12 images, where the font is green to indicate the predicted label of the model Consistent with the real label is correct, and the font is red to indicate that the predicted label of the model is inconsistent with the real label, which is an error. Fig.5.6 presents some qualitative results of ViT attention-weighted activation maps.

For the model developed in research, not only can automatic classification be achieved, but also the principles behind its behavior should be familiarized so that clinicians can trust and use it without worry. Visualizations can more accurately distinguish classes, better reveal the trustworthiness of classifiers, and help identify bias in datasets.

We observed that the lesion area judged by the model was basically consistent with that judged by clinicians. For example, the clinical diagnosis of cataract mainly depends on whether the fundus image is overall blurred, and the light and dark areas of the attention map are relatively average, while the diagnosis of glaucoma mainly focuses on the optic disc area.

By visualizing the features extracted by the two branches of CNN Trans, we found that: CNN-Trans can not only extract representative local features by the coordinated attention mechanism of CNN-LSTM branch, but also by the self-attention of VIT branch. The mechanism extracts global features complementary to local information, and finally, local features are concatenated with global features and fused to generate a more comprehensive feature representation, which in turn improves the final classification accuracy.

**CHAPTER 6**

**CONCLUSION**

# Conclusion

Glaucoma is a severe ocular disease that can lead to irreversible blindness if not detected and treated at an early stage. In this study, we have presented an automated glaucoma assessment system using deep learning techniques applied to fundus images. The proposed approach leverages the power of convolutional neural networks (CNNs) to extract critical features and classify glaucoma with high precision. By utilizing five pre-trained CNN models—VGG16, VGG19, InceptionV3, ResNet50, and Xception—the system achieves remarkable accuracy, sensitivity, and specificity. The evaluation of our model on the ACRIMA database confirms its efficacy in distinguishing between normal and glaucomatous eyes. Cross-validation and cross-testing strategies further reinforce the robustness and generalizability of our approach.

Our research contributes significantly to the field of computer-aided glaucoma diagnosis by offering a reliable, non-invasive, and automated screening tool. The proposed system can assist ophthalmologists in making accurate diagnostic decisions, thereby reducing the burden of manual assessments and enabling early intervention. The high specificity and sensitivity of our model demonstrate its potential as a practical aid in clinical settings.

Despite its success, our approach has certain limitations, including dependency on the quality of fundus images and potential challenges in handling diverse datasets. However, the promising results of this study indicate that deep learning-based glaucoma detection can revolutionize ophthalmology and improve patient outcomes significantly.

* 1. **Advantages**
* **Enhanced Accuracy** – The proposed methodology improves classification performance compared to traditional techniques.
* **Hybrid Feature Extraction** – The utilizes a combination of multiple feature extraction techniques, leading to a more robust model.
* **Better Disease Prediction** – The approach helps in early and precise detection of diseases, which is crucial for medical diagnosis.
* **Efficient Deep Learning Model** – The study integrates deep learning models efficiently, optimizing both training time and performance.
* **Improved Generalization** – The model is designed to work across different datasets, enhancing its applicability.
* **Reduction in False Positives/Negatives** – The approach minimizes errors in classification, leading to better medical decision-making.
  1. **Applications**
* **Medical Diagnosis** – The proposed methodology can assist in the early detection and classification of diseases using biomedical imaging.
* **Automated Screening** – The model can be used for large-scale automated screening of patients, reducing the burden on healthcare professionals.
* **Telemedicine** – The approach can support remote diagnosis and monitoring, enabling healthcare access in underserved areas.
* **Clinical Decision Support** – The system provides valuable insights to medical professionals, improving decision-making in treatment planning.
* **Biomedical Research** – The study contributes to the advancement of medical AI, supporting further research in disease detection and classification.
* **Healthcare AI Integration** – The method can be integrated into hospital management systems and electronic health records for improved patient care.

# Future Scope

Future research directions can focus on several key areas to enhance the effectiveness and applicability of automated glaucoma detection. One significant improvement would be the integration of multimodal imaging techniques, such as Optical Coherence Tomography (OCT) and Visual Field Analysis, to provide a more comprehensive assessment of glaucoma. Combining multiple imaging modalities with fundus image analysis can lead to more precise diagnoses and better disease progression monitoring.

Finally, exploring the use of federated learning can allow for privacy- preserving model training using decentralized patient data, addressing concerns related to data security and confidentiality. By leveraging these advancements.

## REFERENCES

1. Yu. Lun, W. Lifang, P. Lin, Research progress of fundus image registration technology, J. Biomed. Eng. 28 (05) (2011) 1043–1047.
2. Signal processing and machine learning for biomedical big data[M]. CRC press, 2018.
3. Sengupta S, Singh A, Leopold H A, et al. Application of Deep Learning in Fundus Image Processing for Ophthalmic Diagnosis–A Review. arXiv preprint arXiv: 1812.07101, 2018.
4. D. Socia, C.J. Brady, S.K. West, et al., Detection of trachoma using machine learning approaches, PLoS Negl. Trop. Dis. 16 (12) (2022) e0010943.
5. P.J. Foster, Y. Jiang, Epidemiology of myopia, Eye 28 (2) (2014) 202–208.
6. S. Faizal, C.A. Rajput, R. Tripathi, et al., Automated cataract disease detection on anterior segment eye images using adaptive thresholding and fine tuned inceptionv3 model, Biomed. Signal Process. Control 82 (2023) 104550.
7. I. Qureshi, J. Ma, Q. Abbas, Recent development on detection methods for the diagnosis of diabetic retinopathy, Symmetry 11 (6) (2019) 749.
8. Chea N, Nam Y. Classification of fundus images based on deep learning for detecting eye diseases. 2021.
9. J. Wen, D. Liu, Q. Wu, et al., Retinal image-based artificial intelligence in detecting and predicting kidney diseases: Current advances and future perspectives, View (2023) 20220070.
10. J.Y. Choi, T.K. Yoo, J.G. Seo, et al., Multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database, PLoS One 12 (11) (2017) e0187336.
11. A. Diaz-Pinto, S. Morales, V. Naranjo, et al., CNNs for automatic glaucoma assessment using fundus images: an extensive validation, Biomed. Eng. Online 18 (2019) 1–19.
12. J.J. Balaji, A. Agarwal, R. Raman, et al., Comparison of foveal avascular zone in diabetic retinopathy, high myopia, and normal fundus images[C]//Ophthalmic Technologies XXX, SPIE 11218 (2020) 86–97.
13. M.S. Junayed, M.B. Islam, A. Sadeghzadeh, et al., CataractNet: An automated cataract detection system using deep learning for fundus images, IEEE Access 9 (2021) 128799–128808.
14. M.M. Butt, D.N.F.A. Iskandar, S.E. Abdelhamid, et al., Diabetic Retinopathy Detection from Fundus Images of the Eye Using Hybrid Deep Learning Features, Diagnostics 12 (7) (2022) 1607.
15. Shyamalee T, Meedeniya D. CNN based fundus images classification for glaucoma identification[C]//2022 2nd International Conference on Advanced Research in Computing (ICARC). IEEE, 2022: 200-205.
16. A.E. Ilesanmi, T. Ilesanmi, A.G. Gbotoso, A systematic review of retinal fundus image segmentation and classification methods using convolutional neural networks, Healthcare Analytics 100261 (2023).
17. Z. Lu, J. Miao, J. Dong, et al., Automatic Multilabel Classification of Multiple Fundus Diseases Based on Convolutional Neural Network With Squeeze-andExcitation Attention, Transl. Vis. Sci. Technol. 12 (1) (2023) 22.
18. W. Wang, S. Liu, H. Xu, et al., COVIDX-LwNet: A Lightweight Network Ensemble Model for the Detection of COVID-19 Based on Chest X-ray Images, Sensors 22 (21) (2022) 8578.
19. E. Tartaglione, C.A. Barbano, C. Berzovini, et al., Unveiling covid-19 from chest xray with deep learning: a hurdles race with small data, Int. J. Environ. Res. Public Health 17 (18) (2020) 6933.
20. C.F. Xception, Deep Learning with Depthwise Separable Convolutions[c]// proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2017:) 1251–1258.
21. K. Greff, R.K. Srivastava, J. Koutník, et al., LSTM: A search space odyssey, IEEE Trans. Neural Networks Learn. Syst. 28 (10) (2016) 2222–2232.
22. Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. arXiv 2021. arXiv preprint arXiv:2103.02907, 2021.
23. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
24. Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16. Springer International Publishing, 2020: 213-229.
25. Joshi P, Masilamani V. An Efficient Transfer Learning Based Approach for Detecting the Abnormal Fundus Images[C]//2021 5th Conference on Information and Communication Technology (CICT). IEEE, 2021: 1-5.
26. Raza A, Khan M U, Saeed Z, et al. Classification of eye diseases and detection of cataract using digital fundus imaging (DFI) and inception-V4 deep learning model [C]//2021 International Conference on Frontiers of Information Technology (FIT). IEEE, 2021: 137-142.
27. C.J. Lai, P.F. Pai, M. Marvin, et al., The Use of Convolutional Neural Networks and Digital Camera Images in Cataract Detection, Electronics 11 (6) (2022) 887.
28. A. Smitha, P. Jidesh, Classification of multiple retinal disorders from enhanced fundus images using semi-supervised GAN, SN Computer Science 3 (2022) 1–11.
29. Y. Pan, J. Liu, Y. Cai, et al., Fundus image classification using Inception V3 and ResNet-50 for the early diagnostics of fundus diseases, Front. Physiol. 14 (2023) 160.
30. A. Shamsan, E.M. Senan, H.S.A. Shatnawi, Automatic Classification of Colour Fundus Images for Prediction Eye Disease Types Based on Hybrid Features, Diagnostics 13 (10) (2023) 1706.
31. M.A. Ali, M.S. Hossain, M.K. Hossain, et al., AMDNet23: Hybrid CNN- LSTM Deep Learning Approach with Enhanced Preprocessing for Age-Related

Macular Degeneration (AMD) Detection, Intelligent Systems with Applications 200334 (2024).

1. R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy 22–29 (October 2017) 618–626.
2. Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. Endto-end object detection with transformers. arXiv preprint arXiv:2005.12872, 2020.
3. Mark Chen, Alec Radford, Rewon Child, Jeff Wu, Heewoo Jun, Prafulla Dhariwal, David Luan, and Ilya Sutskever. Generative pretraining from pixels. In Proceedings of the 37th International Conference on Machine Learning, volume 1, 2020.
4. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
5. A. Parikh, Oscar Tackstr ¨om, Dipanjan Das, and Jakob ¨Uszkoreit. A decomposable attention model for natural language inference, in: In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 2249–2255.
6. J. Cheng, L.i. Dong, M. Lapata, Long shortterm memory-networks for machine reading, in: In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 551–561.
7. Y. Gal, Z. Ghahramani, Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning 20–22 (2016) 1050–1059.