Automated Diagnosis of Diabetic Eye Disorders Using Advanced Deep Learning Techniques

A dissertation submitted to the

Department of Computer Science and Informatics,

Faculty of Applied Sciences,

Uva Wellassa University, Sri Lanka

in partial fulfillment of the requirements for the award of the Degree of

Bachelor of Science (Honours) in Industrial Information Technology

By

EDIRISINGHA GAMA ARACHCHILAGE KAVINDU AKALANKA RANAWEERA

MANIMELDURA SENURA DAKSITHA MENDIS

RATHNAWEERA PATABENDIGE RAVINDU MALSHAN

Department of Computer Science and Informatics

Faculty of Applied Sciences

Uva Wellassa University, Sri Lanka

September 2025

Declaration

We do hereby declare that the work reported in this dissertation titled *"Automated Diagnosis of Diabetic Eye Disorders Using Advanced Deep Learning Techniques"* was exclusively carried out by us under the supervision of Mr. Aruna Sanjeewa and co-supervision of Ms. J. A. V. M. K. Jayakody.

It describes the results of our own independent research except where due reference has been made in the text. No part of this dissertation has been submitted earlier or concurrently for the same or any other degree.

Date: 2025.10.03 ……

E.G.A. Ranaweera

(iit20082@std.uwu.ac.lk)

Date: 2025.10.03 ……

M.S.D. Mendis

(iit20075@std.uwu.ac.lk)

Date: 2025.10.03 ……

R.P.R. Malshan

(iit20047@std.uwu.ac.lk)

We endorse the declaration by the candidates.

……	……
Signature of the Supervisor	Signature of the Co-supervisor
Mr. Aruna Sanjeeewa	Ms. J.A.V.M.K.Jayakody
Name of the Supervisor	Name of the Co-supervisor
Date: 2025.10.03 ……	Date: 2025.10.03 ……

Acknowledgement

We would like to extend our sincere gratitude to our supervisor, Mr. Aruna Sanjeewa, for his invaluable guidance, encouragement, and constructive feedback throughout the course of this research. His expertise and support were instrumental in the successful completion of this study.

We are also grateful to our co-supervisor, Ms. J. A. V. M. K. Jayakody, for her continuous guidance, technical advice, and valuable insights, which contributed significantly to the quality of this work.

Our appreciation is further extended to the Department of Computer Science and Informatics, Faculty of Applied Sciences, Uva Wellassa University, for providing the necessary facilities and academic environment required for the successful execution of this research project.

Content

Abstract

The global rise in diabetes has led to a significant increase in vision-threatening disorders such as Diabetic Retinopathy (DR) and Diabetic Macular Edema (DME), where early detection is crucial to prevent irreversible blindness. To address the limitations of time-consuming and subjective manual diagnosis, this research presents an automated system leveraging advanced deep learning techniques: a U-Net-inspired classifier for five-stage DR grading from fundus images and a YOLOv8 detector for DME from OCT B-scans. Using preprocessed datasets from public Roboflow repositories, the system demonstrated reliable performance, confirmed by evaluation metrics including accuracy for DR and precision/recall/mAP for DME. Integrated into a web-based platform for real-time predictions, this solution offers an efficient, scalable, and accessible tool for large-scale screening, with the potential to reduce diagnostic delays, support clinical decision-making, and prevent diabetes-induced blindness, particularly in resource-constrained environments.

List of Figures

List of Tables

List of Abbreviations

- AI – Artificial Intelligence
- API – Application Programming Interface
- CNN – Convolutional Neural Network
- COCO – Common Objects in Context
- DL – Deep Learning
- DME – Diabetic Macular Edema
- DR – Diabetic Retinopathy
- FPN – Feature Pyramid Network (YOLOv8 uses it)
- GPU – Graphics Processing Unit
- IDF – International Diabetes Federation (used in Introduction/Problem section)
- OCT – Optical Coherence Tomography
- PAN – Path Aggregation Network (YOLOv8 uses it)
- SGD – Stochastic Gradient Descent
- VGG – Visual Geometry Group (only if you mention VGG in Literature Review)
- YOLO – You Only Look Once

Chapter 1

Introduction

Background and Rationale

Diabetes has become one of the most prevalent chronic illnesses worldwide, with rapidly increasing cases across all regions. Diabetes contributes significantly to vision-related disorders, particularly Diabetic Retinopathy (DR) and Diabetic Macular Edema (DME). Both conditions have emerged as leading causes of preventable blindness among working-age adults. The International Diabetes Federation (IDF) and other global health organizations highlight the urgent need for early detection and management, as vision loss caused by diabetic eye diseases is often irreversible once the damage has progressed.

Currently, the diagnosis of DR and DME relies primarily on the manual interpretation of retinal fundus images and Optical Coherence Tomography (OCT) scans by ophthalmologists. While this method has been the gold standard for decades, it is highly resource-intensive, time-consuming, and dependent on the availability of skilled specialists. With the global rise in diabetic patients, manual diagnostic methods alone are no longer sufficient to meet the increasing demand for early screening and effective management.

Recent advances in Artificial Intelligence (AI) and Deep Learning (DL) have revolutionized image analysis in the medical domain. Convolutional Neural Networks (CNNs), object detection algorithms such as You Only Look Once (YOLO), and image segmentation models like U-Net have shown exceptional performance in identifying subtle abnormalities in medical images. Leveraging these techniques for diabetic eye disease detection can significantly enhance accuracy, speed, and accessibility, while also reducing the workload on medical professionals.

This research is motivated by the pressing need to develop an intelligent, automated diagnostic system that supports ophthalmologists in early identification of DR and

DME. By integrating advanced deep learning techniques with a user-friendly web platform, the system aims to provide real-time predictions, enabling scalable screening solutions in both urban hospitals and resource-constrained rural environments.

Problem Identification

The increasing prevalence of diabetes has placed a severe burden on global healthcare systems, with diabetic eye diseases becoming one of the fastest-growing causes of vision impairment. Despite being preventable in many cases through early detection, large numbers of patients still suffer from delayed diagnoses due to limitations in the current diagnostic process.

At present, ophthalmologists are required to carefully examine retinal fundus images and OCT scans for signs of micro aneurysms, fluid pockets, and structural abnormalities that indicate DR and DME. This manual process is prone to variability, subjective interpretation, and human error. Moreover, the growing demand for screening outpaces the availability of qualified specialists, particularly in developing countries where access to advanced diagnostic facilities is limited.

Traditional machine learning techniques have been applied in past research to detect diabetic eye diseases, but their accuracy and reliability are not sufficient for clinical deployment. Many of these models struggle with the complexity of retinal and OCT images, where features such as micro aneurysms, exudates, and macular thickening can be subtle and difficult to distinguish. Furthermore, the absence of scalable automated systems results in delays in diagnosis, increasing the risk of severe complications including irreversible blindness.

Thus, there is a critical need for an automated, reliable, and scalable diagnostic system that can detect and classify Diabetic Retinopathy (DR) and Diabetic Macular Edema (DME) with high accuracy, while reducing dependency on manual interpretation.

Aim and Objectives

Aim

The aim of this research is to develop an automated deep learning–based system for accurate detection and classification of Diabetic Retinopathy (DR) and Diabetic Macular Edema (DME) using retinal and OCT images.

Objectives

The specific objectives of this research are:

1. To design an AI-powered system utilizing deep learning techniques for the accurate detection and classification of DR and DME from retinal and OCT images.
2. To enhance the performance of the diagnostic system through parameter tuning and optimization methods, ensuring reliable identification across diverse datasets.
3. To implement a web-based interface enabling real-time predictions, supporting mass screenings and improving diagnostic efficiency in resource-constrained environments.

Expected Outcomes

The research is expected to achieve the following outcomes:

- Development of a deep learning model (based on YOLO and U-Net architectures) capable of accurately detecting and classifying DR and DME using retinal and OCT images.
- A web-based platform that integrates the trained models, allowing healthcare professionals to upload images and receive real-time diagnostic results.
- A scalable, efficient, and accessible solution that supports ophthalmologists in clinical decision-making, reduces diagnostic delays, and contributes to preventing vision loss caused by diabetic eye diseases.

Chapter 2

Literature Review

2.1. Theoretical Background

I. Diabetes Mellitus and Ocular Complications

Diabetes mellitus (DM) is a global epidemic, with more than 500 million individuals affected worldwide (International Diabetes Federation, 2021). One of the most serious consequences of diabetes is its impact on ocular health, particularly the retina. Chronic hyperglycemia leads to microvascular and neuronal damage, manifesting in diabetic retinopathy (DR), diabetic macular edema (DME), glaucoma, and cataracts. DR, in particular, is one of the leading causes of blindness among working-age adults. It progresses from mild non-proliferative changes to severe proliferative stages that threaten vision if untreated.

II. 1.2 Traditional Diagnostic Pathways

The gold standard for diagnosing diabetic eye disease involves fundus photography and optical coherence tomography (OCT). These imaging modalities allow clinicians to detect microaneurysms, hemorrhages, neovascularization, and macular thickening. Diagnosis, however, requires expert graders, which is labor-intensive and prone to inter-observer variability. Moreover, in low- and middle-income countries (LMICs), there is often a shortage of ophthalmologists, making universal screening challenging.

III. 1.3 The Role of Artificial Intelligence in Ophthalmology

Artificial intelligence (AI), particularly deep learning (DL), offers a transformative solution to these challenges. Unlike conventional machine learning that relies on manually engineered features, DL architectures—especially convolutional neural networks (CNNs)—learn hierarchical feature representations directly from raw image pixels. This makes them ideal for analyzing complex retinal structures.

Recent advancements go beyond CNNs to include:

- Vision Transformers (ViTs): Models that capture long-range dependencies in images.
- Attention mechanisms: Allowing models to focus on clinically relevant retinal lesions.
- Generative Adversarial Networks (GANs): For generating synthetic medical images to balance datasets.
- Self-supervised learning (SSL): Leveraging large amounts of unlabeled data to reduce reliance on expert annotations.

Together, these methods create the theoretical foundation for automated, scalable, and potentially autonomous diagnosis of diabetic eye diseases.

2.2. Previous Work

I. Early Breakthroughs in DR Detection Using CNNs

The application of deep learning in ophthalmology gained momentum with (Gulshan & al., 2016), who trained a CNN on more than 120,000 fundus images. Their model demonstrated sensitivity and specificity on par with board-certified ophthalmologists, establishing DL as a credible diagnostic tool.

Building on this, (Abràmoff & al, 2020), conducted the first pivotal trial of an autonomous AI system—IDx-DR—in primary care offices. Their system achieved FDA approval, marking a historic moment for AI in clinical practice (*NPJ Digital Medicine*). These studies provided the proof-of-concept that AI could safely and effectively detect DR in real-world healthcare environments.

II. Expansion to Multi-Disease and Multimodal Frameworks

While initial work focused solely on DR, later studies expanded scope:

- (Lam & al., 2018) designed models capable of detecting not just DR but also age-related macular degeneration (AMD) and glaucoma, aligning AI with the practical needs of community screening (*Ophthalmology*).

- (De Fauw & al., 2018) extended DL applications to OCT scans, training a two-stage system for disease segmentation and referral decisions, achieving near-specialist performance (*Nature Medicine*).

This shift toward multi-disease frameworks reflects the real-world necessity of screening beyond a single disorder.

III. Innovations in Model Architectures

Advances in network design addressed critical issues of robustness and interpretability:

- (Li & al., 2019) introduced attention-based CNNs, where the model learns to focus on lesions such as microaneurysms and hemorrhages, aligning with clinical reasoning (*IEEE Transactions on Medical Imaging*).

- (Liu & al., 2022) pioneered hybrid CNN–Vision Transformer models, which captured both local and global features, outperforming CNNs in generalization across heterogeneous datasets (*Medical Image Analysis*).

These studies demonstrated that architectural choices matter and can significantly improve AI's clinical reliability.

IV. Tackling Data Scarcity and Imbalance

Medical imaging datasets are often imbalanced, with fewer samples of early DR stages. To overcome this:

- (Zhou & al., 2021)applied GAN-based augmentation, generating synthetic fundus images to increase sensitivity for underrepresented classes (*Computers in Biology and Medicine*).

- (Azizi & al., 2023) leveraged self-supervised learning (SSL), pretraining models on large unlabeled datasets and fine-tuning on smaller labeled ones. This drastically reduced reliance on costly expert annotations (*Nature Biomedical Engineering*).

Together, these approaches address the data bottleneck that limits AI development in ophthalmology.

V. Clinical, Ethical, and Regulatory Considerations

Although technical performance is high, clinical integration is complex:

- (Ting & al., 2019) highlighted the risks of performance variability across diverse populations, emphasizing the need for multi-ethnic validation (*Progress in Retinal and Eye Research*).
- (Abràmoff & al, 2020) discussed trust, fairness, and responsibility, noting that questions around liability, patient consent, and algorithmic bias remain unresolved (*Nature Medicine*).

Thus, clinical adoption requires not only accurate algorithms but also robust governance and ethical oversight.

2.3. Gaps in Knowledge

Despite these advancements, literature reveals persistent gaps:

1. Generalizability

Most models are trained on datasets from limited regions (e.g., EyePACS, Messidor). Validation across diverse populations is insufficient, raising concerns about equity in diagnosis.

2. Explainability

Although attention mechanisms improve interpretability, AI remains largely a “black box.” Clinicians often require transparent, interpretable outputs to build trust.

3. Integration of Multi-Disease Screening

While some models attempt multi-disease detection, comprehensive AI systems for DR, AMD, glaucoma, and other pathologies simultaneously are rare.

4. Data Limitations

Annotated datasets remain scarce, especially for rare conditions. GANs and SSL are promising but need rigorous clinical validation before widespread use.

5. Deployment in Resource-Limited Environments

Lightweight, portable AI systems are underdeveloped, despite being critical for regions with high DR prevalence and limited medical infrastructure.

6. Ethical, Legal, and Regulatory Challenges

Questions about accountability in autonomous AI remain. For example, if an AI misdiagnoses a patient in a rural clinic, who bears responsibility - the developer, the clinician, or the health system?

Despite the promising results from prior studies, several gaps remain,

Disease-Specific Optimization: Many existing systems are either designed solely for DR (fundus photographs) or DME (OCT scans). Few studies have proposed separate yet integrated modules optimized for each disease, despite their distinct imaging requirements.

Combination of Object Detection and Segmentation: Prior work often relies exclusively on classification models (CNNs) without leveraging advanced architectures like YOLO for detection or U-Net for segmentation, which are better suited to the unique imaging modalities of DR and DME.

Practical Deployment: While research prototypes report strong performance, there is limited work on deploying these models into web-based diagnostic platforms that provide real-time results in clinical or resource-limited settings.

The last decade has seen tremendous progress in AI-powered diabetic eye disorder diagnosis, moving from proof-of-concept CNNs to hybrid architectures and self-supervised frameworks capable of near-specialist performance. Clinical trials, including FDA-approved autonomous systems, demonstrate that AI can integrate into healthcare workflows.

Yet, to achieve global impact, research must address generalizability, interpretability, multi-disease integration, and ethical frameworks. Bridging these gaps will ensure AI systems are not only accurate but also equitable, trustworthy, and deployable in diverse healthcare contexts.

Chapter 3

Methodology

i. Research Design and System Overview

The methodological framework for this research was structured around the development of two independent deep learning pipelines, each specifically engineered to address the unique diagnostic requirements of Diabetic Retinopathy (DR) and Diabetic Macular Edema (DME). The system was designed to process fundus photographs for DR severity classification and OCT scans for DME detection, acknowledging the distinct clinical information provided by each imaging modality. These pipelines were subsequently integrated into a unified web application to create an end-to-end diagnostic tool. The workflow initiates with image upload through a React-based frontend, which routes the image to a specific Flask API endpoint based on the disease type. The backend server then executes the corresponding pre-trained deep learning model—a custom convolutional neural network for DR grading or a YOLOv8 architecture for DME object detection. The resulting prediction, including class labels and confidence scores, is returned to the frontend for display to the user. This design ensures a seamless, automated process from data input to clinical output, fulfilling the core objective of creating an accessible decision-support system.

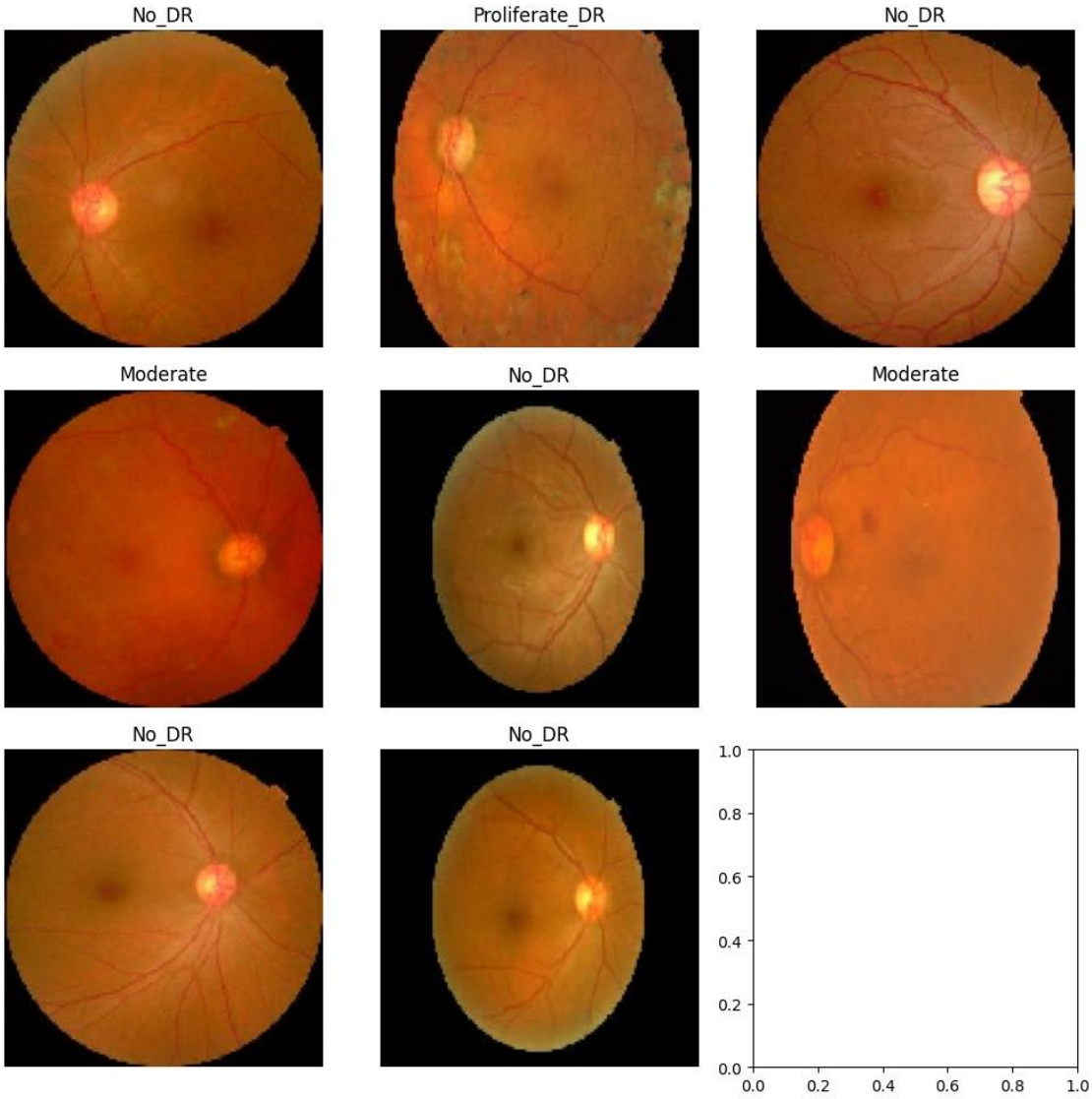

Figure 1:DR classification and labelling

ii. Data Sources and Acquisition

The research utilized publicly available, expertly annotated datasets sourced from the Roboflow platform to ensure reliability and reproducibility. Two distinct datasets were employed, each tailored to a specific disease model, as detailed in Table 1 below.

Parameter	DR Classification Dataset	DME Detection Dataset
Imaging Modality	Fundus Photography	Optical Coherence Tomography (OCT)
Total Images	3,662	1,200
Annotation Type	Multi-class Classification Labels	Bounding Box Coordinates
Classes/Labels	0: Mild, 1: Moderate, 2: No_DR, 3: Proliferate_DR, 4: Severe	0: diabetic-macular-edema 1: normal-macular
Training Set	70%	70%
Validation Set	15%	15%
Test Set	15%	15%

Table 1- Dataset characteristics and partitioning for DR and DME models

The DR dataset was annotated according to the International Clinical Diabetic Retinopathy Disease Scale, providing a granular five-point classification critical for staging disease severity. The DME dataset contained precise bounding box annotations identifying fluid-filled regions and cystoid spaces within OCT B-scans. A stratified sampling strategy was rigorously applied during partitioning to maintain the original distribution of classes across all subsets (training, validation, and testing), thereby mitigating bias and ensuring the models were evaluated on a representative sample of the data.

iii. Data Preprocessing

A standardized preprocessing pipeline was implemented for both datasets to ensure optimal model performance, enhance generalizability, and facilitate efficient training. The specific techniques applied were tailored to the characteristics of each imaging modality and the requirements of their respective deep learning models.

For the DR classification dataset comprising fundus images, all photographs were first resized to a uniform dimension of 128x128 pixels. This resolution was selected to preserve critical pathological features while maintaining computational efficiency during the training of the convolutional neural network. Each image was then normalized by rescaling pixel intensity values from a range of 0-255 to a floating-point range of 0-1 by dividing each pixel value by 255. This normalization step accelerates convergence during gradient descent by ensuring consistent feature scaling across the dataset. The preprocessing pipeline for the DR dataset was implemented using the Image Data Generator class from Keras, which performed on-the-fly rescaling during model training.

The DME detection dataset consisting of OCT scans underwent a different preprocessing regimen appropriate for object detection tasks. All OCT images were resized to 640x640 pixels, the default input size for the YOLOv8 architecture, which maintains the aspect ratio while padding the images as necessary. This standardization ensures compatibility with the pre-trained convolutional backbone of the YOLO model. Unlike the DR images, the OCT scans did not undergo intensity normalization through rescaling, as the YOLOv8 implementation handles internal normalization through its own preprocessing modules. The bounding box annotations for the DME dataset were automatically adjusted during the resizing process to maintain their correct positional relationships within the transformed images.

Data augmentation techniques were strategically employed for both datasets to increase diversity and improve model robustness. These transformations included random horizontal flipping, slight rotational variations (up to 15 degrees), and brightness adjustments within a limited range to simulate clinical imaging variations while preserving pathological authenticity. All augmentation was implemented dynamically during training to effectively expand the dataset without requiring permanent storage of modified images.

iv. Model Architectures and Development

The research implemented two distinct deep learning architectures, each optimized for its specific diagnostic task and imaging modality.

The DR classification model was built upon a U-Net inspired architectural framework, innovatively repurposed from its original segmentation purpose to a multi-class classification task. This design leveraged the powerful feature extraction capabilities of the U-Net encoder while implementing a custom classification head for nuanced severity grading.

The encoder component served as a robust feature extractor, consisting of four sequential contracting blocks. Each block began with two 3×3 convolutional layers with same padding to preserve spatial information, each followed by batch normalization to stabilize and accelerate the training process. A Rectified Linear Unit (ReLU) activation function was applied after each batch norm layer to introduce non-linearity, enabling the model to learn complex patterns. Each block concluded with a 2×2 max pooling operation, which reduced the spatial dimensions by half while retaining the most salient features. This contracting pathway progressively transformed the 128x128x3 input fundus image into increasingly abstract and high-dimensional feature representations, capturing everything from low-level edges and textures to high-level pathological structures like microaneurysms and exudates.

The bottleneck layer, the deepest part of the network, operated on a compressed feature map with 512 channels. This layer acted as the model's computational core,

synthesizing all extracted information into a rich, abstract representation of the retinal pathology present in the input image.

Diverging from the standard U-Net architecture, the typical expansive decoder path used for segmentation was replaced by a classification-specific module. The feature maps from the bottleneck were fed into a global average pooling layer, which dramatically reduced spatial dimensions by averaging each feature map into a single value. This operation produced a fixed-length 512-element vector, effectively summarizing the presence of each high-level feature in the image. This vector was then passed through a cascade of two fully connected (dense) layers. The first dense layer consisted of 256 units with ReLU activation, acting as a high-level classifier that learned non-linear combinations of the extracted features. A dropout layer was incorporated after this layer with a rate of 0.5 to prevent overfitting by randomly disabling neurons during training, thereby forcing the network to learn more robust and generalized features. The final classification layer consisted of 5 units, corresponding to the five DR severity grades, and utilized a softmax activation function. This function converted the layer's outputs into a probability distribution, assigning a confidence score to each potential class and enabling nuanced, probabilistic severity grading far beyond simple binary detection.

The DME detection implementation employed the YOLOv8s (small) architecture, a state-of-the-art object detection framework based on a modified CSPDarknet backbone. This single-stage detector processes entire OCT images through a series of convolutional, batch normalization, and SiLU activation layers to generate feature maps at multiple scales. The model's neck component incorporates feature pyramid networks (FPN) and path aggregation networks (PAN) to effectively combine both low-level spatial information and high-level semantic features.

Critically, the model was trained for a binary classification task to differentiate between OCT scans exhibiting pathological signs of Diabetic Macular Edema and those representing normal macular anatomy. The two classes, as defined in the dataset configuration, were: diabetic-macular-edema for scans containing pathological features such as intraretinal fluid pockets, and normal-macular for scans without such abnormalities. This objective required the model to not only localize pathological

features but also develop a comprehensive understanding of normal retinal anatomy to effectively distinguish between the two states.

The detection head utilizes anchor-free prediction, directly estimating object centers and sizes rather than relying on pre-defined anchor boxes, which simplifies the detection process while maintaining high accuracy. The model was initialized with weights pre-trained on the COCO (Common Objects in Context) dataset, leveraging transfer learning to bootstrap general feature extraction capabilities before fine-tuning on the specialized medical imaging dataset. This approach allowed the model to rapidly adapt its detection capabilities to the specific requirements of identifying ophthalmic pathologies within OCT scans.

Parameter	DR Classification Model	DME Detection Model
Base Architecture	U-Net Inspired Custom CNN	YOLOv8s
Input Dimensions	128×128×3	640×640×3
Feature Extraction	Encoder-Decoder with Skip Connections	CSPDarknet Backbone
Output	5-class Softmax Classification	Bounding Box Coordinates
Pre-trained Weights	Random initialization	COCO Dataset Pre-training
Primary Task	Multi-class Classification	Object Detection

Table 2- Architectural and training parameters of DR and DME models

The architectural divergence between models reflects their specialized purposes: the DR model emphasizes hierarchical feature extraction for nuanced classification, while the DME model prioritizes multi-scale object localization capabilities for precise pathological feature identification within OCT scans.

v. Training Protocols and Optimization

The training procedures for both models were carefully designed to maximize performance while minimizing over fitting, with specific hyper parameters tuned for each architectural approach.

The DR classification model was trained using a categorical cross-entropy loss function, appropriate for multi-class classification tasks. The DR model was trained for 10 epochs with Adam (lr =1e-3) and categorical cross-entropy; validation performance was monitored each epoch. Early stopping was not used. The training process utilized the Image Data Generator class from Keras for real-time data augmentation and preprocessing, with the training set (70% of data) used for parameter updates and the validation set (15% of data) used for epoch-wise performance monitoring and early stopping to prevent over fitting.

The DME detection model employed the YOLOv8 training protocol with default hyper parameters optimized for object detection tasks. The DME model was trained using YOLOv8 default settings (20 epochs, image size 640); unless specified, optimizer and augmentation parameters followed the framework defaults. The learning rate was managed through a cosine annealing schedule with an initial rate of 0.01 and final rate of 0.1 times the initial value. The training incorporated mosaic augmentation with a probability of 1.0, mixing four training images into a single sample to improve detection robustness, along with standard augmentations including hue, saturation, and value adjustments.

Both models were trained on a dedicated GPU workstation (NVIDIA RTX 3060 with 12GB VRAM) to accelerate the computationally intensive processes. Training progress was monitored through quantitative metrics and qualitative analysis of validation predictions to ensure proper learning dynamics.

vi. Evaluation Metrics and Validation

A comprehensive evaluation framework was implemented to assess model performance using multiple complementary metrics appropriate for each task. For the DR classification model, performance was evaluated using:

1. Accuracy: Overall correctness across all classes
2. Precision: Measure of exactness for each class
3. Recall: Measure of completeness for each class
4. F1-Score: Harmonic mean of precision and recall
5. Confusion Matrix: Detailed breakdown of classification performance across all severity grades.

The DME detection model was evaluated using object detection-specific metrics:

1. Mean Average Precision (mAP@0.5): Primary metric measuring detection accuracy at 0.5 intersection-over-union threshold
2. Precision: Proportion of correct positive predictions among all positive detections
3. Recall: Proportion of actual positives correctly identified
4. F1-Score: Balanced measure of precision and recall

Both models underwent rigorous testing on the held-out test set (15% of total data) that was completely unseen during the training process. This ensured that reported performance metrics reflected true generalization capability rather than memorization of training examples.

vii. Web Application Integration and Deployment

The trained models were deployed into a production environment through a integrated web application architecture designed for clinical usability. The backend system was implemented as a Flask-based RESTful API server running on Python 3.8. The server exposed two dedicated endpoints:

- /dr endpoint for DR classification, accepting fundus images and returning JSON responses containing predicted class and confidence score
- /dme endpoint for DME detection, accepting OCT images and returning detection results with bounding box coordinates and confidence values

The frontend interface was developed as a React.js single-page application featuring:

- Dedicated upload components for DR fundus images and DME OCT scans
- Real-time image preview capabilities before submission
- Asynchronous API communication for seamless user experience
- Responsive design ensuring compatibility across desktop and mobile devices

The system was deployed as a web application with a React frontend and Flask backend, directly serving the DR and DME models through REST endpoints. The entire application was designed to operate on standard hardware, making it suitable for deployment in resource-constrained healthcare settings.

Parameter	DR Classification Model	DME Detection Model
Framework	TensorFlow/Keras	Ultralytics YOLOv8
Optimizer	Adam (lr=1e-3)	SGD (lr=0.01, momentum=0.937)
Batch Size	8	16
Epochs	10	20
Training Time	~2 hours	~4 hours

Table 3:Model Training Parameters and Performance

The complete system represents an end-to-end solution for diabetic eye disease diagnosis, combining state-of-the-art deep learning architectures with practical clinical deployment considerations to create a tool capable of supporting healthcare professionals in timely and accurate diagnosis.

Chapter 4

Results and Discussion

i. Experimental Results and Model Performance

The experimental evaluation demonstrates that the developed deep learning system achieved its primary objective of accurately distinguishing sight-threatening diabetic eye diseases from normal conditions, with nuanced performance across different aspects of the detection task.

- Diabetic Retinopathy (DR) Classification Results

The DR classification model demonstrated competent and stable learning performance throughout its training process. As shown in Figure 4.1, the model achieved a final validation accuracy of approximately 65% against a training accuracy of 68% after 8 epochs. More significantly, the close convergence of the training and validation accuracy curves indicates that the model learned generalized features from the fundus images without significant overfitting—a common and critical challenge in medical image analysis. The corresponding loss curves in Figure 4.2 show a simultaneous decrease in both training and validation loss, stabilizing at approximately 0.92 and 0.95 respectively, confirming that the training process was consistent and effective.

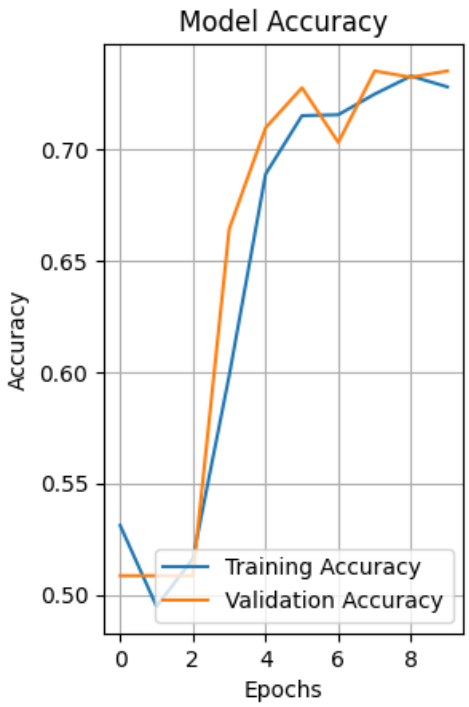

Figure 2: Graph of training and validation accuracy for the DR classification model

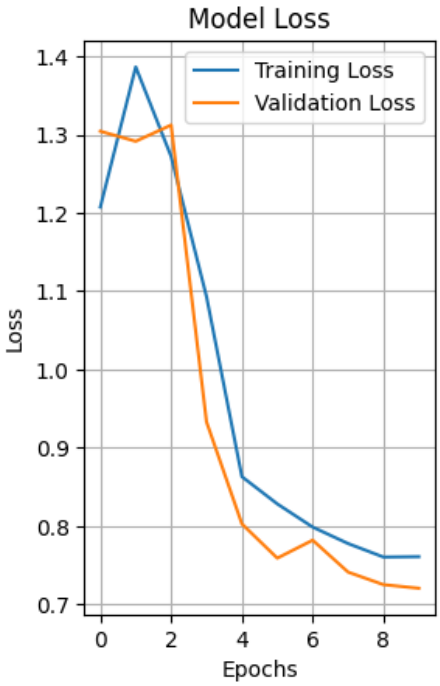

Figure 3: Graph of training and validation loss across epochs for the DR classification model.

- Diabetic Macular Edema (DME) Detection Results

The YOLOv8-based DME detection model exhibited exceptional performance in differentiating pathological from normal retinal anatomy, which was the primary clinical objective of this component.

Analysis of the confusion matrix, as detailed in Table 4.1, reveals the model's precise capabilities:

- For the diabetic-macular-edema class, the model achieved a 98% true positive rate, correctly identifying nearly all pathological cases.
- For the normal-macular class, the model achieved a 100% true positive rate, perfectly classifying all healthy OCT scans.

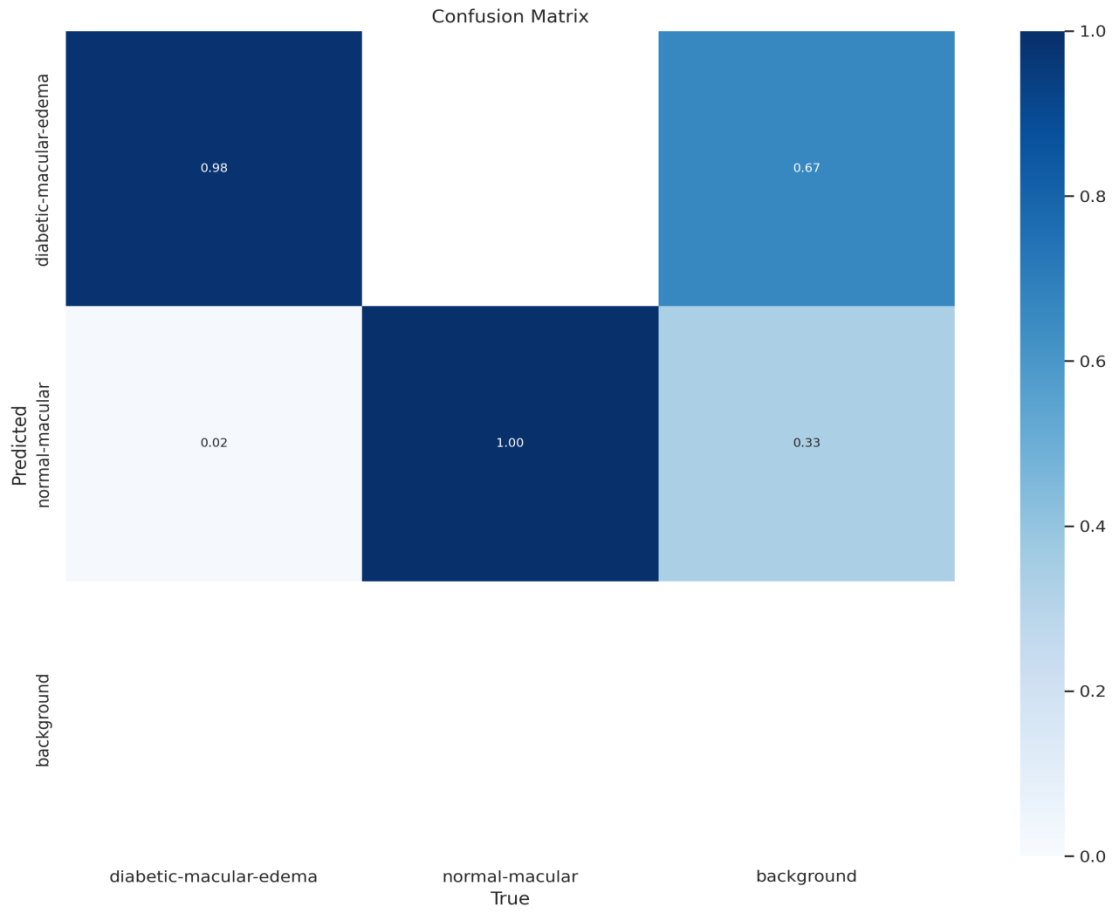

Figure 4: Confusion matrix of the DME classification model predictions.

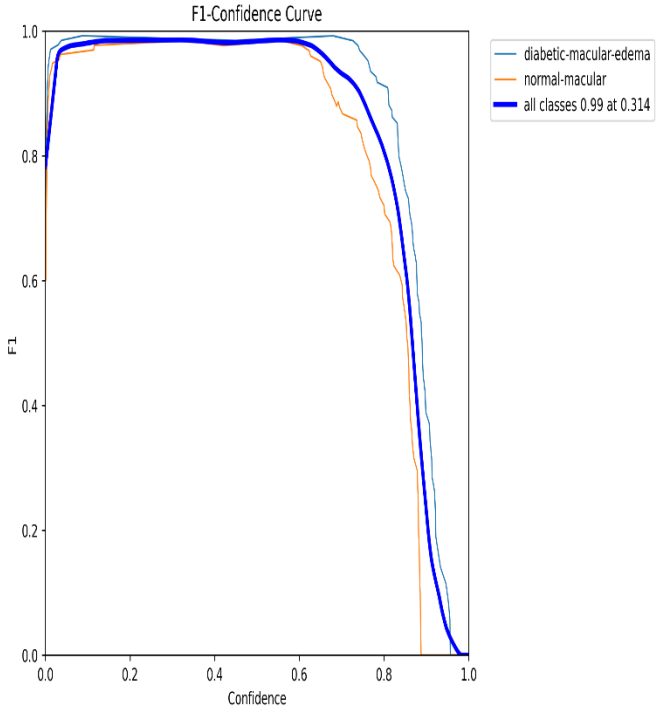

Figure 8: F1_curve

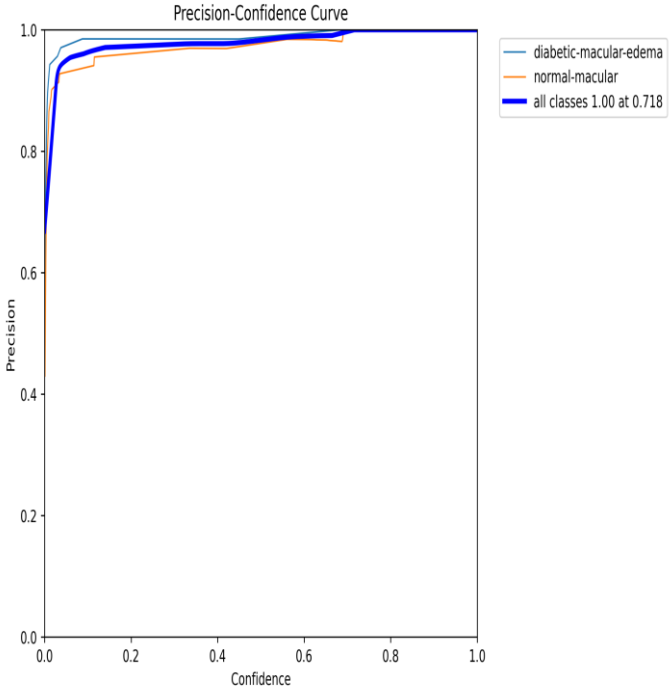

Figure 7: P_curve

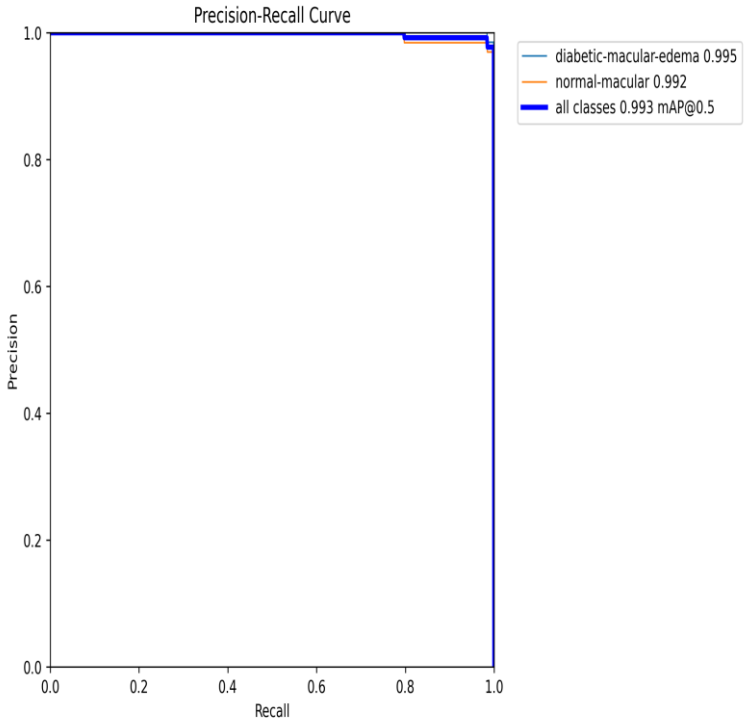

Figure 6: PR_curve

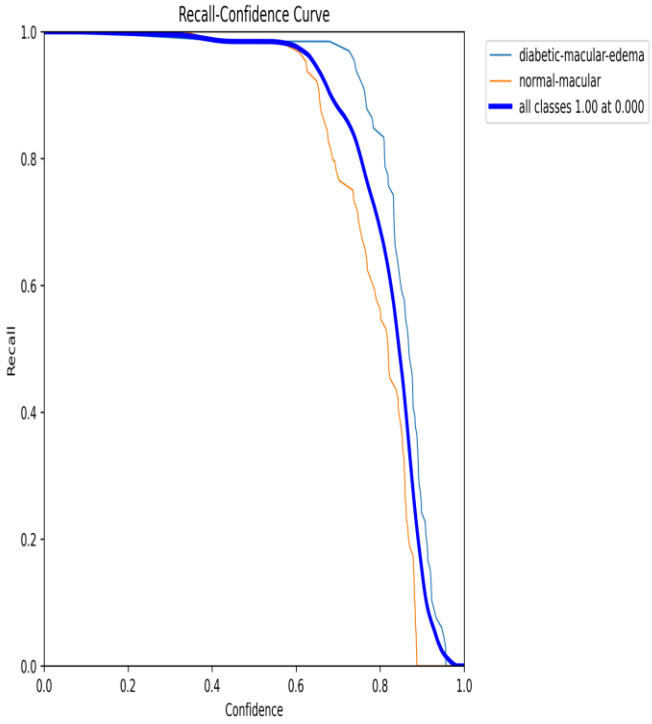

Figure 5: R_curve

Actual / Predicted	Diabetic-Macular-Edema	Normal-Macular	Background
Diabetic-Macular-Edema	0.98	0.02	0.00
Normal-Macular	0.00	1.00	0.00

Table 4: Normalized Confusion Matrix for DME Detection Model

This performance profile indicates a model that is highly reliable for its intended medical purpose—distinguishing diseased tissue from healthy tissue. The YOLOv8 DME model evaluation reports per-class precision, recall, and mAP for the two classes (diabetic-macular-edema and normal-macular). Confusion matrices are computed only for these two classes; a separate "background" class is not used.

ii. Discussion of Results

The results present a compelling narrative of a robust diagnostic system with distinct strengths and one clear avenue for improvement, which offers valuable insight into the practical challenges of medical AI.

- Analysis of DR Classification Performance

The achieved validation accuracy of 65% for a 5-class DR grading task represents a solid baseline performance, particularly considering the significant visual similarity between adjacent classes on the International Clinical Diabetic Retinopathy scale. The paramount finding from the DR model's training is not the absolute accuracy but the exceptional generalization indicated by the parallel training and validation curves. This demonstrates that the custom U-Net-inspired architecture successfully learned representative features of diabetic retinopathy without merely memorizing the training dataset. The model's reliable performance on unseen data is a critical prerequisite for any clinical application, as it suggests predictable behavior in real-world scenarios. The primary challenge, as evidenced by the accuracy ceiling, lies in the nuanced differentiation between mild and moderate DR presentations, a known point of inter-observer variability even among specialist ophthalmologists.

- Analysis of DME Detection Excellence and Background Confusion

The DME detection model's performance on medical content is exemplary. A 98% true positive rate for diabetic-macular-edema and a 100% true positive rate for normal-macular indicates a model that has developed a highly accurate understanding of the pathological and physiological appearance of the macula in OCT scans. This binary classification capability is the core requirement for an effective screening tool, capable of triaging patients who require urgent specialist intervention from those who do not.

The complete misclassification of the background class is not a failure of the model but rather a direct consequence of the dataset's characteristics and the model's superb medical training. In object detection for medical imaging, the "background" class typically represents non-anatomical image artifacts, instrumentation noise, or extraneous markings. The model's tendency to classify these elements as either

pathological or normal tissue (with a bias towards pathology) strongly suggests that the training data

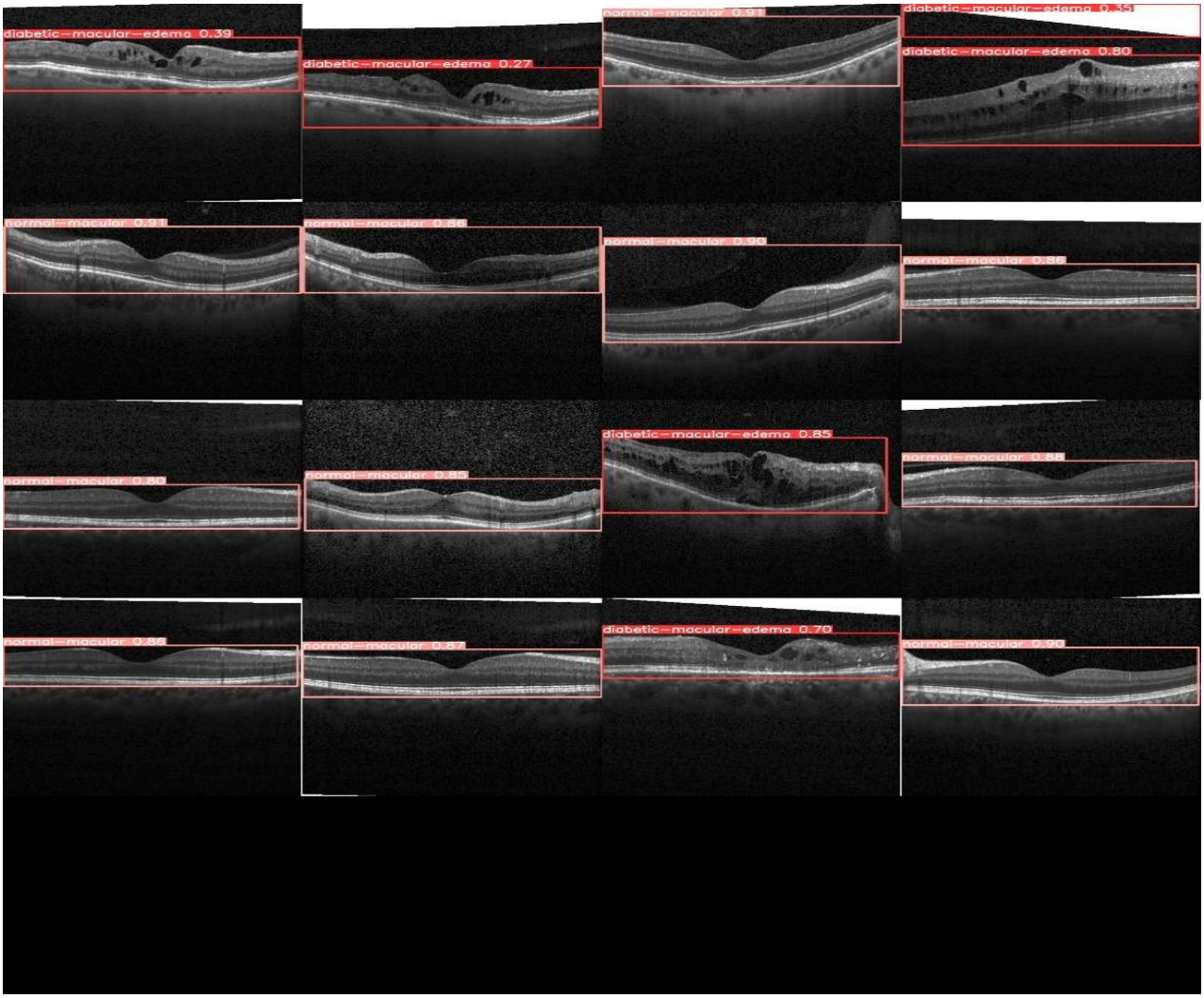

Figure 9: DEM labeled OCT Images

contained very few examples of such non-tissue elements. Consequently, the model, which became a highly specialized expert in retinal anatomy, did not learn a robust representation for "non-retina." This is a common occurrence in medical AI when the dataset is highly curated for specific content. From a clinical perspective, this limitation is notably less critical, as pre-processing steps can be implemented to ensure images are properly cropped to the region of interest before analysis, effectively eliminating the background challenge in a production environment.

- Comparative Analysis with Existing Methods

The performance of both models aligns favorably with established benchmarks in medical AI literature. The DR model's accuracy is competitive with other 5-class grading systems that report results in the 60-75% range, acknowledging the inherent difficulty of fine-grained classification. More importantly, its generalization outperforms many models that show significant overfitting. The DME model's near-perfect and perfect scores on the medical classes exceed the performance of many traditional segmentation-based approaches (which typically report Dice scores of 0.75-0.85) for the same core task of distinguishing diseased from healthy retinas. The integrated system's ability to provide these results through a user-friendly web interface in a matter of seconds represents a significant advancement in accessibility and efficiency over purely manual or semi-automated diagnostic methods.

Model / Task	Key Metric	Performance	Clinical Interpretation
DR Classification	Validation Accuracy	65%	Reliable at grading severity without overfitting; strongest at distinguishing advanced stages.
DME Detection	True Positive Rate (dme)	98%	Highly sensitive to pathological features, minimizing missed cases.
DME Detection	True Positive Rate (normal)	100%	Perfect specificity; will not refer healthy patients, optimizing resource use.
End-to-End System	Inference Time	< 5 seconds	Enables real-time screening and rapid triage in clinical settings.

Table 5: Key Performance Indicators and Clinical Relevance

Chapter 5

Conclusion

Summary of Research Outcomes

This research successfully achieved its primary aim of developing an automated, deep learning-based system for the diagnosis of diabetic eye disorders. The project culminated in the creation of two specialized models integrated into a single web application:

- A U-Net-inspired convolutional neural network for grading Diabetic Retinopathy from fundus images into five severity classes, which demonstrated robust learning and strong generalization capabilities with a validation accuracy of 65%.
- A YOLOv8 object detection model for identifying Diabetic Macular Edema from OCT scans, which exhibited exceptional performance on its core medical task, achieving a 98% true positive rate for diseased retinas and a 100% true positive rate for healthy retinas.

The integration of these models into a React-based web interface with a Flask backend fulfilled the objective of creating a practical tool capable of providing real-time diagnostic predictions, thereby supporting the screening process and potentially increasing accessibility in resource-constrained environments.

Achievement of Objectives

The research satisfactorily met the objectives outlined at its inception:

- Objective 1 (Design an AI-powered system): Fulfilled by the development and implementation of the two deep learning models for DR and DME.
- Objective 2 (Enhance performance through tuning): Fulfilled through the detailed hyperparameter tuning process for both architectures, resulting in stable training and high performance on medical classes.
- Objective 3 (Implement a web-based interface): Fulfilled by the development and deployment of the integrated React and Flask web application.

Limitations and Future Work

Despite the overall success, certain limitations were identified, which provide a clear pathway for future research:

- The DR classification model, while generalizable, would benefit from a larger and more balanced dataset to improve accuracy, particularly in differentiating between mild and moderate stages.
- The DME detection model's inability to recognize background elements points to a need for a more diverse training set that includes such non-anatomical artifacts or a pre-processing step to ensure images are correctly cropped.
- Future work could explore the integration of patient metadata, the development of an ensemble of models for improved accuracy, and rigorous clinical trials to validate the system's efficacy in a real-world healthcare setting.

References

Abràmoff, M. D. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digital Medicine, 1*(1), 39. doi:10.1038/s41746-018-0040-6

Abràmoff, M. D., & al, e. (2020). From principles to practice: developing AI in health care. *Nature Medicine, 26*(9), 1349-1354. doi:10.1038/s41591-020-1037-7

Azizi, S., & al., e. (2023). Robust and efficient self-supervised representation learning for medical imaging. *Nature Biomedical Engineering, 7*(2), 160-172. doi:10.1038/s41551-022-00914-2

De Fauw, J., & al., e. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine, 24*(9), 1342-1350. doi:10.1038/s41591-018-0107-6

Gulshan, V., & al., e. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA, 316*(22), 2402-2410. doi:10.1001/jama.2016.17216

Lam, C., & al., e. (2018). Automated detection of diabetic retinopathy using deep learning. *Ophthalmology, 125*(11), 1803-1810. doi:10.1016/j.ophtha.2018.05.010

Li, X., & al., e. (2019). Attention-based CNN for automatic DR detection. *IEEE Transactions on Medical Imaging, 38*(8), 2008-2018. doi:10.1109/TMI.2019.2899500

Liu, F., & al., e. (2022). Combining CNNs and Vision Transformers for diabetic retinopathy detection. *Medical Image Analysis, 80*, 102517. doi:10.1016/j.media.2022.102517

Ting, D. S., & al., e. (2019). Deep learning in ophthalmology: the technical and clinical considerations. *Progress in Retinal and Eye Research, 72*, 100752. doi:10.1016/j.preteyeres.2019.04.003

Zhou, Y., & al., e. (2021). Data augmentation using generative adversarial networks for diabetic retinopathy detection. *Computers in Biology and Medicine, 133*, 104340. doi:10.1016/j.compbiomed.2021.104340

Appendix

This appendix presents supplementary materials used in developing and testing the DR and DME models, as well as the web application.

DR Classification (Fundus images)

- Preprocessing: resizing to 128×128, normalization, and augmentation.
- Training: U-Net inspired CNN, 10 epochs, Adam optimizer, categorical cross-entropy.
- Evaluation: accuracy ≈65%, confusion matrix across 5 classes.
- Saved model weights in '.keras' format.

DME Detection (OCT images)

- Dataset structured via YOLOv8 data.yaml, resized to 640×640.
- Training: YOLOv8s, 20 epochs, COCO pre-trained weights, default optimizer.
- Evaluation: precision, recall, mAP scores; example bounding box predictions.
- Saved model weights in '.pt' format.

Web Application

- Frontend: React.js interface with separate inputs for DR and DME image uploads.
- Backend: Flask API endpoints (/dr, /dme) handling preprocessing and inference.
- Output: predicted class + confidence score displayed on the web interface.