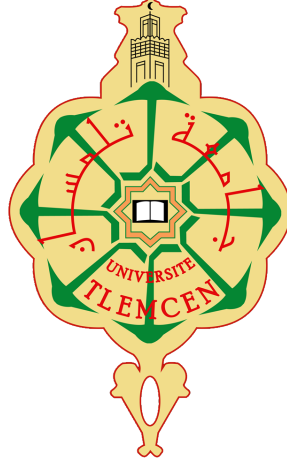


الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education and Scientific Research
جامعة أبي بكر بلقايد تلمسان
Abou Bekr Belkaid University - Tlemcen
Faculty of Technology



Thesis

Presented to obtain the degree of
DOCTORATE Third Cycle Degree
Field of Study: **Biomedical Engineering**
Specialization: **Medical Imaging**

Theme

**Intelligent Retinal Image Analysis
for Medical Biometrics applications**

Presented by
MOKHTARI Aicha

In front of the jury composed of

Mr. DJEBBARI Abdelghani	Professor	Univ. Tlemcen	President
Mr. HADJ SLIMANE Zine-Eddine	Professor	Univ. Tlemcen	Supervisor
Mr. BETAOUAF Talib Hichem	MCA	Univ. Tlemcen	Examinator
Mr. BENGHNIA Hadj AbdelKader	MCA	Univ. Mascara	Examinator

Academic Year: 2025-2026

Abstract

The rapid integration of artificial intelligence into medical imaging has created new opportunities for improving both clinical diagnosis and secure patient identification. Among medical imaging modalities, retinal fundus photography is particularly valuable due to the uniqueness, stability, and diagnostic richness of the retinal vascular network. Retinal images simultaneously encode highly discriminative biometric patterns and clinically significant biomarkers associated with ocular and systemic diseases. However, existing studies typically address biometric recognition and pathology classification as separate tasks, limiting the development of unified intelligent systems.

This thesis proposes an integrated framework for intelligent retinal image processing in medical biometric applications, combining deep learning, evolutionary hyperparameter optimization, and multitask learning. A robust retinal authentication pipeline is developed, including advanced preprocessing, residual attention U-Net vessel segmentation, Gaussian-weighted patch reconstruction, and multi-modal feature extraction incorporating morphological, topological, and texture-based descriptors. Furthermore, hyperparameter optimization strategies are systematically evaluated, demonstrating the effectiveness of the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) in achieving improved convergence and generalization compared to conventional methods. Finally, a unified multitask architecture with dual backbones and attention mechanisms is introduced to simultaneously perform biometric identification and multi-label pathology classification.

Experimental validation on multiple public and clinical datasets confirms high recognition accuracy, strong robustness to image perturbations, and competitive diagnostic performance. The results demonstrate that retinal fundus images can serve as dual-purpose data sources, supporting secure authentication and automated clinical assessment within a single intelligent framework.

Keywords: Medical Biometrics, Retinal Authentication, Hyperparameter Optimization, Multitask Learning, deep learning, retina.

Résumé

L'intégration rapide de l'intelligence artificielle dans l'imagerie médicale ouvre de nouvelles perspectives pour le diagnostic clinique et l'identification sécurisée des patients. Parmi les modalités d'imagerie, la photographie du fond d'œil se distingue par l'unicité, la stabilité et la richesse diagnostique du réseau vasculaire rétinien. Les images rétiniennes contiennent simultanément des motifs biométriques hautement discriminants et des biomarqueurs cliniquement pertinents associés à diverses pathologies oculaires et systémiques. Toutefois, la majorité des travaux existants traitent la reconnaissance biométrique et la classification des pathologies comme des tâches indépendantes, limitant ainsi le développement de systèmes intelligents unifiés.

Cette thèse propose un cadre intégré de traitement intelligent des images rétiniennes pour des applications biométriques médicales, combinant apprentissage profond, optimisation évolutive des hyperparamètres et apprentissage multitâche. Un pipeline robuste d'authentification rétinienne est développé, incluant un prétraitement avancé, une segmentation vasculaire basée sur un U-Net à attention résiduelle, une reconstruction pondérée gaussienne et une extraction multi-modale de caractéristiques morphologiques, topologiques et texturales. L'optimisation des hyperparamètres est étudiée de manière comparative, mettant en évidence l'efficacité de la méthode CMA-ES en termes de convergence et de généralisation. Enfin, une architecture multitâche unifiée, fondée sur une double architecture backbone avec mécanismes d'attention, permet l'identification biométrique et la classification multi-étiquettes des pathologies.

Les résultats expérimentaux confirment la précision biométrique, la robustesse face aux perturbations et des performances diagnostiques compétitives.

Mots-clés : Biométrie médicale, Authentification rétinienne, Optimisation des hyperparamètres, Apprentissage multitâche, Apprentissage profond, Rétine.

ملخص

أدى التطور السريع في تقنيات الذكاء الاصطناعي وتطبيقها في التصوير الطبي إلى إتاحة فرص جديدة لتحسين التشخيص السريري وضمان التعرف الآمن على هوية المرضى. وتُعد صور قاع العين من أكثر وسائل التصوير تميزاً نظراً لتفرد الشبكة الوعائية الشبكية واستقرارها وغناها بالمعلومات التشخيصية. إذ تحتوي هذه الصور على أنماط بيومترية عالية الدقة، إضافة إلى مؤشرات حيوية مهمة ترتبط بأمراض عينية وجهازية متعددة. ومع ذلك، فإن معظم الدراسات الحالية تتعامل مع التعرف البيومتري وتصنيف الأمراض كجالين منفصلين، مما يحد من تطوير أنظمة ذكية موحدة. تقدم هذه الأطروحة إطاراً متكاملًا للمعالجة الذكية لصور الشبكية في التطبيقات البيومترية الطبية، يجمع بين التعلم العميق، والتحسين التطوري للمعاملات الفائقة، والتعلم متعدد المهام. تم تطوير نظام متكامل للمصادقة الشبكية يشمل المعالجة المسبقة المتقدمة، وتقسيم الأوعية الدموية باستخدام شبكة U-Net بآلية انتباه متبقي، وإعادة البناء بالترجيح الغاوسي، إضافة إلى استخراج خصائص متعددة الأنماط تشمل الخصائص الشكلية والطوبولوجية والنسيجية. كما تم تقييم استراتيجيات تحسين المعاملات الفائقة، حيث أثبتت خوارزمية CMA-ES فعاليتها من حيث سرعة التقارب وقدرة التعميم. وأخيراً، تم اقتراح بنية تعلم متعددة المهام تعتمد على معماريتين أساسيتين مع آليات انتباه، مما يسمح بالتعرف البيومتري وتصنيف أمراض الشبكية في نموذج موحّد. تؤكد النتائج التجريبية دقة عالية في التعرف البيومتري، ومثانة أمام التشويش، وأداءً تشخيصياً تنافسياً.

الكلمات المفتاحية: البيومترية الطبية، المصادقة الشبكية، تحسين المعاملات الفائقة، التعلم متعدد المهام، التعلم العميق، الشبكية.

Acknowledgements

First and foremost, I would like to express my sincere and deepest gratitude to my supervisor, **Professor HADJ SLIMANE Zine-Eddine**, for his invaluable guidance, constant support, and insightful advice throughout the development of this research. His scientific rigor, patience, and encouragement have been fundamental to the completion of this thesis.

I am highly honored by the presence of the distinguished members of the jury and would like to express my sincere appreciation to:

Professor DJEBBARI Abdelghani for kindly accepting to preside over this jury and for his valuable time and consideration. **Mr. BETAOUAF Talib Hichem** for accepting to examine this thesis and for his constructive remarks. **Mr. BENGHNIA Hadj AbdelKader** for his willingness to serve as examiner and for his thoughtful evaluation.

I would also like to express my sincere gratitude to **Mr. LAZOUNI Mohamed Amine** for his valuable contribution to the development and organization of the local retinal database used in this study. His support and collaboration were greatly appreciated and significantly contributed to the experimental validation of this work.

My sincere thanks also go to my colleagues and members of the research laboratory for their collaboration, discussions, and supportive academic environment.

Finally, I would like to express my heartfelt gratitude to my family for their unconditional support, patience, and encouragement throughout my academic journey. Their belief in me has been a constant source of strength and motivation.

To all those who contributed directly or indirectly to this work, I extend my sincere appreciation.

Dedication

I dedicate this thesis to my beloved family, whose unwavering support, encouragement, and love have been the foundation of my academic journey.

To my parents, for instilling in me the values of perseverance, curiosity, and integrity; to my siblings, for their constant motivation and understanding during the most challenging moments of this research.

This work is also dedicated to all mentors, teachers, and colleagues who have guided, inspired, and believed in me, helping me transform curiosity into knowledge.

Finally, I dedicate this achievement to all patients and healthcare professionals whose dedication and trust motivate the pursuit of innovation in medical science.

Contents

List of Figures	xi
List of Tables	xiii
List of Abbreviations	xv
General Introduction	1
1 Fundamentals of Medical Biometrics	4
1.1 Introduction	4
1.2 Definition and Scope of Biometrics	5
1.2.1 Physiological Biometrics	6
1.2.2 Behavioral Biometrics	7
1.3 Biometric System Architecture and Processing Pipeline	8
1.3.1 Data Acquisition and Enrollment	9
1.3.2 Feature Extraction: From Handcrafted to Deep Representations . .	10
1.3.3 Template Storage and Security Considerations	10
1.3.4 Matching and Decision-Making	11
1.4 Identification, Verification, and Authentication	12
1.4.1 Verification (1:1 Comparison)	12
1.4.2 Identification (1:N Search)	13
1.4.3 Authentication (Security Process)	13
1.5 Medical Biometrics: Motivation and Advantages	14
1.6 Retinal Biometrics	16
1.6.1 Anatomical Basis of Retinal Biometrics	16
1.6.2 Retinal Imaging Modalities	17
1.6.3 Historical Development of Retinal Biometric Systems	18
1.7 Challenges in Clinical Retinal Biometric Deployment	20
1.8 State of the Art in Retinal Biometrics: A Synthetic Review	21
1.8.1 From Handcrafted Features to Deep Representations	22
1.8.2 The Role of Segmentation in Retinal Biometrics	22

1.8.3	Metric Learning for Biometric Embeddings	23
1.8.4	Addressing Data Scarcity: Self-Supervised and Semi-Supervised Learning	23
1.8.5	Domain Adaptation and Federated Learning for Clinical Deployment	23
1.8.6	The Persistent Divide: Biometric Recognition vs. Pathology Classification	24
1.8.7	Hyperparameter Optimization: An Underexplored Dimension . . .	24
1.9	Research Gaps and Thesis Positioning	25
1.10	Chapter Summary	26
2	Theoretical Foundations of Artificial Intelligence for Biometric Systems	27
2.1	Introduction	27
2.2	Fundamentals of Machine Learning for Biometrics	27
2.2.1	Definition of Machine Learning	27
2.2.2	Classical Machine Learning Algorithms for Retinal Biometrics . . .	28
2.2.3	Artificial Neural Networks	31
2.3	Deep Learning Foundations	32
2.3.1	Convolutional Neural Networks	32
2.3.2	Parameters of Deep Learning	34
2.3.3	Regularization Techniques in Deep Learning	35
2.3.4	Training of Deep Neural Networks	36
2.4	Transfer Learning	37
2.4.1	ResNet Architecture	37
2.4.2	DenseNet Architecture	41
2.5	Retinal Image Segmentation Theory	44
2.5.1	U-Net Architecture for Retinal Segmentation	44
2.6	Attention Mechanisms: Theoretical Foundations	46
2.7	Multitask Learning	47
2.7.1	Hard and Soft Parameter Sharing	47
2.7.2	Benefits of Multitask Learning	48
2.8	Common Datasets Overview	49
2.8.1	Publicly Available Datasets	49
2.8.2	Local Dataset	51
2.9	Evaluation Metrics for Retinal Image Analysis	52
2.9.1	Pixel-Level Segmentation Metrics	52
2.9.2	Classification and Identification Metrics	53
2.9.3	Biometric-Specific Metrics	54
2.9.4	Multi-Label Classification Metrics	55
2.10	Conclusion	55

3	Retinal Authentication System Using Deep Learning and Multi-Modal Feature Extraction	57
3.1	Introduction	57
3.2	Methodology	59
3.2.1	Image Preprocessing	60
3.2.2	Patch Extraction Strategy	61
3.2.3	Model Development	64
3.2.4	Loss Function and Training Protocol	65
3.2.5	Gaussian-Weighted Reconstruction	66
3.2.6	Retinal Vessel Feature Extraction	67
3.2.7	Biometric Gallery Construction and Pair Generation	71
3.2.8	Biometric Verification Model	72
3.3	Results and Discussion	73
3.3.1	Retinal Vessel Segmentation Results	73
3.3.2	Biometric Verification Results	75
3.3.3	Comparison with State-of-the-Art Verification Methods	81
3.4	Conclusion	83
4	Hyperparameter Optimization for Retinal Biometrics	84
4.1	Introduction	84
4.2	Motivation and Research Hypothesis	85
4.3	Description of the Proposed System	85
4.4	Datasets	86
4.5	Preprocessing	86
4.5.1	Geometric Transformations	86
4.5.2	Photometric Augmentations	86
4.5.3	Image Preprocessing Pipeline	87
4.6	Model Development	88
4.7	Hyperparameter Optimization Strategies	89
4.7.1	Grid Search Methodology	90
4.7.2	Random Search	91
4.7.3	Bayesian Optimization	93
4.7.4	Covariance Matrix Adaptation Evolution Strategy	95
4.8	Results and Discussion	100
4.8.1	Grid Search	100
4.8.2	Random Search	102
4.8.3	Bayesian Optimization	103
4.8.4	CMA-ES Method	106
4.8.5	Comparative Analysis of Optimization Methods	110

4.8.6	Performance Evaluation Metrics	114
4.8.7	Learning Curves Analysis	115
4.8.8	Robustness Analysis	117
4.8.9	Comparative Analysis with State-of-the-Art	119
4.9	Conclusion	122
5	Multitask Learning for Retinal Analysis	123
5.1	Introduction	123
5.2	Deep Learning Revolution in Retinal Image Analysis	123
5.3	Retinal Fundus Imaging Beyond Diagnosis: Biometric Recognition	124
5.4	Privacy and Ethical Considerations	124
5.5	The Persistent Divide Between Diagnosis and Biometrics	125
5.6	Multitask Learning as a Unifying Paradigm	125
5.7	Contributions of This Work	126
5.8	Materials and Methods	126
5.8.1	Data Preprocessing and Augmentation	127
5.8.2	Proposed Unified Multitask Architecture	128
5.8.3	Training Protocol	134
5.9	Results and Discussion	135
5.9.1	Results on ODIR-2019 Dataset	135
5.9.2	Results on 39-CAS Dataset	137
5.9.3	Results on FIVES Dataset	142
5.9.4	Comparative Discussion Across Datasets	144
5.9.5	Comparative Analysis with State-of-the-Art Methods	146
5.10	Conclusion	149
6	General Conclusion and Future Perspectives	150
6.1	Introduction	150
6.2	Synthesis of Core Contributions	150
6.2.1	Robust Retinal Authentication Framework	150
6.2.2	Systematic Hyperparameter Optimization	151
6.2.3	Unified Multitask Learning Framework	151
6.3	Overall Contribution to the Field	152
6.4	Limitations of the Current Work	152
6.5	Future Research Directions	153
6.6	Final Remarks	153

List of Figures

1.1	Hierarchical classification of biometric modalities	8
1.2	Data acquisition and enrollment process	9
1.3	Multi-sample enrollment strategy	10
1.4	Conceptual illustration of the verification process	12
1.5	Conceptual illustration of the identification process	13
1.6	Conceptual illustration of the authentication process	14
1.7	Anatomical structure of the human retina	17
1.8	Concept of a retinal biometric system	18
1.9	Historical retinal imaging system for medical biometric applications	20
2.1	Convolution operation in CNNs. A filter (kernel) slides across the input image to produce a feature map.	33
2.2	Example of max pooling and average pooling operations with a 2×2 filter and stride 2.	34
2.3	Architectural diagram of ResNet-152, illustrating the convolutional stem, four residual stages, and the final classification layers.	40
2.4	Architectural diagrams of DenseNet-169 and DenseNet-201 backbones, showing the progression of dense blocks and transition layers.	43
3.1	Proposed retinal authentication system pipeline.	59
3.2	Preprocessing pipeline applied to a sample retinal fundus image.	61
3.3	Patch extraction methodology applied to a sample retinal fundus image from the DRIVE dataset.	62
3.4	Model architecture.	65
3.5	Illustration of Gaussian-weighted reconstruction. Overlapping 64×64 patches contribute to each pixel in the final segmentation mask, with higher weights assigned near patch centers.	67
3.6	Morphological Features: (a) Input Binary Mask, (b) Distance Transform, (c) Tortuosity Map.	68
3.7	Topological Features: (a) Skeletonization, (b) Bifurcation Detection, (c) Spatial Grid, (d) Orientation Histogram.	69

3.8	Texture-Based Descriptors: (a) LBP Texture Map, (b) LBP Histogram, (c) Fractal Dimension Analysis.	71
3.9	Qualitative segmentation results across multiple datasets.	75
3.10	Biometric Gallery Construction and Pair Generation for Verification	76
3.11	Qualitative segmentation results across multiple datasets.	80
4.1	Overview of the model development pipeline.	85
4.2	Results of Retinal Image Processing. (a) Original input image, (b) L color space representation, (c) Contrast-Limited Adaptive Histogram Equalization (CLAHE) enhanced image, (d) Merged processed image.	88
4.3	Schematic representation of the proposed model.	89
4.4	Effect of Dropout Rate p_{dropout} and Learning Rate η on Validation Loss $\mathcal{L}_{\text{val}}$: A Three-Dimensional Visualization across Multiple Datasets.	106
4.5	Training and Validation Learning Curves for the RIDB Dataset.	115
4.6	Training and Validation Learning Curves for the ARIA Dataset.	116
4.7	Training and Validation Learning Curves for the STARE Dataset.	116
4.8	Training and Validation Learning Curves for the DRIVE Dataset.	116
4.9	Training and Validation Learning Curves for the VARIA Dataset.	117
4.10	Training and Validation Learning Curves for the LOCAL Dataset.	117
4.11	Example of a Perturbed Input Image.	118
5.1	Unified multitask learning framework for simultaneous retinal biometric identification and pathology classification.	127
5.2	Dual-backbone feature extraction module for the proposed multitask retinal analysis system.	129
5.3	Attention-Based Feature Refinement and Aggregation process.	130
5.4	Selective Feature Aggregation process.	132
5.5	Shared Representation Learning Stage.	133
5.6	Task-Specific Output Heads Step.	134
5.7	Comparative pathology classification performance across ODIR-2019, 39-CAS, and FIVES datasets.	145
5.8	Comparative biometric identification performance across ODIR-2019, 39-CAS, and FIVES datasets.	146

List of Tables

3.1	Cross-Dataset Evaluation (Trained on Combined Datasets, Tested Separately)	74
3.2	Performance Comparison of Verification Classifiers on RIDB Dataset . . .	77
3.3	Performance Comparison of Verification Classifiers on DRIVE Dataset . .	77
3.4	Performance Comparison of Verification Classifiers on ARIA Dataset . . .	78
3.5	Performance Comparison of Verification Classifiers on VARIA Dataset . . .	78
3.6	Performance Comparison of Verification Classifiers on STARE Dataset . .	79
3.7	Comparison of Retinal Biometric Verification Performance with State-of-the-Art Methods	81
4.1	Grid Search Optimization Outcomes Across Retinal Datasets	102
4.2	Performance Metrics and Optimal Hyperparameters Identified through Random Search Across Six Retinal Image Datasets	103
4.3	Bayesian Optimization Results Across Retinal Image Datasets	104
4.4	Consolidated CMA-ES Optimization Results Across All Datasets	109
4.5	Best Validation Loss $\mathcal{L}_{\text{best}}$ Achieved by Each Optimization Strategy Across Datasets	111
4.6	Comparative Analysis of Hyperparameter Optimization Methods	112
4.7	Performance Evaluation Metrics Across Datasets	114
4.8	Robustness Evaluation Under Noise Perturbations	118
4.9	Comparative Performance of Retinal Biometric Approaches	120
5.1	Overall Evaluation Metrics on ODIR-2019	136
5.2	Per-Class Pathology Classification Results on ODIR-2019	137
5.3	Biometric Identification Results on ODIR-2019	137
5.4	Overall Evaluation Metrics on 39-CAS	138
5.5	Classification Performance for Normal / Baseline Categories on 39-CAS . .	138
5.6	Classification Performance for Diabetic Retinopathy Categories on 39-CAS	139
5.7	Classification Performance for Glaucoma and Optic Nerve-Related Pathologies on 39-CAS	140
5.8	Classification Performance for Retinal Vascular Diseases on 39-CAS	140
5.9	Classification Performance for Macular & Retinal Disorders on 39-CAS . .	141

5.10	Classification Performance for Retinal Dystrophies & Deposits on 39-CAS .	141
5.11	Classification Performance for Other Retinal Conditions on 39-CAS	142
5.12	Biometric Identification Performance Metrics on 39-CAS	142
5.13	Overall Pathology Classification Performance on FIVES	143
5.14	Per-Class Pathology Classification Report on FIVES	144
5.15	Biometric Identification Performance on FIVES	144
5.16	Comparison of the proposed method with state-of-the-art pathology classi- fication approaches on fundus photography datasets	147
5.17	Comparison of the proposed method with state-of-the-art retinal biometric identification approaches	149

List of Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
AMD	Age-related Macular Degeneration
ARIA	Automatic Retinal Image Analysis
AUC	Area Under the Curve
BO	Bayesian Optimization
CBAM	Convolutional Block Attention Module
CLAHE	Contrast Limited Adaptive Histogram Equalization
CMA-ES	Covariance Matrix Adaptation Evolution Strategy
CNN	Convolutional Neural Network
DR	Diabetic Retinopathy
DRIVE	Digital Retinal Images for Vessel Extraction
DT	Decision Tree
EER	Equal Error Rate
FAR	False Acceptance Rate
FIVES	Fundus Image Vessel Segmentation
FRR	False Rejection Rate
GAN	Generative Adversarial Network
GAP	Global Average Pooling
GAR	Genuine Acceptance Rate
KNN	k-Nearest Neighbors
LBP	Local Binary Pattern
ML	Machine Learning
MTL	Multitask Learning
OCT	Optical Coherence Tomography
ODIR	Ocular Disease Intelligent Recognition
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
ResNet	Residual Network
RF	Random Forest
RIDB	Retinal Identification Database
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
U-Net	U-shaped Network
VARIA	Vascular Analysis for Retinal Identification Applications
XGBoost	Extreme Gradient Boosting

General Introduction

1. Context

Medical imaging has become indispensable in modern clinical practice, providing critical insights into human health and disease. Among various modalities, retinal fundus photography is particularly valuable due to the unique, stable, and diagnostically rich nature of the retinal vascular network. Beyond its primary role in ophthalmic diagnosis, the retina encodes highly discriminative individual-specific patterns, giving rise to a novel research domain: medical biometrics. Unlike traditional biometrics, which rely on external features, medical biometrics leverages internal physiological markers that are difficult to forge, are routinely acquired during clinical care, and rely on the longitudinal stability of retinal imaging as a cornerstone of emerging patient identification systems. To realize this potential, advanced computational methods, particularly artificial intelligence and deep learning, have been increasingly applied to retinal image analysis.

2. Research Problem

Despite advances in artificial intelligence and deep learning for medical image analysis, challenges persist that hinder the development of reliable, integrated retinal biometric systems.

First, many existing retinal recognition approaches lack robustness in real-world clinical environments. Retinal images often suffer from changes in illumination and differences in acquisition devices. Patient movement and noise can also affect image quality. These issues degrade vascular patterns, which are the primary biometric features. Earlier systems often used handcrafted descriptors or simple deep learning pipelines. Such methods may not capture the complex vascular structures needed for accurate identification.

Second, deep learning models designed for retinal analysis are highly sensitive to hyperparameter configurations, such as learning rates and regularization strategies, which play a crucial role in determining model performance. However, in many studies, these parameters are selected through manual experimentation, which is often time-consuming

and subjective. Such practices limit the reproducibility of results and reduce models' ability to generalize across heterogeneous datasets from different imaging devices or clinical populations.

Biometric recognition and retinal pathology analysis were once studied separately. However, both rely on the same anatomical structures, especially the vascular network. Diseases like diabetic retinopathy can alter these structures over time. This affects the stability of biometric features. Clinical datasets are often heterogeneous, imbalanced, and only partially annotated. This makes model development and evaluation more difficult. Retinal images contain both biometric identity and sensitive medical biomarkers. This raises concerns about privacy and data security.

These limitations show the need for new methods. Such methods must address both biometric and clinical goals. There is a growing demand for robust authentication that works under varied conditions. Systematic hyperparameter optimization can make models more reliable and reproducible. Also, unified deep learning frameworks could combine biometric identification and retinal pathology analysis. These frameworks would use a single multitask architecture that adapts to clinical variability.

3. Research Objectives

The main objective of this thesis is to develop intelligent retinal image analysis techniques that enable both reliable biometric recognition and automated detection of retinal pathology. To achieve this objective, the research focuses on three key goals:

1. Designing robust vascular feature extraction techniques capable of capturing discriminative retinal patterns for biometric identification
2. Investigate efficient hyperparameter optimization to boost deep learning performance and generalization
3. Develop a unified multitask framework for biometric identification and retinal pathology classification.

4. Thesis Contributions

This thesis presents three main contributions to medical retinal biometrics .

The first contribution is a retinal authentication framework using robust vascular feature extraction. This framework integrates advanced preprocessing, vessel segmentation via a Residual Attention U-Net, and multi-modal feature descriptors based on morphological,

topological, and texture features of the retinal vascular network.

The second contribution is a systematic study of hyperparameter optimization strategies for retinal biometric systems. Specifically, it evaluates grid search, random search, Bayesian optimization, and CMA-ES to identify configurations that enhance model performance and generalization.

The third contribution is a unified multitask deep learning architecture that performs both biometric identification and multi-label retinal pathology classification. By leveraging shared representations from retinal images, this approach boosts both biometric and diagnostic results.

5. Significance

This research contributes to clinical practice by enabling automated patient identification directly from retinal images, reducing administrative errors, and supporting longitudinal monitoring. Methodologically, it demonstrates the mutual benefit of combining biometric and diagnostic objectives, extending multitask learning to medical imaging. Ethically, it addresses privacy concerns through responsible AI design. Practically, it bridges the gap between research prototypes and deployable clinical systems.

6. Thesis Structure

This dissertation is organized into six chapters.

Chapter 1 introduces the fundamental concepts of medical biometrics and presents the main characteristics of retinal biometric systems.

Chapter 2 presents the theoretical foundations of artificial intelligence and deep learning techniques used in biometric recognition and medical image analysis.

Chapter 3 describes the proposed retinal authentication framework based on vascular segmentation and multi-modal feature extraction.

Chapter 4 investigates hyperparameter optimization strategies for improving the performance of retinal biometric models.

Chapter 5 presents the proposed multitask deep learning architecture for unified retinal biometric identification and pathology classification.

Finally, **Chapter 6** concludes and summarizes the main findings of the thesis and discusses potential directions for future research.

Chapter 1

Fundamentals of Medical Biometrics

1.1 Introduction

Accurate patient identification is a fundamental prerequisite for safe and effective healthcare. Misidentification remains a persistent challenge worldwide, with studies reporting that up to one in ten patients may experience errors during hospital admission or registration (Li et al., 2020). Such errors can have severe consequences, including incorrect medication administration, diagnostic delays, inappropriate treatment, and compromised patient safety. Therefore, ensuring reliable patient identification is not merely an administrative requirement but a critical component of clinical care quality and patient safety.

The digitization of healthcare systems has led to the systematic acquisition and long-term storage of large volumes of patient physiological and medical imaging data. Commonly collected modalities include retinal fundus images and optical coherence tomography (OCT) scans, which are routinely acquired for diagnosis, screening, and longitudinal patient monitoring. These extensive repositories, integrated within modern hospital information systems, provide unprecedented opportunities to exploit medical data for secondary applications, including biometric recognition systems. The convergence of large-scale clinical data availability and advances in computational technologies, therefore opens new perspectives for developing integrated solutions that combine healthcare analytics with secure patient identification.

Parallel to this growth in medical data, artificial intelligence (AI), particularly deep learning, has advanced significantly. Deep learning-based representation learning enables the automatic extraction of highly discriminative and invariant features from complex medical data, surpassing traditional handcrafted methods in terms of robustness and scalability (Esteva et al., 2021). Additionally, multitask learning frameworks have emerged as a promising paradigm, allowing the simultaneous optimization of multiple related objectives within a single model (Chen et al., 2021). In the context of medical biometrics,

this approach enables systems to perform patient identification while extracting clinically relevant information, thereby creating a dual-use framework that aligns with both security and healthcare objectives.

Traditional biometric systems primarily rely on externally observable traits such as fingerprints, facial appearance, and voice. Although these modalities have achieved widespread success in civil and commercial applications, their deployment in healthcare environments faces limitations. External traits are vulnerable to spoofing, subject to degradation from injury or aging, and may conflict with clinical workflows and hygiene requirements. In contrast, medical biometrics exploit internal physiological characteristics that are intrinsically linked to human anatomy and clinical examination. Endogenous identifiers, such as retinal vascular patterns, are inherently difficult to forge, exhibit strong individuality, and often demonstrate high longitudinal stability (Simon and Goldstein, 1935; Yu et al., 2025).

The integration of AI and deep learning with medical biometrics enables the extraction of complex, high-dimensional representations directly from physiological and imaging data. Modern learning architectures facilitate robust feature extraction from modalities such as fundus images, OCT scans, and physiological signals, addressing the limitations of traditional handcrafted features (Uhl et al., 2020; Singh et al., 2026). Moreover, multimodal fusion and intelligent integration strategies enhance system reliability by accommodating intra-subject variability, acquisition noise, and modality-specific limitations (Poornima and Subramanian, 2022). This combination of data availability, AI-based representation learning, and multimodal optimization forms the foundation of this thesis.

This chapter establishes the theoretical and practical foundations of medical biometric systems. It begins by defining biometrics and its core concepts, then examines the architecture of biometric systems, and finally provides a focused discussion of medical and retinal biometrics. The chapter concludes by identifying the research gaps that motivate the contributions of this thesis.

1.2 Definition and Scope of Biometrics

The term "biometrics" is derived from the Greek words "bios" (life) and "metron" (measure), and broadly refers to the measurement and analysis of biological characteristics. In the context of identification and authentication systems, biometrics denotes the science and technology of recognizing individuals based on distinctive physiological, behavioral, or functional traits. Unlike conventional authentication mechanisms that rely on knowledge-based credentials such as passwords or personal identification numbers (PINs), or on possession-based tokens such as physical keys, identification cards, or smart cards, biometric systems establish identity through intrinsic human characteristics that are inherently linked

to the individual.

This intrinsic association provides several fundamental advantages. Biometric traits cannot be easily forgotten, lost, or transferred, thereby reducing common security risks associated with traditional authentication methods. Moreover, biometric systems enhance user convenience by eliminating the need to memorize multiple credentials or manage physical tokens. When properly designed and implemented, they also offer strong resistance to impersonation by tying authentication to the individual's physical presence.

Biometric identifiers encompass a diverse range of measurable attributes that can be broadly classified into physiological and behavioral categories.

1.2.1 Physiological Biometrics

Physiological biometrics are derived from the anatomical structure or internal biological functions of the human body. Well-established physiological biometric modalities include:

- **Fingerprint patterns**, formed by unique epidermal ridge structures on the fingertips
- **Palmprint patterns**, capturing principal lines, wrinkles, and textural features of the palm
- **Facial geometry**, obtained through two-dimensional (2D) or three-dimensional (3D) imaging techniques
- **Iris texture**, revealed using near-infrared illumination to capture fine and stable iris patterns
- **Retinal vascular patterns**, visualized through fundus imaging of the ocular blood vessel network
- **Hand vein patterns**, extracted using near-infrared transillumination to reveal subcutaneous vein structures

In addition to these external anatomical traits, emerging physiological biometrics exploit internal biological signals acquired through medical sensors and diagnostic instruments, such as:

- **Electrocardiography (ECG) signals**, reflecting individual-specific cardiac morphology and electrical conduction properties
- **Electroencephalography (EEG) signals**, capturing distinctive neural activity patterns
- **Photoplethysmography (PPG) waveforms**, encoding cardiovascular characteristics through optical measurements of blood volume variations

Physiological biometric traits typically exhibit high inter-subject distinctiveness and strong temporal stability, making them particularly suitable for accurate, reliable, and secure identity recognition systems.

1.2.2 Behavioral Biometrics

Behavioral biometrics capture patterns related to human actions and learned behaviors, reflecting the dynamic interaction between an individual and their environment. Prominent behavioral biometric modalities include:

- **Gait dynamics**, characterizing walking patterns through parameters such as stride length, cadence, joint motion, and body movement
- **Handwritten signature dynamics**, capturing temporal and spatial features including pen pressure, writing velocity, stroke sequence, and trajectory
- **Voice characteristics**, encompassing acoustic features such as fundamental frequency, formant structure, spectral energy distribution, and speaking rhythm
- **Keystroke dynamics**, analyzing typing behavior through metrics such as key press duration, inter-key latency, and typing rhythm

Behavioral biometrics enable non-intrusive, transparent, and continuous authentication, allowing identity verification to operate unobtrusively during normal user activities. However, compared to physiological traits, behavioral characteristics exhibit greater intra-subject variability. They may be influenced by emotional state, fatigue, physical injury, environmental conditions, or deliberate behavioral modification, leading to increased performance variability and higher error rates.

The evolution of identification systems illustrates the paradigm shift introduced by biometric technologies. Historically, identity verification relied on knowledge-based and token-based mechanisms. While these approaches are relatively simple to deploy and offer revocability, they are vulnerable to well-known threats, including guessing attacks, social engineering, phishing, theft, loss, duplication, and unauthorized sharing. Biometric authentication addresses these limitations by verifying "who" an individual is, rather than "what they know" or "what they possess." This intrinsic binding of identity to the individual has driven the widespread adoption of biometric technologies in security-sensitive domains, where both reliability and usability are critical.

Advances in sensing technologies, data acquisition systems, computational power, and machine learning algorithms have significantly expanded the scope of biometric applications beyond traditional security contexts. Modern biometric systems are now embedded in everyday consumer devices, such as smartphones equipped with fingerprint sensors and

facial recognition cameras, laptop computers with integrated biometric authentication, and wearable devices capable of user recognition through physiological signals such as ECG patterns or wrist vein structures. This widespread deployment reflects increasing societal acceptance and demonstrates the technological maturity of biometric systems.

Beyond authentication and access control, biometric measurement and analysis are increasingly applied in diverse domains, including healthcare for patient identification and monitoring, sports science for performance assessment, precision agriculture for livestock tracking, and biological and ecological research for individual-level identification. Within this broader landscape, medical biometrics has emerged as a specialized and rapidly growing subfield. Figure 1.1 illustrates different biometric modalities.

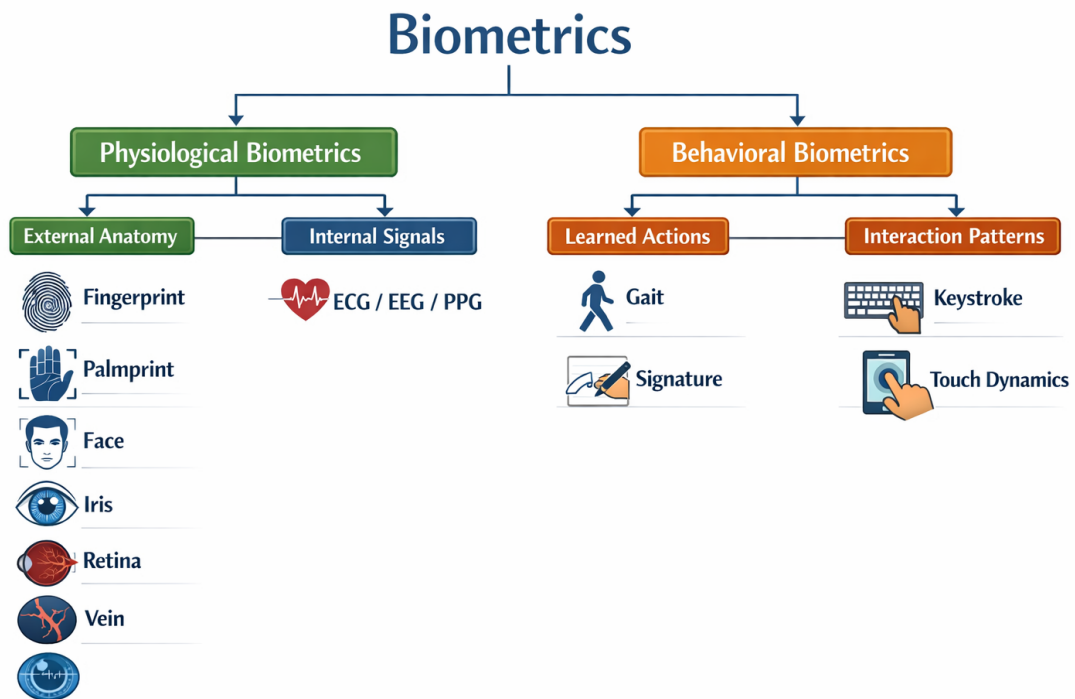


Figure 1.1. Hierarchical classification of biometric modalities

1.3 Biometric System Architecture and Processing Pipeline

A generic biometric system follows a well-defined processing pipeline that systematically transforms raw biometric data into identity decisions. Understanding this architecture is essential for analyzing system performance, identifying potential vulnerabilities, and designing improvements. The standard biometric pipeline comprises five fundamental stages.

1.3.1 Data Acquisition and Enrollment

The data acquisition stage captures raw biometric measurements using appropriate sensors or imaging devices. The quality of the acquired data plays a critical role, as it directly influences all subsequent processing stages and ultimately determines the overall recognition performance. Several factors affect acquisition quality, including sensor resolution and signal-to-noise ratio, which determine the level of detail captured; acquisition protocols and environmental conditions, which influence consistency and reproducibility; and subject cooperation and positioning, which affect the completeness and accuracy of the captured data.

During enrollment, one or more high-quality biometric samples are collected from an individual. These samples are then processed to generate a biometric template that serves as a reference for future recognition or verification. The enrollment process typically involves several steps, including quality assessment to ensure that captured samples meet minimum standards, acquisition of multiple samples to capture intra-subject variability, extraction of discriminative features, and secure storage of the resulting template in a database along with the associated identity information. The enrollment process establishes the baseline against which all future authentication attempts will be compared. Therefore, it is essential that enrollment data be of the highest possible quality and that the individual's identity be verified through reliable external procedures. Figure 1.2 illustrates the overall data acquisition and enrollment process.



Figure 1.2. Data acquisition and enrollment process

In addition, multi-sample enrollment, illustrated in Figure 1.3, involves capturing multiple biometric samples from the same individual to create a composite template or a set of templates. This strategy improves subsequent matching performance by better representing the natural variability in biometric presentations and reducing the risk of recognition errors.



Figure 1.3. Multi-sample enrollment strategy

1.3.2 Feature Extraction: From Handcrafted to Deep Representations

Feature extraction is a core computational component of biometric systems, as it determines the discriminative power, robustness, and efficiency of the resulting biometric templates. This stage transforms raw or preprocessed biometric data into compact representations that enable reliable and efficient matching.

Early biometric systems relied primarily on handcrafted features designed using domain-specific knowledge and expert intuition. These features are often interpretable and computationally efficient; however, they are typically sensitive to noise, acquisition variability, preprocessing or segmentation errors, and inter-subject differences. As a result, handcrafted approaches may struggle to capture complex or subtle discriminative patterns, limiting their generalization in real-world scenarios (Uhl et al., 2020).

Recent advances in machine learning, particularly deep learning, have fundamentally transformed feature extraction in biometric systems. Modern architectures such as convolutional neural networks (CNNs) and transformer-based models can automatically learn hierarchical and highly discriminative representations directly from raw or minimally processed data through supervised or self-supervised training (Dosovitskiy et al., 2020). These learned feature embeddings are generally more robust to noise, acquisition variability, and environmental changes, and they exhibit improved generalization across different sensors, devices, and populations.

Extensive experimental studies have demonstrated that deep-learned representations consistently outperform handcrafted features in terms of recognition accuracy, robustness, and scalability (Madenda et al., 2021; Jalilian et al., 2022). Consequently, deep feature extraction has become the dominant paradigm in modern biometric systems, particularly for complex, high-dimensional biometric modalities, where manual feature design is insufficient to capture the full discriminatory information in the data.

1.3.3 Template Storage and Security Considerations

After feature extraction, biometric templates are stored for future matching. Template storage and protection are critical aspects of biometric system security because biometric

characteristics are inherently non-revocable. Unlike passwords or cryptographic keys, biometric templates cannot be easily replaced if compromised, which exposes individuals to long-term security and privacy risks.

To address these challenges, biometric systems incorporate a range of protection mechanisms. These include strong encryption and strict access control policies for template databases, as well as the use of secure hardware platforms or trusted execution environments to protect templates during storage and matching. In addition, cancelable biometric schemes apply non-invertible transformations to biometric templates, enabling compromised templates to be revoked and replaced without exposing the original biometric data.

Privacy-preserving approaches have also gained increasing attention, particularly in sensitive application domains. Techniques such as secure computation and federated learning allow biometric models to be trained or matched without direct access to raw biometric data, thereby reducing privacy risks while preserving system utility (Yang et al., 2025). Collectively, these strategies aim to balance recognition performance with robust security and privacy protection, ensuring compliance with regulatory and ethical requirements.

1.3.4 Matching and Decision-Making

The matching stage compares a query biometric sample with stored templates to compute a similarity or distance score. The choice of the matching metric depends on the feature representation and the recognition task. Commonly used metrics include Euclidean distance, cosine similarity, correlation-based measures, and probabilistic models that account for variation within and between individuals.

In modern biometric systems, particularly those based on deep learning, matching is performed in a learned embedding space. In this space, samples from the same individual are mapped close to each other, while samples from different individuals are mapped farther apart. Metric learning methods, such as contrastive loss and triplet loss, are widely used to optimize this embedding space by minimizing intra-class variation and maximizing inter-class separation (Hadsell et al., 2006; Schroff et al., 2015). Siamese and related network architectures support end-to-end learning of similarity functions, reducing reliance on handcrafted features and improving robustness to acquisition variability, temporal changes, and differences between sensing devices (Agnihotri et al., 2019).

The decision-making stage interprets the matching scores using predefined rules or thresholds. In verification scenarios, a claimed identity is accepted if the score exceeds a selected threshold; otherwise, it is rejected. In identification scenarios, the system outputs a ranked list of candidate identities. The decision threshold determines the trade-off between the false acceptance rate (FAR) and the false rejection rate (FRR). It must be selected based

on the application's security and usability requirements.

1.4 Identification, Verification, and Authentication

Biometric systems operate mainly in two recognition modes: verification and identification. These modes differ in terms of computational complexity, error behavior, and application scenarios. Understanding these operating modes and their performance measures is essential for the proper design, evaluation, and deployment of biometric systems.

1.4.1 Verification (1:1 Comparison)

Verification, also known as one-to-one (1:1) matching, is used to confirm an individual's identity. During verification, a user provides a biometric sample along with a claimed identity, such as a user ID, username, or a smart card. The system compares the presented biometric sample with the stored template corresponding to the claimed identity and produces a similarity score. This score is then compared with a predefined threshold to generate a binary decision: accept or reject.

Verification is computationally efficient because only a single comparison is required, regardless of the database size. This makes it suitable for systems with a large number of enrolled users. Common applications of verification include unlocking personal devices, controlling access to secure buildings, authorizing financial transactions, and logging into computer systems or online services. Figure 1.4 illustrates the verification process.

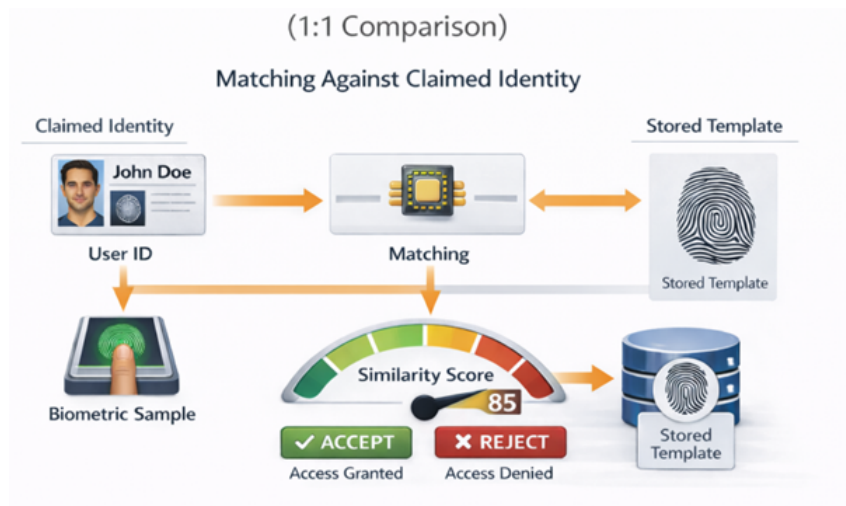


Figure 1.4. Conceptual illustration of the verification process

1.4.2 Identification (1:N Search)

Identification, also called one-to-many (1:N) recognition, aims to determine the identity of an unknown individual without requiring a claimed identity. In this mode, the biometric sample is compared against all enrolled templates in the database. The system returns the most similar match or declares that the individual is not enrolled.

Identification is computationally more demanding than verification because the query sample must be compared with all stored templates. As database sizes increase, processing time and computational cost rise, making efficient search and indexing techniques essential for large-scale systems.

Identification is commonly used in applications such as law enforcement investigations, border control and watchlist screening, missing person identification, and patient identification in healthcare environments. Figure 1.5 illustrates the identification process.



Figure 1.5. Conceptual illustration of the identification process

1.4.3 Authentication (Security Process)

Authentication is a higher-level security process that determines whether access to a system or resource should be granted or denied. Biometric authentication typically relies on a single factor and can be combined with additional methods, such as passwords or smart cards, in multi-factor authentication systems. These systems enhance security by requiring two or more independent factors, commonly categorized as "something you know" (e.g., a password or PIN), "something you have" (e.g., a smart card or token), and "something you are" (e.g., a biometric trait). For example, a secure system may require both biometric verification and a one-time password. In this context, biometric verification functions as a technical component within the broader authentication framework rather than as a separate recognition mode. Figure 1.6 illustrates the authentication process.



Figure 1.6. Conceptual illustration of the authentication process

1.5 Medical Biometrics: Motivation and Advantages

Medical biometrics is a specialized field of biometric recognition that utilizes internal physiological or anatomical features, typically captured during routine medical examinations or through clinical monitoring devices. Unlike conventional biometrics, such as fingerprints or facial recognition, which rely on externally observable traits, medical biometrics derives its identifiers from data acquired via specialized medical equipment, including imaging systems like retinal scanners, diagnostic tools like electrocardiogram (ECG) machines, and neurological monitors.

This fundamental distinction provides the core motivation for its development: to address critical and persistent challenges in healthcare, primarily the accurate and secure identification of patients. In an environment where misidentification can lead to severe medical errors, duplicate records, and administrative inefficiencies, medical biometrics offers a solution by creating a direct, intrinsic link between a patient and their medical identity.

The primary advantages of this approach are multifaceted, beginning with enhanced security and resistance to fraud. External biometric traits can be susceptible to replication through photographs, artificial molds, or other spoofing techniques. In contrast, internal features such as the pattern of retinal blood vessels or the unique waveform of an individual's ECG are exceptionally difficult to forge, as they require access to specialized medical technology. This inherent barrier significantly raises the security threshold for identity verification.

Furthermore, these medical biometric traits exhibit high uniqueness and long-term stability. For instance, the vascular structure of the retina forms during fetal development in a complex, random process that ensures distinctiveness even between identical twins.

Similarly, an ECG reflects the unique size, position, and electrical activity of a person's heart. Crucially, these internal patterns remain largely stable over decades, unlike faces or voices, which change with age, enabling reliable identification throughout a person's lifetime (Simon and Goldstein, 1935; Akram et al., 2020).

A particularly compelling advantage is the principle of "dual utility." Data collected for standard clinical purposes, such as a retinal image for an ophthalmological exam or an ECG trace for cardiac assessment, can simultaneously serve as a robust biometric identifier without necessitating additional procedures or effort from patients or clinical staff. This convergence streamlines workflows, reduces administrative burden, and enhances practical efficiency by allowing identity verification to occur seamlessly within the diagnostic process.

Concurrent technological and healthcare trends strongly support the growing feasibility of medical biometrics. The widespread digitization of health records creates a pressing need for secure patient matching to ensure data integrity. Simultaneously, advances in medical imaging and portable monitoring devices enable high-quality data acquisition across diverse settings. Perhaps most critically, the advent of sophisticated deep learning methods provides the necessary computational tools to extract meaningful, discriminative patterns from complex medical data, enabling accurate and scalable biometric recognition (Esteva et al., 2021).

However, implementing medical biometrics must carefully navigate significant privacy, ethical, and regulatory considerations. Medical data is inherently sensitive and protected by strict legal frameworks such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States (United States Congress, 1996; Annas, 2003; Cohen and Mello, 2018) and the General Data Protection Regulation (GDPR) in the European Union (European Parliament and Council of the European Union, 2016; Voigt and Von dem Bussche, 2017; Shabani and Borry, 2020). Consequently, any biometric system must be designed with robust safeguards, including strong encryption, secure storage protocols, and transparent mechanisms for patient consent, particularly if data is to be shared across institutions. Promising technological approaches, such as privacy-preserving federated learning, which allows model training on decentralized data without transferring the raw information, are emerging to help balance utility with confidentiality (Kumar et al., 2022; Arya et al., 2024; Rieke et al., 2020).

In conclusion, driven by modern artificial intelligence, medical biometrics presents a powerful paradigm. It moves beyond simple identification towards integrated systems that can enhance both the security of patient records and the delivery of healthcare itself, offering a synergistic solution to some of the most fundamental challenges in modern medicine.

1.6 Retinal Biometrics

Retinal biometrics refers to the use of retinal patterns for automatic human identification. Among physiological biometric modalities, retinal recognition is considered highly secure due to its internal location and the difficulty of reproducing its structural characteristics. The reliability of retinal biometrics arises from the uniqueness, permanence, and complexity of retinal features, particularly those derived from the vascular network. These properties enable retinal imaging to provide highly discriminative information for identity verification and authentication in high-security and medical applications.

1.6.1 Anatomical Basis of Retinal Biometrics

The retina is a highly specialized neural tissue located at the back of the eye that converts light into neural signals transmitted to the brain via the optic nerve. It consists of layered structures, including photoreceptors, intermediate neural cells, and ganglion cells, supported by a dense, intricate vascular network that meets the high metabolic demands of retinal tissue. This vascular system originates at the optic disc and forms a complex branching pattern across the retinal surface.

Retinal biometric recognition relies primarily on the uniqueness and stability of the retinal vascular network. Individual differences in vessel branching topology, diameter, tortuosity, and spatial distribution create highly distinctive patterns that can be captured through fundus imaging. Additional anatomical landmarks, such as the optic disc, fovea, macula, and nerve fiber layer, further enhance discriminative capability.

The key advantage of retinal biometrics is long-term stability, as the major vascular structure is established during early development and remains largely unchanged throughout life. Although certain pathologies, such as diabetic retinopathy, hypertensive retinopathy, and age-related macular degeneration, can locally alter retinal appearance, robust biometric systems can maintain performance by exploiting redundant and healthy regions of the vascular network. Figure 1.7 presents the anatomical structure of the human retina.

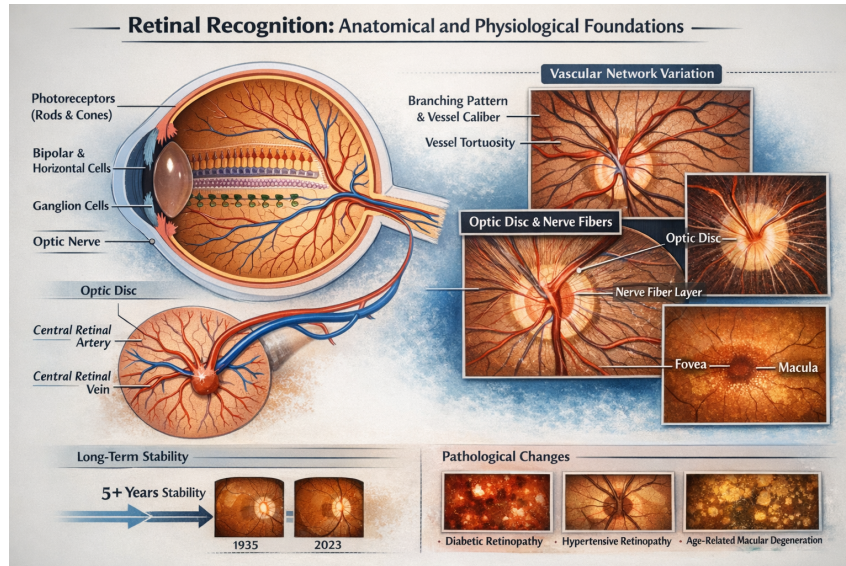


Figure 1.7. Anatomical structure of the human retina

1.6.2 Retinal Imaging Modalities

Retinal biometric systems rely on medical imaging technologies to capture the retina's anatomical and vascular structures. Fundus photography is the most commonly used modality, producing two-dimensional color images that clearly reveal blood vessels, the optic disc, the macula, and surrounding retinal tissue. It is non-invasive, cost-effective, and widely used in clinical practice, making it well-suited for large-scale biometric applications.

Modern digital fundus cameras typically provide a 30-50° field of view and sufficient resolution to visualize fine vascular structures. Non-mydratic fundus imaging, which does not require pharmacological pupil dilation, has further improved usability and user acceptance while maintaining image quality adequate for biometric recognition.

Beyond fundus photography, optical coherence tomography (OCT) represents a promising advanced imaging modality for retinal biometrics. OCT provides high-resolution cross-sectional and volumetric imaging of retinal layers via low-coherence interferometry, revealing detailed structural information, including retinal thickness variations, nerve fiber layer organization, and three-dimensional vascular architecture. While OCT devices are more expensive and less widely available than fundus cameras, the rich anatomical information they provide may enhance discriminative capability and offer robustness against certain surface-level pathologies (Akram et al., 2020).

Advanced modalities such as ultra-widefield fundus imaging and adaptive optics imaging provide extended retinal coverage and ultra-high-resolution visualization, respectively. While currently limited by cost and complexity, these techniques may further improve recognition performance by capturing additional vascular landmarks and fine anatomical features.

Retinal image quality can vary due to acquisition conditions, patient factors, and device differences, introducing intra-subject variability. Robust biometric systems address these challenges through quality control, data augmentation, and feature-extraction methods that focus on stable retinal regions. Overall, the richness and uniqueness of retinal vascular patterns provide high inter-subject discriminability, supporting the reliability of retinal biometrics. Figure 1.8 presents the concept of a retinal biometric system.

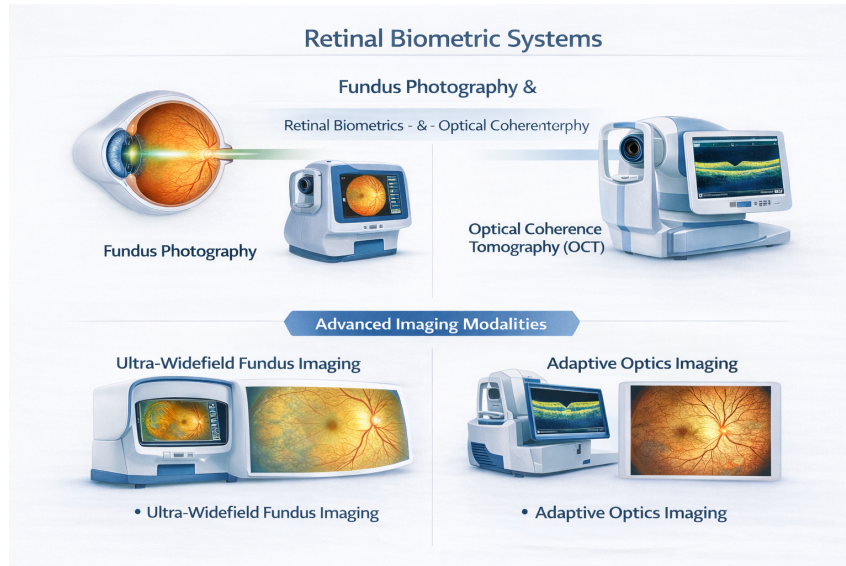


Figure 1.8. Concept of a retinal biometric system

1.6.3 Historical Development of Retinal Biometric Systems

The concept of using retinal vascular patterns for personal identification dates back to early medical examinations of the human eye. In 1935, Simon and Goldstein reported that the vascular structure of the retina exhibits substantial inter-individual variability while remaining remarkably stable over time (Simon and Goldstein, 1935). They suggested that the unique branching pattern of retinal blood vessels could be used to identify individuals. However, at that time, there were no suitable imaging devices or computers to make this practical.

In the decades that followed, retinal imaging remained largely confined to clinical diagnostics. Ophthalmologists relied on retinal examinations to detect and monitor pathological changes associated with systemic diseases such as hypertension and diabetes mellitus (Timberlake and Kennedy, 2005). Although advances in optical design, illumination techniques, and photographic sensors during the mid-twentieth century gradually improved image quality, retinal imaging continued to depend primarily on manual interpretation, and automated retinal-based identification remained largely theoretical.

The first practical retinal biometric systems emerged in the 1970s, notably through the pioneering work of Robert Buzz Hill. Hill developed one of the earliest operational retinal

identification prototypes, demonstrating that retinal vascular patterns could be captured and analyzed for reliable identity verification. This work ultimately led to the commercial deployment of retinal biometric systems by EyeDentify Inc. in the 1980s (Hill, 1986). These systems employed low-intensity infrared illumination to scan the retina and extracted discriminative features from the vascular network, particularly vessel bifurcation and crossover points.

EyeDentify systems achieved exceptionally low false acceptance rates, establishing retinal recognition as one of the most secure biometric modalities available at the time (Womack, 1994). Nevertheless, these early devices suffered from several practical limitations, including the requirement for precise eye positioning at very short working distances, mechanically complex scanning mechanisms involving rotating optical components, user discomfort due to perceived intrusiveness, high system cost, and limited acquisition throughput. Consequently, retinal biometrics remained largely restricted to high-security environments such as military installations, nuclear facilities, and other critical infrastructure sites (Kumar et al., 2021). In contrast, fingerprint and iris recognition systems gained broader commercial adoption due to their lower cost, faster acquisition, and improved user acceptance.

During the 1990s and early 2000s, research interest in retinal biometrics declined, while iris recognition emerged as a more practical compromise between security and usability. In parallel, however, medical retinal imaging underwent rapid technological development. The widespread adoption of digital fundus cameras, non-mydratic imaging techniques, and later wide-field and ultra-wide-field systems significantly enhanced image resolution, coverage, and acquisition speed (Kumar et al., 2021).

The convergence of these medical imaging advances with modern computing capabilities in the 2010s led to renewed interest in retinal biometrics. High-resolution CMOS sensors, compact infrared illumination sources, and increasingly powerful image processing pipelines enabled more robust and user-friendly retinal acquisition. At the same time, machine learning and deep learning techniques began to outperform traditional handcrafted feature extraction approaches, allowing automated detection of vascular landmarks and improved matching accuracy under variable imaging conditions (Azad et al., 2024).

More recently, retinal imaging has shifted toward automation and portability. Handheld and mobile fundus cameras, initially developed for teleophthalmology and large-scale screening programs, have demonstrated that high-quality retinal images can be acquired outside conventional clinical environments (Hogarty et al., 2020). These developments have significantly reduced the operational gap between medical-grade imaging systems and the requirements of biometric deployment.

Against this historical backdrop, modern platforms such as EYRINA represent a new generation of retinal imaging systems. Unlike earlier retinal biometric scanners, EYRINA

is designed for fully automatic acquisition, eliminating the need for trained operators and supporting non-mydriatic operation to enhance user comfort and acceptance. It features high-resolution visible and infrared imaging for both medical and biometric analysis, along with multi-axis mechanical alignment to ensure reliable capture under less constrained conditions (EYRINA Inc., 2023).

A critical modern advancement enabled by such platforms is the generation of large-scale, standardized retinal datasets. These data resources support the development of deep learning models and help overcome the limited database sizes that constrained earlier research. By integrating decades of progress in ophthalmic imaging with state-of-the-art automation and artificial intelligence, these systems represent a transition from highly constrained niche security applications toward scalable retinal biometric solutions suitable for real-world environments. Figure 1.9 illustrates a historical retinal imaging system used in the medical biometric domain.

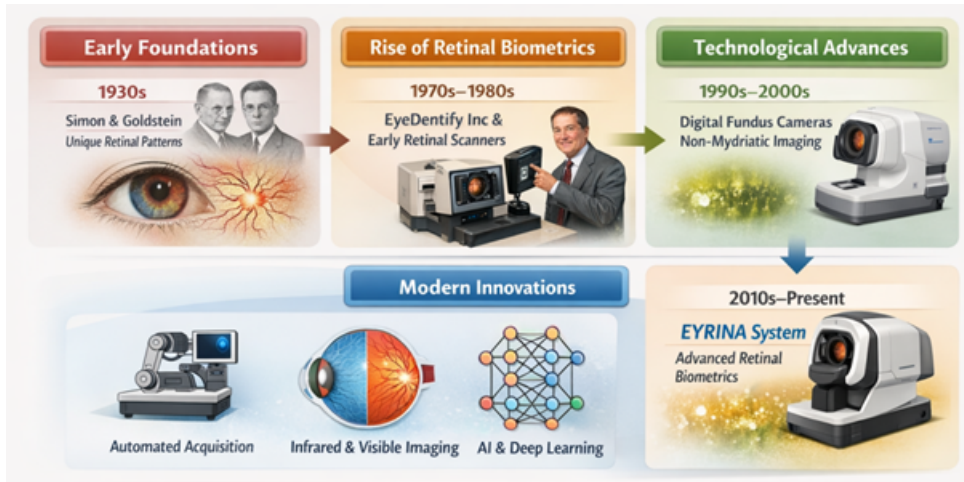


Figure 1.9. Historical retinal imaging system for medical biometric applications

1.7 Challenges in Clinical Retinal Biometric Deployment

Despite its strong discriminative capability and long-term stability, retinal biometrics faces several challenges that limit its large-scale clinical deployment. These challenges arise from biological variability, imaging conditions, device heterogeneity, data limitations, and operational constraints.

First, image quality plays a central role in recognition performance. Involuntary eye movements, illumination variability, reflections, and media opacities such as cataracts can degrade vascular visibility and reduce matching accuracy. Consequently, robust preprocessing strategies and quality-aware models are essential to ensure reliable feature

extraction under non-ideal acquisition conditions.

Second, pathological changes may influence biometric stability. While major vascular structures remain relatively stable in healthy individuals and in early disease stages, advanced conditions such as proliferative diabetic retinopathy or severe degeneration can alter vascular topology and compromise the reliability of recognition. For this reason, pathology-aware systems capable of adapting decision thresholds or updating templates are increasingly necessary.

Third, technical heterogeneity across imaging devices further complicates deployment. Differences in sensor resolution, field of view, spectral sensitivity, and proprietary processing pipelines can affect vessel appearance and reduce cross-device matching performance. Therefore, device-invariant feature learning and domain adaptation techniques are critical for achieving robust generalization.

Fourth, progress in this field is constrained by the scarcity of large, standardized, and demographically diverse retinal biometric datasets. Limited data availability restricts system validation and scalability. Ethical data-sharing frameworks and privacy-preserving learning strategies are thus essential to support broader development and clinical integration.

Finally, deep learning models for retinal analysis are highly sensitive to hyperparameter configurations. Learning rates, network depths, regularization strategies, and architectural choices significantly impact performance. However, systematic hyperparameter optimization remains underutilized, with many studies relying on manual or heuristic tuning that limits reproducibility and generalization across datasets.

Addressing these interconnected biological, technical, and operational constraints is fundamental for the safe, reliable, and scalable adoption of retinal biometric systems in healthcare environments.

1.8 State of the Art in Retinal Biometrics: A Synthetic Review

This section synthesizes the evolution of retinal biometric recognition research, tracing the transition from traditional image processing approaches to modern deep learning frameworks. Rather than providing an exhaustive list of studies, it identifies the key methodological shifts and remaining gaps that motivate this thesis.

1.8.1 From Handcrafted Features to Deep Representations

Early retinal recognition methods relied on traditional image processing techniques. Researchers employed multi-scale filtering, morphological operations, and matched filtering to enhance and extract vascular structures. Handcrafted descriptors such as Gabor filters, curvature measures, bifurcation mapping, and graph-based vascular models were then used to represent retinal patterns for matching (Madenda et al., 2021). While these methods achieved moderate accuracy under controlled conditions, their performance consistently degraded under variable illumination, noise, inter-device differences, and anatomical diversity. Machine learning classifiers improved performance incrementally but remained fundamentally constrained by the quality and discriminative power of the manually engineered features.

The advent of convolutional neural networks (CNNs) transformed retinal recognition by enabling hierarchical feature learning directly from raw images (Litjens et al., 2017). CNNs proved capable of capturing both local vascular details and global retinal structures, such as the optic disc and macula, effectively eliminating the need for handcrafted descriptors. Early studies by Zanlorensi et al. (2020) and Liu et al. (2022); Jalilian et al. (2022) demonstrated the feasibility of end-to-end deep learning for retinal biometric recognition, achieving significant improvements in accuracy and robustness over traditional approaches.

Subsequent research focused on architectural innovations to further enhance performance. Attention-enhanced CNNs were developed to focus on salient vascular regions while suppressing background noise (Agnihotri et al., 2019). Networks such as ResNet, DenseNet, and SENet addressed specific limitations, including vanishing gradients, limited feature reuse, and non-adaptive channel weighting (He et al., 2016; Huang et al., 2017; Hu et al., 2018). More recently, Vision Transformers (ViTs) and hybrid CNN-Transformer models have been explored for their ability to model long-range dependencies and capture global retinal context through self-attention mechanisms (Dosovitskiy et al., 2020; Li et al., 2026).

1.8.2 The Role of Segmentation in Retinal Biometrics

Retinal vessel segmentation emerged as a foundational preprocessing step in many biometric pipelines. Accurate segmentation enables the extraction of vascular topology, which can then be used for graph-based matching or template generation. U-Net and its variants became the standard architecture for this task due to their ability to combine contextual information with precise localization (Ronneberger et al., 2015). Subsequent improvements, including residual connections and attention mechanisms, further enhanced segmentation accuracy, particularly for thin and low-contrast vessels (Bhanuse et al., 2025).

However, a limitation of many existing approaches is the separation of segmentation from

the recognition task. Pipelines that first segment vessels and then extract features for matching cannot be optimized end-to-end for biometric performance. This decoupling may lead to suboptimal representations, as segmentation errors propagate to subsequent stages.

1.8.3 Metric Learning for Biometric Embeddings

A significant advancement in modern biometrics has been the adoption of metric learning to optimize embedding spaces. Rather than training networks for classification and using intermediate features, metric learning approaches directly optimize the feature space such that samples from the same individual cluster together while samples from different individuals are separated.

Contrastive loss (Hadsell et al., 2006) and triplet loss (Schroff et al., 2015) have been successfully applied to retinal biometrics, enabling more discriminative and robust representations (Agnihotri et al., 2019). Advanced strategies, including center loss, margin-based learning, and adaptive distance functions, have further enhanced performance (Liu et al., 2023). These approaches are particularly valuable for verification tasks, where the system must determine whether two samples belong to the same individual.

1.8.4 Addressing Data Scarcity: Self-Supervised and Semi-Supervised Learning

A persistent challenge in retinal biometric research is the scarcity of large, annotated datasets. Collecting retinal images from many subjects with multiple acquisitions per subject is logistically difficult and raises privacy concerns. To address this, researchers have increasingly turned to self-supervised and semi-supervised learning paradigms.

Self-supervised learning leverages proxy tasks such as rotation prediction, contrastive learning (SimCLR, MoCo), or masked image modeling to learn meaningful representations from unlabeled data (Chen et al., 2020; He et al., 2020; Gidaris et al., 2018). These representations can then be fine-tuned for biometric tasks with limited labeled data. Semi-supervised learning combines small labeled datasets with larger unlabeled collections, improving performance in clinically realistic scenarios where annotations are expensive to obtain (Zhang et al., 2024).

1.8.5 Domain Adaptation and Federated Learning for Clinical Deployment

Real-world deployment of retinal biometric systems must contend with domain shifts caused by different imaging devices, acquisition protocols, and patient populations. Domain adaptation techniques aim to learn representations that are invariant to these shifts,

enabling models trained on one dataset to generalize to others (Lyu et al., 2022).

Federated learning has emerged as a particularly promising approach for medical applications, enabling model training across multiple institutions without requiring the direct sharing of sensitive patient data (Yang et al., 2025). In the federated paradigm, models are trained locally at each institution, and only parameter updates are shared with a central server. This preserves privacy while enabling the development of more robust and generalizable models.

1.8.6 The Persistent Divide: Biometric Recognition vs. Pathology Classification

Despite these advances, a critical gap remains in the literature. Biometric recognition and retinal pathology classification have largely evolved as separate research tracks, with different datasets, evaluation protocols, and model architectures. This separation overlooks a fundamental clinical reality: both tasks depend on the same vascular substrate, and individuals requiring reliable identification are often those affected by progressive retinal diseases.

Only a limited number of studies have explored the intersection of these domains. Multitask learning frameworks that jointly optimize identity recognition and disease classification have shown promise in preliminary investigations (He et al., 2022; Zhao et al., 2023). By learning shared representations, these models can leverage complementary information: stable vascular patterns support identification, while deviations from these patterns may indicate pathology. However, systematic investigation of such unified architectures remains limited, particularly for multi-label pathology classification across diverse disease categories.

1.8.7 Hyperparameter Optimization: An Underexplored Dimension

Another underexplored dimension in retinal biometric research is systematic hyperparameter optimization. Deep learning models are highly sensitive to configuration choices, such as learning rates, network depths, regularization strategies, and architectural parameters, yet most studies select these through manual experimentation or heuristic approaches.

Grid search and random search remain common, despite their inefficiency in high-dimensional spaces. Bayesian optimization has been applied in some medical imaging contexts, but its use in retinal biometrics is limited. Evolutionary strategies, particularly CMA-ES, offer powerful alternatives capable of navigating complex, non-convex optimization landscapes, yet their application to retinal biometric model tuning remains largely unexplored. This gap is significant, as suboptimal hyperparameter choices can

substantially impact model performance and generalization across heterogeneous datasets.

1.9 Research Gaps and Thesis Positioning

The synthetic review presented in the previous section reveals several interconnected gaps in the current literature:

Gap 1: Lack of Robust, End-to-End Retinal Authentication Pipelines. While individual components have been studied extensively, their integration into a coherent pipeline optimized for verification performance remains underexplored. Many existing approaches separate segmentation from recognition, preventing end-to-end optimization and allowing segmentation errors to propagate.

Gap 2: Insufficient Attention to Hyperparameter Optimization. Despite the well-established sensitivity of deep learning models to hyperparameter configurations, systematic optimization remains rare in retinal biometric research. Manual tuning and basic search methods limit reproducibility and generalization across datasets, and advanced strategies, such as CMA-ES, have not been systematically evaluated in this domain.

Gap 3: Separation of Biometric and Diagnostic Objectives. The most fundamental gap is the historical division between biometric recognition and pathology classification research. This separation ignores the shared anatomical substrate of these tasks and the clinical reality that individuals requiring reliable identification are often those with progressive retinal diseases. Unified frameworks that jointly optimize both objectives remain scarce.

Gap 4: Limited Validation Across Heterogeneous Clinical Conditions. Many studies evaluate methods on single, curated datasets, providing limited insight into performance under real-world conditions with variable image quality, device heterogeneity, and diverse patient populations. Validation across multiple datasets with different characteristics is essential for establishing clinical applicability.

This thesis directly addresses these gaps through three complementary contributions:

1. **A robust retinal authentication pipeline** that integrates preprocessing, Residual Attention U-Net segmentation, Gaussian-weighted reconstruction, and multi-modal feature extraction, followed by machine learning-based verification. This pipeline is designed for end-to-end optimization and evaluated across multiple datasets.
2. **A systematic investigation of hyperparameter optimization strategies**, comparing grid search, random search, Bayesian optimization, and CMA-ES. This analysis provides practical guidance for training robust models and demonstrates the advantages of evolutionary optimization for retinal biometric tasks.

3. **A unified multitask learning framework** that performs simultaneous biometric identification and multi-label pathology classification. By leveraging shared representations, this architecture addresses the historical divide between these domains and demonstrates that both objectives can be jointly optimized within a single model.

Through these contributions, the thesis aims to advance the state of the art in retinal image analysis, demonstrating that biometric recognition and pathology classification are not mutually exclusive objectives but rather complementary tasks that can be jointly optimized to enhance both patient safety and clinical outcomes.

1.10 Chapter Summary

This chapter has established the theoretical and practical foundations of medical biometric systems, situating them within the broader context of identity recognition and healthcare applications. We began by defining biometrics and contrasting classical external modalities with emerging internal medical approaches that leverage physiological and anatomical signals acquired through medical sensing technologies. The advantages of medical biometrics, including intrinsic uniqueness, longitudinal stability, and robustness against spoofing attacks, were emphasized, particularly in the context of retinal imaging.

The operational principles of biometric systems were systematically presented, including data acquisition, feature extraction, template storage, matching, and decision-making. The evolution from traditional handcrafted features to deep learning-based hierarchical representations was emphasized, illustrating how modern AI techniques enhance robustness, scalability, and clinical relevance.

Retinal biometrics was presented as a central focus, with a detailed discussion of its anatomical basis, imaging modalities, and historical development. Key challenges, such as image quality variability, pathological changes, device heterogeneity, data scarcity, and hyperparameter sensitivity, were examined, demonstrating the need for robust, adaptable solutions.

A synthetic review of the literature traced the evolution of retinal biometric research and identified three critical gaps: the lack of integrated authentication pipelines, insufficient attention to hyperparameter optimization, and the persistent separation of biometric and diagnostic objectives. These gaps directly motivate the contributions of this thesis.

The following chapters build on this foundation, detailing the methodological framework, experimental evaluations, and key contributions of this research, with the overarching goal of advancing the state of the art in secure, clinically integrated medical biometric systems.

Chapter 2

Theoretical Foundations of Artificial Intelligence for Biometric Systems

2.1 Introduction

Artificial intelligence (AI) has become one of the most transformative technological paradigms in modern biomedical engineering, enabling the development of intelligent systems capable of extracting meaningful information from complex and high-dimensional medical data. In particular, machine learning and deep learning approaches have demonstrated remarkable success in visual recognition tasks, significantly surpassing traditional rule-based techniques in both accuracy and adaptability. These advances have led to a growing interest in applying AI methodologies to biometric systems, where the objective is to reliably authenticate or identify individuals based on unique physiological characteristics.

This chapter establishes the theoretical foundations necessary for AI methodologies developed throughout this thesis. It provides a structured overview of machine learning, deep neural architectures, segmentation theory, metric learning for authentication, transfer learning, and advanced optimization paradigms. By grounding the proposed contributions within a rigorous theoretical framework, this chapter ensures a coherent transition from fundamental AI principles to practical retinal biometric system development.

2.2 Fundamentals of Machine Learning for Biometrics

2.2.1 Definition of Machine Learning

Machine learning (ML) is a subfield of artificial intelligence that enables computational systems to learn patterns and relationships from data rather than relying solely on manually defined rules. Learning, in this context, refers to the ability of a system to improve its

performance on a specific task based on experience. Mitchell (1997) formally defines learning in computer science as follows:

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .”

In the context of retinal biometrics, learning corresponds to extracting structural descriptions from data that can be used to recognize individuals or detect pathological changes. These structural descriptions are formalized as predictive models that approximate the underlying relationships between retinal features and target labels.

A machine learning algorithm is any computational procedure that can improve its performance through experience. Learning algorithms can be viewed as optimization problems, where the goal is to find parameters θ_p that minimize a given objective function:

$$\theta_p = \arg \min_{\theta} [\mathcal{L}(\theta_p; X, Y) + \alpha_{reg} \mathcal{R}(\theta_p)] \quad (2.2.1)$$

Here, $\mathcal{L}(\theta_p; X, Y)$ is the loss function, quantifying the model’s error on the training data, $\mathcal{R}(\theta_p)$ is a regularization term controlling model complexity, and α_{reg} balances the trade-off between accuracy and overfitting.

Machine learning encompasses a wide range of algorithms for tasks including classification, regression, clustering, and pattern discovery. Three primary learning paradigms are commonly employed:

- **Supervised learning:** Learning from labeled examples (X, Y) .
- **Unsupervised learning:** Detecting patterns in unlabeled data.
- **Reinforcement learning:** Learning optimal policies through interaction with an environment using a reward signal.

2.2.2 Classical Machine Learning Algorithms for Retinal Biometrics

Classical machine learning algorithms provide foundational tools for biometric recognition and pathological detection in retinal images. Despite the rise of deep learning, these methods remain relevant, especially when datasets are limited or interpretability is critical. This section summarizes the key algorithms and their theoretical underpinnings.

Support Vector Machine (SVM)

Support Vector Machine (SVM), introduced by Cortes and Vapnik (1995), is a supervised machine learning algorithm widely used for classification and regression. The fundamental objective of SVM is to determine an optimal decision boundary that separates data points belonging to different classes. This boundary, called a hyperplane, is positioned to maximize the margin between samples from different classes, thereby improving the model's ability to generalize to unseen data.

In practical applications, many datasets are not perfectly separable by a simple linear boundary. To address this limitation, SVM employs kernel functions that transform the original data into a higher-dimensional space, where class separation becomes easier. Commonly used kernels include the linear, polynomial, and radial basis function (RBF) kernel.

k-Nearest Neighbors (k-NN)

The k-Nearest Neighbors (k-NN) algorithm (Cover and Hart, 1967) is a simple yet effective supervised machine learning method that classifies a new sample by examining the labels of the k most similar samples in the training dataset. The predicted class is typically determined through majority voting among these neighboring samples.

Similarity between samples is measured using a distance metric that quantifies how close two data points are in the feature space. Commonly used distance measures include Euclidean distance, cosine similarity, and Mahalanobis distance, each capturing different aspects of similarity depending on the data structure.

Decision Trees (DT)

Decision Trees, as formalized by Quinlan (1986), are supervised learning models that perform classification by recursively partitioning the feature space into homogeneous regions. Each internal node of the tree represents a decision based on a feature threshold, while each leaf node corresponds to a class label.

The construction of a decision tree involves selecting the feature and threshold that best splits the data at each node according to a predefined criterion, such as Gini impurity or information gain (Breiman et al., 1984). This process is repeated recursively until a stopping condition is reached, such as a maximum tree depth or a minimum number of samples per node.

Decision Trees are intuitive and easy to interpret, making them suitable for understanding decision boundaries in biometric verification tasks. However, they are prone to overfitting when grown to full depth. To mitigate this issue, constraints such as limiting the maximum

depth or minimum samples per leaf are typically applied.

Extra Trees (Extremely Randomized Trees)

Extra Trees, also known as Extremely Randomized Trees and introduced by Geurts et al. (2006), are an ensemble learning method that builds multiple decision trees using the entire training dataset, while introducing additional randomness during the splitting process. Unlike Random Forests, where splits are optimized based on criteria such as Gini impurity or information gain, Extra Trees select split thresholds randomly for each feature.

This increased randomness reduces variance and often improves generalization performance, particularly in high-dimensional feature spaces. During prediction, the outputs of all trees are aggregated using majority voting for classification tasks. Extra Trees are computationally efficient and provide robust performance when applied to complex biometric feature representations.

Random Forests

Random Forests, introduced by Breiman (2001), are ensemble learning methods that combine multiple decision trees to improve predictive performance and model stability. Each tree in the forest is trained using a bootstrap sample of the original dataset. During prediction, the outputs of individual trees are aggregated, typically through majority voting for classification tasks, to produce the final decision.

An important characteristic of Random Forests is the random selection of a subset of features at each splitting node during tree construction. This mechanism reduces the correlation among individual trees and increases ensemble diversity, thereby improving robustness to noise and reducing the risk of overfitting.

Gradient Boosting Machines

Gradient Boosting Machines (GBMs), introduced by Friedman (2001), are powerful ensemble learning techniques that construct predictive models by sequentially combining multiple weak learners, typically decision trees. In this approach, each new model is trained to correct errors made by previous models, thereby allowing the ensemble to improve its predictive performance gradually. By iteratively focusing on difficult or misclassified samples, gradient boosting can capture complex relationships within the data.

Several optimized implementations of gradient boosting have been widely adopted in machine learning applications. Notable examples include XGBoost developed by Chen and Guestrin (2016), as well as LightGBM introduced by Ke et al. (2017). These frameworks improve computational efficiency and scalability while maintaining high predictive accuracy.

Gradient boosting is particularly effective for structured retinal features, combining multiple feature types (vessel geometry, texture descriptors, and lesion statistics) to achieve high discrimination.

AdaBoost (Adaptive Boosting)

AdaBoost, proposed by Freund and Schapire (1997), is an ensemble learning technique that combines multiple weak learners, typically decision stumps, into a strong classifier. The algorithm operates iteratively by assigning higher weights to misclassified samples at each step, forcing subsequent learners to focus on difficult cases.

At each iteration, a weak classifier is trained, and its contribution to the final model is weighted according to its accuracy. The final prediction is obtained as a weighted combination of all weak learners. AdaBoost is particularly effective in improving classification performance when base learners are only slightly better than random guessing, and it is well-suited for structured feature representations extracted from retinal biometrics.

2.2.3 Artificial Neural Networks

Artificial Neural Networks (ANNs) are computational models inspired by the structure and functioning of biological neural systems. These networks consist of interconnected processing units, called neurons, organized into layers. A typical neural network includes an input layer that receives the raw data, one or more hidden layers that perform intermediate computations, and an output layer that produces the final prediction.

Each neuron processes input signals by combining them through weighted connections and applying a nonlinear transformation known as an activation function. Through this process, neural networks can learn complex relationships between inputs and outputs. During training, the network parameters are gradually adjusted via an optimization procedure that iteratively updates the model to minimize prediction errors on the training data. This optimization process is commonly performed using gradient-based algorithms such as the Adam optimizer.

Although early neural networks were relatively shallow, advances in computing power and data availability have enabled the development of much deeper architectures capable of learning highly complex representations.

ANNs provide a bridge to deep learning: they demonstrate that computational models can learn from data rather than rely exclusively on manually engineered features, setting the stage for architectures capable of handling high-dimensional image data, such as fundus photographs.

2.3 Deep Learning Foundations

Deep learning represents a major paradigm shift in machine learning, enabling computational models to learn hierarchical representations directly from raw data through multilayer neural architectures. Unlike traditional machine learning approaches that rely on handcrafted features, deep learning automatically learns relevant features during training. This capability has significantly improved performance in various computer vision tasks, including image classification, object detection, medical image analysis, and biometric recognition.

In the context of retinal biometric systems, deep learning plays a crucial role in several stages of the processing pipeline, including retinal vessel segmentation, feature extraction, biometric identity recognition, and disease detection. Deep neural networks are particularly effective in retinal imaging due to their ability to capture complex spatial patterns such as vascular structures, branching configurations, and pathological features.

The following subsections introduce the fundamental principles of artificial neural networks and convolutional neural networks, which form the backbone of most modern deep learning systems for medical imaging and biometric authentication.

2.3.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are specialized deep neural networks designed for image analysis. CNNs have become the dominant approach for computer vision tasks because they can efficiently capture spatial patterns in images.

Unlike traditional neural networks that process images as one-dimensional vectors, CNNs preserve the two-dimensional structure. This allows them to analyze spatial relationships between pixels and detect visual patterns such as edges, textures, and shapes.

A typical CNN architecture is composed of several types of layers, including:

- **Convolutional layers** are the primary building blocks of CNNs and are responsible for extracting hierarchical features from input images. Each convolutional layer applies a set of learnable filters that slide across the image to detect specific visual patterns, producing feature maps that highlight the presence and spatial location of these patterns. Early layers typically capture simple and low-level features such as edges, corners, or color gradients. As the network goes deeper, the layers combine these simple patterns to detect more complex and semantically meaningful structures. Feature maps provide a visual representation of what the network sees at each layer, revealing how the network progressively abstracts information from raw pixel values to high-level concepts. Figure 2.1 depicts one example of the convolution operations with a layer

of size $6 \times 6 \times 1$ and a $3 \times 3 \times 1$ filter in the top row. The modified image will have better feature representation than the original image. A CNN will typically have many convolution layers and each convolution layer will have its own set of learned kernels or filters (Salvaris et al., 2018).

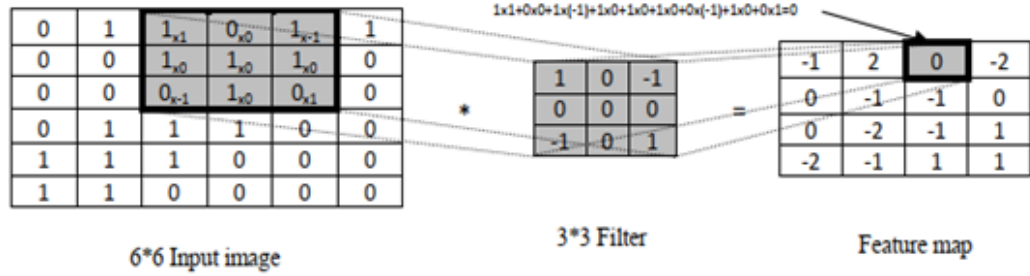


Figure 2.1. Convolution operation in CNNs. A filter (kernel) slides across the input image to produce a feature map.

- **Activation layers** introduce non-linearity into neural networks, allowing them to learn complex patterns that cannot be captured by linear combinations of inputs alone. Common activation functions include:

- **Sigmoid:** used for binary classification, it maps any real-valued input into the range $[0, 1]$, representing a probability.

$$\sigma(x)_{sig} = \frac{1}{1 + e^{-x}} \quad (2.3.1)$$

where x is the input to the neuron.

- **Softmax:** used for multi-class classification, generating a probability distribution over multiple classes. Mathematically, it is defined as:

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{C_{class}} e^{x_j}} \quad (2.3.2)$$

where x_i is the input score (logit) for class i , and C_{class} is the total number of classes.

- **Rectified Linear Unit (ReLU):** widely used in hidden layers, which sets all negative values to zero while keeping positive values unchanged, improving computational efficiency and mitigating the vanishing gradient problem.

$$\text{ReLU}(x) = \max(0, x) \quad (2.3.3)$$

Among these, the ReLU function has become the most widely used in deep learning due to its computational efficiency and ability to mitigate the vanishing gradient problem.

- **Pooling layers** reduce the spatial dimensions of feature maps. By downsampling these maps, pooling lowers computational and memory requirements while improving the network's robustness to small translations, rotations, or distortions in the input images. The two most commonly used pooling operations are:
 - **Max pooling:** selects the maximum value within a specified window, emphasizing the most salient features.
 - **Average pooling:** computes the mean value within the window, providing a smoother representation of the features.

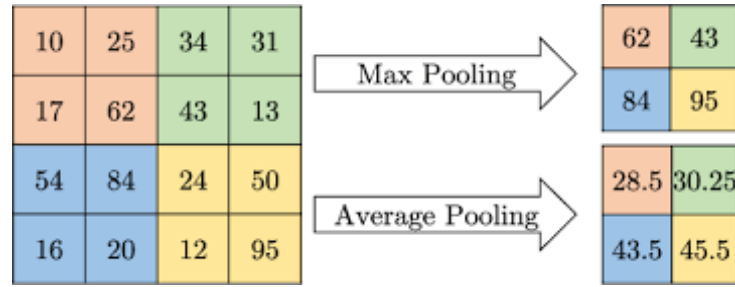


Figure 2.2. Example of max pooling and average pooling operations with a 2×2 filter and stride 2.

- **Fully connected layers** are typically located near the end of the network and perform high-level reasoning based on the features extracted by previous layers. These layers combine the learned features to produce the final classification or prediction.

CNNs are particularly effective for retinal image analysis because they can automatically detect vascular structures and anatomical patterns that are difficult to describe using handcrafted features. As a result, CNN-based approaches have significantly improved the performance of retinal disease detection and biometric identification systems.

2.3.2 Parameters of Deep Learning

Deep learning models are characterized by two types of parameters: learnable parameters, which are adjusted automatically during training, and hyperparameters, which are predefined before training and guide the learning process. Both categories play a crucial role in determining the performance, generalization, and efficiency of deep neural networks.

Learnable Parameters

- **Weights:** These are coefficients associated with the connections between neurons. Weights determine the influence of inputs on each neuron's output and are iteratively adjusted during training to minimize the loss function.

- **Biases:** Bias terms allow neurons to activate independently of their input, providing flexibility in learning and enabling the network to model complex functions more effectively.

Hyperparameters

Hyperparameters are set before training and control various aspects of the learning process and network architecture:

- **Learning rate:** Specifies the magnitude of updates applied to learnable parameters during optimization.
- **Number of layers and neurons per layer:** Determines the depth and capacity of the network, influencing its ability to capture complex patterns.
- **Batch size:** Defines the number of samples processed before updating the network parameters that impact training stability and convergence speed.
- **Number of epochs:** Indicates the total number of complete passes through the training dataset.
- **Dropout rate:** Sets the probability of randomly deactivating neurons during training, helping to prevent overfitting.
- **Convolution-specific parameters:** Include kernel number and size, stride, and padding, which controls how features are extracted from input images.

2.3.3 Regularization Techniques in Deep Learning

Deep neural networks often contain millions of parameters, which enables them to learn highly complex representations from large datasets. However, this high capacity also increases the risk of overfitting, a phenomenon in which the model learns the training data too closely and fails to generalize well to unseen data. Regularization techniques are therefore introduced to improve the generalization ability of deep learning models by preventing the network from memorizing the training data. Several regularization strategies are commonly used in deep learning systems for medical image analysis and biometric recognition.

- **Dropout:** randomly turns off a subset of neurons in a neural network with a predefined probability. As a result, the network cannot rely on specific neurons and must learn more distributed, robust feature representations. This process effectively creates multiple sub-networks during training, which are implicitly averaged during inference. By reducing the co-adaptation of neurons, dropout helps prevent overfitting and improves the robustness of deep learning models.

- **Batch Normalization:** is another important regularization and training stabilization technique. The main objective of batch normalization is to normalize the inputs of each layer so that they maintain a stable distribution during training.
- **Data Augmentation:** is a regularization strategy that artificially increases the size and diversity of the training dataset by applying various transformations to the original images. Instead of collecting additional data, new training samples are generated by modifying existing images through operations such as rotation, flipping, scaling, cropping, or intensity adjustments.
- **Early Stopping:** is used to prevent overfitting by monitoring the model's performance on a validation dataset during training. Instead of training the model for a fixed number of epochs, the training process is terminated when the validation performance stops improving.

2.3.4 Training of Deep Neural Networks

Training deep neural networks is the process through which the model learns to map input data to the correct outputs by adjusting its internal parameters. During training, the network receives input samples and produces predictions through a sequence of layers in a process known as forward propagation. The predicted output is then compared with the true label using a loss function, which measures the difference between the predicted and actual values.

To minimize this error, the network uses the backpropagation algorithm, which computes the gradients of the loss function with respect to the model parameters (weights and biases). These gradients indicate how each parameter should be adjusted to reduce the prediction error. An optimization algorithm, such as stochastic gradient descent (SGD) or Adam, updates the parameters iteratively based on computed gradients and a predefined learning rate.

Training is performed over multiple iterations, called epochs, during which the model processes the entire training dataset several times. The data are typically divided into smaller groups, called batches, to enable efficient computation and more stable parameter updates. Throughout the training process, the network gradually improves its ability to extract relevant features and make accurate predictions.

Proper training of deep neural networks also involves techniques to improve generalization and prevent overfitting, such as regularization methods, dropout, and data augmentation. These techniques ensure that the trained model performs well not only on the training data but also on unseen data. In medical imaging applications, including retinal image analysis, careful training procedures are essential for obtaining reliable, robust models

that can detect subtle anatomical patterns and pathological features.

2.4 Transfer Learning

Transfer learning is a powerful approach in deep learning that enables knowledge acquired from one task to be reused for a related task. Instead of training deep neural networks from scratch, transfer learning allows pre-trained models to serve as feature extractors or initializations for new tasks. This strategy is particularly advantageous in medical imaging, where labeled data are often limited.

In retinal image analysis, transfer learning is widely used to accelerate training, improve model generalization, and mitigate overfitting. Pre-trained convolutional neural networks, typically trained on large-scale natural image datasets such as ImageNet, have learned hierarchical feature representations. Early layers capture low-level features like edges, textures, and color gradients, while deeper layers encode complex, high-level patterns. By adapting these representations to retinal images, networks can effectively detect intricate vascular structures, lesions, and other anatomical features, even with relatively small datasets.

Transfer learning can be implemented in several ways:

- **Feature extraction:** The pre-trained network is used as a fixed feature extractor, and only the task-specific layers (e.g., fully connected classification heads) are trained on retinal images.
- **Fine-tuning:** Some or all layers of the pre-trained network are updated using retinal data, allowing the model to adapt learned features to the specific domain.
- **Hybrid strategies:** Early layers of the network are frozen to preserve general visual features, while deeper layers are fine-tuned to capture domain-specific retinal patterns.

Several well-established deep learning architectures are commonly used for transfer learning in medical image analysis, such as ResNet, DenseNet, and EfficientNet.

2.4.1 ResNet Architecture

Residual Networks (ResNet), introduced by He et al. (2016), were proposed to address the degradation problem observed during the training of very deep neural networks. As network depth increases, accuracy may saturate and then degrade, not necessarily due to overfitting but rather because of optimization difficulties. To overcome this limitation, ResNet introduces residual (shortcut) connections that enable direct information and gradient flow across layers during backpropagation.

Instead of directly approximating a desired mapping $\mathcal{H}(x)$, ResNet reformulates the learning objective to estimate a residual function:

$$\mathcal{F}(x) = \mathcal{H}(x) - x \quad (2.4.1)$$

Thus, the block output becomes:

$$y = \mathcal{F}(x) + x \quad (2.4.2)$$

This identity mapping facilitates gradient propagation, alleviates vanishing gradient issues, and enables the successful training of extremely deep architectures such as ResNet-50, ResNet-101, and ResNet-152. Due to their strong representational capacity and stable convergence behavior, ResNet models have become widely adopted in computer vision and medical image analysis applications.

In our work, we employ ResNet-152 as the backbone network for feature extraction from retinal images. The architecture consists of an initial convolutional stem, followed by four hierarchical residual stages composed of bottleneck blocks, and concludes with a global pooling layer and a fully connected classifier. The network contains 152 learnable layers and approximately 60 million parameters.

Input Layer and Convolutional Stem

The network accepts an RGB input image with a spatial resolution of $224 \times 224 \times 3$. The initial processing stage, referred to as the convolutional stem, performs early feature extraction and spatial downsampling. It consists of:

- A 7×7 convolutional layer with 64 filters and stride 2.
- Batch normalization.
- Rectified Linear Unit (ReLU) activation.
- A 3×3 max-pooling layer with stride 2.

This configuration reduces the spatial resolution from 224×224 to 56×56 , while increasing channel depth to 64. The resulting feature tensor has dimensions $56 \times 56 \times 64$. The stem captures low-level visual primitives such as edges, contours, gradients, and basic textures, forming the foundation for subsequent hierarchical feature extraction.

Residual Learning Framework

The fundamental building block of ResNet-152 is the bottleneck residual block, designed to maintain computational efficiency while enabling deep representations. Each bottleneck block comprises three convolutional layers arranged sequentially: a 1×1 convolution for channel reduction, a 3×3 convolution for spatial feature extraction, and a 1×1 convolution for channel expansion.

Each convolutional layer is followed by batch normalization and ReLU activation, except for the final convolution preceding the residual addition. The shortcut connection either performs identity mapping or uses a 1×1 projection when the spatial or channel dimensions change.

By learning residual mappings rather than direct transformations, the architecture significantly improves training stability and enables effective optimization at extreme depths.

Hierarchical Residual Stages

The residual blocks are organized into four main stages: conv2_x, conv3_x, conv4_x, and conv5_x. Across these stages, spatial resolution decreases while feature dimensionality increases, producing progressively abstract representations.

- **Stage conv2_x:** Contains 3 bottleneck blocks. The spatial resolution remains 56×56 , and the output dimensionality expands to 256 channels. This stage refines low-level structural representations.
- **Stage conv3_x:** Includes 8 bottleneck blocks. The first block downsamples with a stride of 2, reducing the spatial resolution to 28×28 . The number of channels increases to 512. This stage captures more complex spatial configurations and intermediate-level patterns.
- **Stage conv4_x:** The deepest stage, comprising 36 bottleneck blocks, which accounts for the majority of the network depth. The spatial resolution is reduced to 14×14 , and channel dimensionality increases to 1024. This stage is responsible for learning high-level semantic representations and complex structural features.
- **Stage conv5_x:** Contains 3 bottleneck blocks. Downsampling in the first block reduces the feature maps to 7×7 with 2048 channels. At this level, the network encodes highly abstract and globally discriminative features suitable for classification.

Global Average Pooling

Following the final residual stage, the feature tensor of size $7 \times 7 \times 2048$ is processed through global average pooling, which computes the spatial mean across each channel.

This produces a compact representation of $1 \times 1 \times 2048$, equivalent to a 2048-dimensional feature vector.

Global average pooling reduces parameter count, mitigates overfitting, and enforces correspondence between feature maps and semantic categories.

Fully Connected Classification Layer

In the original ImageNet classification configuration, the pooled feature vector is fed into a fully connected layer with 1000 neurons, corresponding to the 1000 object categories in the ImageNet dataset. A Softmax activation function generates normalized class probabilities.

The architecture follows a pyramidal design, progressively decreasing spatial resolution from 224×224 to 7×7 , while increasing channel dimensionality from 64 to 2048. This structural progression enables the transition from local feature extraction to high-level semantic abstraction.

Despite its extreme depth, ResNet-152 remains trainable due to residual connections that facilitate efficient gradient flow. However, it is computationally intensive and memory demanding compared to shallower variants such as ResNet-50 and ResNet-101. The bottleneck structure significantly reduces computational complexity by compressing and expanding channel dimensions around the central 3×3 convolution, thereby maintaining efficiency while preserving representational power. Figure 2.3 summarizes the architecture diagram.

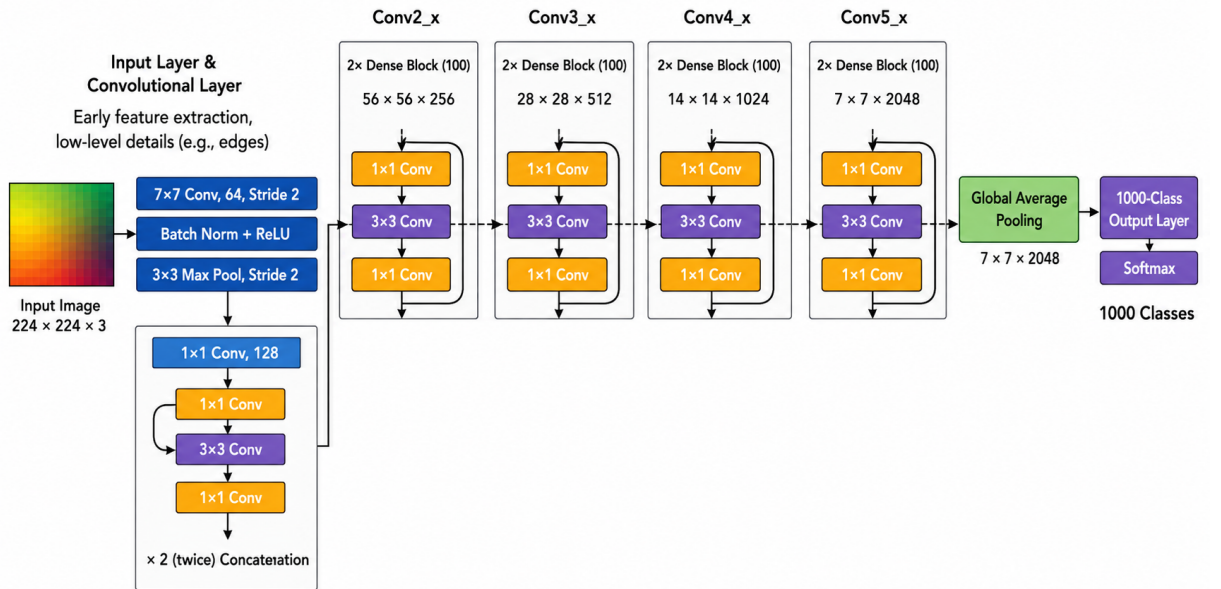


Figure 2.3. Architectural diagram of ResNet-152, illustrating the convolutional stem, four residual stages, and the final classification layers.

2.4.2 DenseNet Architecture

Densely Connected Convolutional Networks (DenseNet), introduced by Huang et al. (2017), were designed to improve information flow and gradient propagation in very deep networks. Unlike traditional convolutional architectures where each layer feeds its output only to the next layer, DenseNet establishes direct connections from each layer to all subsequent layers within the same dense block. This dense connectivity alleviates vanishing gradient problems, encourages feature reuse, strengthens feature propagation, and substantially reduces the number of parameters compared to conventional deep networks.

Formally, for a dense block with ℓ layers, each layer receives as input the concatenation of feature maps produced by all preceding layers:

$$x_\ell = H_\ell([x_0, x_1, \dots, x_{\ell-1}]) \quad (2.4.3)$$

where $[\cdot]$ denotes concatenation and H_ℓ represents a composite function consisting of Batch Normalization (BN), ReLU activation, and convolution. This design ensures maximum information flow between layers and promotes implicit deep supervision.

In our work, we employ DenseNet-169 and DenseNet-201 as backbone architectures for retinal image feature extraction. Both models follow the same structural principles but differ in depth and number of layers per dense block.

Input Layer and Convolutional Stem

The network accepts an RGB image of size $224 \times 224 \times 3$. The initial convolutional stem is similar to that of ResNet and consists of:

- A 7×7 convolution with 64 filters and stride 2.
- Batch Normalization.
- ReLU activation.
- 3×3 max pooling with stride 2.

This stage reduces spatial resolution from 224×224 to 56×56 , producing an initial feature tensor of size $56 \times 56 \times 64$. The stem extracts low-level visual features such as edges, contours, and texture primitives.

Dense Learning Framework

The fundamental building unit of DenseNet is the dense layer, which employs a bottleneck design consisting of a 1×1 convolution for channel reduction followed by a 3×3 convolution

for spatial feature extraction. Each dense layer receives as input the concatenation of all preceding feature maps within the same dense block, and if the growth rate is denoted by k , each layer produces k new feature maps. Consequently, the number of channels increases linearly within each block. This dense connectivity strategy provides several advantages: it improves gradient flow, mitigating vanishing gradient problems in very deep networks; it promotes strong feature reuse by allowing layers to access all previously learned features; it reduces redundancy in feature representations; and it enhances parameter efficiency. Unlike ResNet, which performs feature summation through residual addition, DenseNet uses feature concatenation, thereby preserving information from all previous layers and supporting richer hierarchical feature learning.

Transition Layers

Between dense blocks, DenseNet introduces transition layers to control model complexity and reduce spatial dimensions. Each transition layer consists of a 1×1 convolution and 2×2 average pooling with stride 2. These layers reduce both the number of feature maps and spatial resolution, ensuring computational efficiency.

DenseNet-169 Architecture

DenseNet-169 comprises 169 learnable layers and approximately 14 million parameters, striking a balance between network depth and computational efficiency. The architecture is structured into four dense blocks, containing 6, 12, 32, and 32 layers, respectively. Spatial resolution follows a pyramidal progression, decreasing from 56×56 after the first block to 28×28 , 14×14 , and finally 7×7 after the fourth block. Due to the dense concatenation of feature maps, channel dimensionality increases progressively throughout the network, enabling the extraction of rich hierarchical features. This design allows DenseNet-169 to capture detailed representations while remaining computationally tractable, which is particularly advantageous for high-resolution retinal image analysis.

DenseNet-201 Architecture

DenseNet-201 builds upon the same architectural principles as DenseNet-169 but extends the network to 201 learnable layers, with approximately 20 million parameters. Its four dense blocks are composed of 6, 12, 48, and 32 layers, respectively, with the third dense block substantially deeper than in DenseNet-169. This increased depth enables richer hierarchical feature extraction and stronger high-level representation learning. The spatial resolution follows a consistent pyramidal reduction from 224×224 to 56×56 , 28×28 , 14×14 , and finally 7×7 . Through dense concatenation, channel dimensionality progressively increases, allowing the network to capture highly detailed and discriminative features while

maintaining parameter efficiency relative to other networks of comparable depth, making DenseNet-201 particularly suitable for complex retinal image analysis tasks.

Global Average Pooling and Classification Layer

After the final dense block, the resulting feature tensor of size $7 \times 7 \times C_{chan}$ where C_{chan} denotes the number of feature channels (1664 for DenseNet-169 and 1920 for DenseNet-201) undergoes global average pooling. This operation reduces each 7×7 feature map to a single value, producing a compact feature vector of length equal to the number of channels. In the original ImageNet configuration, this feature vector is subsequently connected to a fully connected layer with 1000 neurons, followed by a Softmax activation for multi-class classification. By aggregating spatial information via global average pooling, the network captures the most salient features while reducing dimensionality, thereby enabling efficient, discriminative representations for downstream tasks.

In retinal image analysis and medical imaging tasks, DenseNet backbones demonstrate strong performance due to their ability to preserve fine-grained structural information across layers, which is critical for vascular pattern recognition and lesion detection. Figure 2.4 summarizes the architectural diagrams of DenseNet-169 and DenseNet-201.

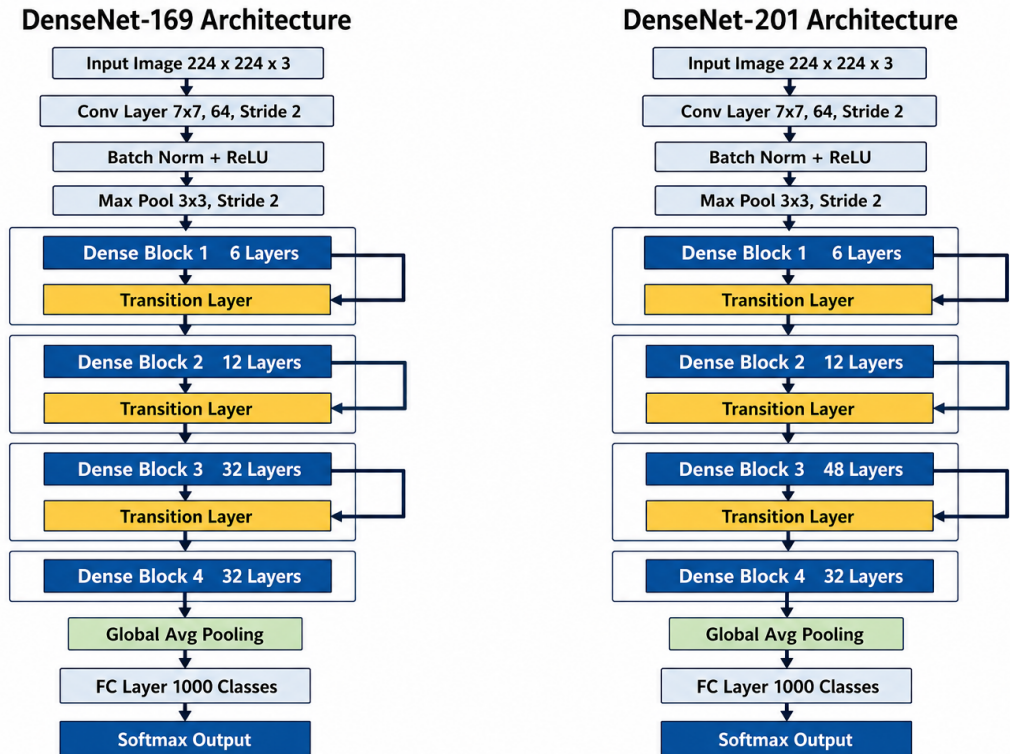


Figure 2.4. Architectural diagrams of DenseNet-169 and DenseNet-201 backbones, showing the progression of dense blocks and transition layers.

2.5 Retinal Image Segmentation Theory

Segmentation is a fundamental step in retinal biometric systems, aiming to isolate meaningful anatomical structures from the background of retinal images. Accurate segmentation of blood vessels, optic discs, fovea regions, and pathological lesions provides structural representations that are essential for both identity recognition and disease detection. By converting raw retinal images into structured maps, segmentation reduces data complexity while preserving discriminative features necessary for subsequent analysis.

The primary goal of retinal segmentation in biometric systems is to extract unique vascular patterns and individual-specific anatomical landmarks. These features include:

- **Vascular morphology:** Branching patterns, lengths, widths, and bifurcation angles, which exhibit high inter-subject variability.
- **Optic disc and fovea localization:** Serve as spatial reference points to normalize vascular structures across images.
- **Lesion detection:** Microaneurysms, exudates, and hemorrhages, while often indicative of pathology, can also contribute to distinguishing retinal images over time.

High-quality segmentation ensures that extracted features are robust, discriminative, and invariant to imaging conditions, making them suitable for biometric authentication and longitudinal monitoring.

2.5.1 U-Net Architecture for Retinal Segmentation

The U-Net architecture, introduced by Ronneberger et al. (2015), has become the standard approach for medical image segmentation, including retinal vessel extraction, due to its ability to integrate contextual understanding with precise localization. Its structure consists of two main paths: an encoder and a decoder.

- **Encoder:** is designed to extract hierarchical features from the input retinal image progressively. It is composed of sequential convolutional blocks, each typically consisting of two convolutional layers followed by Rectified Linear Unit (ReLU) activations and batch normalization. These operations enable the network to learn increasingly complex representations while maintaining training stability. Max pooling layers are interleaved between convolutional blocks to reduce the spatial resolution of feature maps, thereby allowing the encoder to capture broader contextual information and larger receptive fields. This hierarchical feature extraction is essential for accurately modeling the intricate vascular structures present in retinal images, as it provides both fine-grained local details and global structural context necessary for reliable segmentation.

- **Decoder:** is responsible for reconstructing the segmentation map from the abstract feature representations produced by the encoder. It consists of upsampling layers, typically implemented via transposed convolutions or interpolation, followed by convolutional blocks that refine the upsampled features. This process progressively restores the spatial resolution while preserving the semantic information captured in the encoder. Crucially, skip connections link corresponding layers of the encoder and decoder, allowing the network to combine high-resolution spatial information with deep contextual features. These connections mitigate information loss caused by downsampling and enable precise localization of retinal structures, such as blood vessels and lesions. By integrating both global context and fine-grained details, the encoder-decoder architecture ensures accurate and robust segmentation performance, particularly in complex retinal images with intricate vascular patterns.
- **Output Layer:** serves as the final stage of the segmentation network, translating the learned feature representations into pixel-wise class predictions. A 1×1 convolution is employed to reduce the multi-channel feature maps produced by the decoder to the desired number of classes, ensuring that each pixel is assigned a probability distribution over possible labels. For binary segmentation tasks, such as distinguishing retinal vessels from the background, a Sigmoid activation function is applied to produce values between 0 and 1, representing the likelihood of each pixel belonging to the target class. In multi-class segmentation scenarios, such as simultaneously delineating vessels, optic disc, and pathological lesions, a Softmax activation function is used to generate a normalized probability distribution across all classes for each pixel. This design enables the network to make precise, per-pixel predictions while accommodating both simple and complex segmentation objectives, ensuring accurate delineation of retinal structures critical for biometric and diagnostic applications.

Several variants of the classical U-Net architecture have been proposed to enhance segmentation performance, particularly in challenging retinal imaging tasks. Residual U-Net integrates shortcut connections within convolutional blocks, facilitating improved gradient propagation and enabling deeper network construction. This design is particularly effective at detecting thin, low-contrast vessels that are otherwise difficult to segment. Attention U-Net incorporates attention mechanisms that selectively emphasize relevant retinal regions while suppressing background noise, thereby improving segmentation accuracy in images with complex structures or pathological artifacts. The multi-scale U-Net employs parallel processing across multiple spatial resolutions, allowing the network to simultaneously capture fine vessel details and the global topology of the vascular network. Collectively, these variants address limitations of the standard U-Net and provide more robust and precise segmentation of retinal structures, which is crucial for both biometric recognition and medical diagnosis.

2.6 Attention Mechanisms: Theoretical Foundations

Attention mechanisms are a fundamental concept in modern deep learning, inspired by human cognitive processes that allow selective focus on relevant information while ignoring irrelevant stimuli. In neural networks, attention provides a dynamic means of weighting features, enabling the model to prioritize informative regions or channels of the input data. This concept is especially valuable in high-dimensional data such as images, where critical information may be sparsely distributed.

Formally, consider an input feature representation F . An attention mechanism computes a weighting map A , which can be spatial, channel-wise, or both, and generates a refined representation F' as:

$$F' = F \odot A \quad (2.6.1)$$

where $\odot$ denotes element-wise multiplication. The attention map A is learned during training, allowing the network to adaptively emphasize features that are most relevant for the task at hand.

Attention can be categorized into several types:

- **Spatial Attention:** Focuses on "where" important features are located in the input space. It assigns higher weights to critical spatial regions, facilitating the identification of localized structures in complex patterns.
- **Channel Attention:** Focuses on "what" features are most relevant across different channels or feature dimensions. By modulating the contribution of each channel, the network emphasizes the most discriminative feature maps.
- **Self-Attention:** Computes relationships between all elements of the input, capturing long-range dependencies and contextual interactions. This is widely used in transformer architectures and increasingly applied in vision tasks.

Attention mechanisms offer several theoretical advantages in deep learning:

- **Dynamic Feature Selection:** Allows the network to focus on salient information without hard-coded rules.
- **Improved Representational Power:** Facilitates the learning of complex, context-dependent features.
- **Enhanced Gradient Flow:** By emphasizing important features, attention can improve training stability and convergence.

- **Task Adaptivity:** Particularly in multi-task models, attention enables the network to allocate computational focus differently for each objective, such as segmentation, classification, or detection.

In summary, attention mechanisms provide a mathematically grounded, flexible framework for selective information processing, which has become a cornerstone of modern deep learning, particularly in vision and biomedical applications.

2.7 Multitask Learning

Multitask Learning (MTL) is a machine learning paradigm in which a single model is trained to perform multiple related tasks simultaneously. Instead of developing separate models for each task, multitask learning enables the network to share representations across tasks, allowing it to learn more general and informative features.

The central idea behind multitask learning is that related tasks often share underlying structures or patterns. By learning these tasks jointly, the model benefits from additional supervision signals, which improves generalization and reduces the risk of overfitting. This is particularly useful in medical and biometric image analysis, where anatomical and pathological characteristics may overlap.

Formally, assume that a model is trained to solve T tasks with corresponding loss functions $\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_T$. The overall objective function is defined as:

$$\mathcal{L}_{total} = \sum_{i=1}^T \lambda_i \mathcal{L}_i \quad (2.7.1)$$

where λ_i represents the weighting factor assigned to each task. These coefficients control the relative importance of the tasks during training.

2.7.1 Hard and Soft Parameter Sharing

Multitask learning architectures are generally categorized into two main strategies: hard parameter sharing and soft parameter sharing. These strategies define how information is shared between different tasks within a unified model.

Hard Parameter Sharing

Hard parameter sharing is the most commonly used approach in deep multitask learning. In this strategy, the initial layers of the neural network are shared among all tasks, forming a common feature extraction backbone. The later layers are task-specific and produce

independent outputs for each objective.

The shared layers learn a general representation of the input data that captures common structures and patterns relevant to all tasks. For example, in image analysis, early convolutional layers typically extract low-level features such as edges, textures, and shapes. These features are often useful across multiple tasks, making shared learning efficient and effective.

Hard sharing significantly reduces the number of model parameters compared to training separate networks for each task. This reduction acts as a regularization mechanism, decreasing the risk of overfitting. By forcing the model to learn representations that are useful for multiple objectives, the network becomes more robust and generalizable.

However, hard parameter sharing assumes that the tasks are sufficiently related. If the tasks are weakly correlated, shared learning may lead to suboptimal performance due to interference between task objectives.

Soft Parameter Sharing

In contrast to hard sharing, soft parameter sharing maintains separate model parameters for each task. Instead of directly sharing layers, the models are encouraged to learn similar representations through regularization constraints.

In this approach, each task has its own neural network, but a penalty term is added to the loss function to minimize the difference between corresponding parameters across tasks. This encourages knowledge transfer while preserving task-specific flexibility.

Soft parameter sharing provides greater adaptability compared to hard sharing, as each task retains its own feature extraction pathway. This is particularly beneficial when tasks are related but not identical, allowing the network to capture subtle differences between objectives.

However, soft sharing increases computational complexity and requires more parameters, which may lead to higher training cost.

2.7.2 Benefits of Multitask Learning

Multitask learning provides several key advantages for complex image analysis systems. By sharing representations across multiple objectives, the model can improve generalization and reduce overfitting, as the learned features capture patterns that are broadly useful rather than specific to a single task. This approach also enhances training efficiency, since a single unified network can replace multiple independent models, reducing computational and memory requirements. Moreover, multitask learning facilitates feature sharing, allowing low-level structures such as edges, textures, or anatomical patterns to be reused across

tasks, which is particularly beneficial in retinal image analysis. Finally, learning multiple objectives simultaneously exerts a regularization effect, stabilizing training and producing more robust and discriminative feature representations. Collectively, these properties make multitask learning an effective strategy for developing accurate, efficient, and reliable image analysis systems.

2.8 Common Datasets Overview

The development and validation of retinal biometric algorithms rely heavily on the availability of standardized, high-quality datasets. This section provides an overview of the publicly available datasets used in this thesis, along with a description of the proprietary clinical dataset that serves as a core component of our experimental work.

2.8.1 Publicly Available Datasets

1. **Retinal Identification Database (RIDB)** is a standardized public benchmark specifically designed for retinal biometric recognition research. It comprises 100 high-resolution fundus images (1504×1000 pixels) captured under controlled conditions using a TOPCON TRC retinal camera. The dataset includes five samples from each of 20 distinct subjects, all free of diagnosed retinal pathology, ensuring anatomical homogeneity and minimizing confounding variability during biometric feature extraction. This structure facilitates the study of intra-subject variation while providing a consistent framework for evaluating verification and identification performance. In this thesis, RIDB serves as an external validation set, enabling assessment of model generalizability across different acquisition devices and populations beyond the primary proprietary dataset (Akram et al., 2020).
2. **ARIA (Automatic Retinal Image Analysis)** is a clinically curated public resource widely employed for ophthalmic image analysis and medical biometric research. It contains three categories of retinal fundus images: 61 from healthy subjects, 92 from patients with Age-related Macular Degeneration (AMD), and 59 from patients with Diabetic Retinopathy (DR). This distribution enables research on both healthy anatomical patterns and pathological variations, making the dataset suitable for both biometric recognition and disease classification tasks. All images have a resolution of 768×576 pixels and were acquired under standardized clinical conditions at St. Paul's Eye Unit (University of Liverpool, UK) (Farnell et al., 2008).
3. **STARE (Structured Analysis of the Retina)** represents a foundational resource in ophthalmic image analysis, originally developed by Michael Goldbaum at the University

of California, San Diego. Primarily designed for automated medical diagnostics, the dataset contains 400 retinal fundus images of high resolution, providing substantial anatomical detail for the analysis of vascular structures and pathological features. A distinctive strength of STARE lies in its expert annotations, particularly the pixel-level segmentation masks of retinal vasculature, which have established it as a standard benchmark for vessel segmentation algorithms. Additionally, the dataset includes diagnostic labels categorizing images as either healthy or pathological (Hoover et al., 2000).

4. **DRIVE (Digital Retinal Images for Vessel Extraction)** is a specialized retinal image collection used to develop and evaluate blood vessel segmentation methods. Created by the Image Sciences Institute in the Netherlands, the dataset comprises 40 high-resolution color fundus images sourced from a diabetic retinopathy screening program. Its standardized structure, with 20 training and 20 test images in fixed partitions, ensures consistent and reproducible comparisons across studies. Each image is accompanied by meticulously crafted manual annotations that provide pixel-level ground truth for retinal vasculature, enabling precise quantitative evaluation of segmentation accuracy (Staal et al., 2004).
5. **VARIA (Vascular Analysis for Retinal Identification Applications)** is designed for retinal biometric research, distinguishing itself from clinically oriented databases by focusing exclusively on identity verification rather than disease diagnosis. Managed by the VARPA research group, this collection comprises 233 retinal fundus images from 139 distinct individuals, with multiple samples per subject enabling critical analysis of intra-subject variability and the longitudinal stability of vascular features. All images were acquired using a non-mydratic TopCon NW-100 camera with consistent optic disc centering at a resolution of 768×584 pixels (Ortega et al., 2009).
6. **39-CAS Dataset** represents a comprehensive multi-class retinal imaging resource specifically designed for fine-grained pathological classification. Derived from the clinical archives of the Joint Shantou International Eye Centre (JSIEC) in Guangdong, China, this curated collection comprises 1,000 color fundus images systematically annotated across 39 distinct retinal disease categories. As a carefully selected subset from a larger clinical repository exceeding 209,000 images, the dataset provides a balanced representation of multiple ocular pathologies while maintaining single-label annotations essential for supervised classification frameworks (Cen et al., 2021).
7. **FIVES (Fundus Image Vessel Segmentation)** constitutes a comprehensive dataset for vascular segmentation, comprising 800 high-resolution color fundus images with

meticulously curated pixel-level annotations of retinal vessels produced through expert-guided standardized protocols. Spanning an exceptionally broad demographic range (ages 4–83) and including diverse pathological presentations (diabetic retinopathy, glaucoma, and age-related macular degeneration), the dataset offers significant value for evaluating model robustness across heterogeneous populations and clinical conditions (Jin et al., 2022).

8. **OIA-ODIR (Ophthalmic Image Analysis – Ocular Disease Intelligent Recognition)** represents a comprehensive clinical resource for ophthalmic image analysis, featuring binocular color fundus photography paired with demographic metadata (age, sex) and expert-provided diagnostic keywords for each patient. Originally developed by Nankai University’s Intelligent Computing System Lab in collaboration with Peking University and Shangong Medical Technology, this standardized research dataset contains 10,000 fundus images sampled from over 1.6 million clinical images across 400+ hospitals in 26 Chinese provinces. The dataset is organized at the image level for multi-disease simultaneous diagnosis, with each fundus image annotated with eight disease labels: normal (N), diabetes (D), glaucoma (G), cataract (C), age-related macular degeneration (A), hypertension (H), myopia (M), and other diseases/abnormalities (O). The age distribution focuses on the 30–80 year range, representing over 96% of samples. OIA-ODIR is recognized as the first international multi-lesion detection dataset based on single fundus images and serves as a standardized benchmark for multi-label ocular disease classification research (Li et al., 2020; Nankai University Intelligent Computing System Lab, 2019).

2.8.2 Local Dataset

is a proprietary clinical retinal image collection acquired at the LAZOUNI ophthalmology clinic. It consists of 425 high-resolution color fundus photographs from 81 subjects, collected during routine ophthalmic examinations. All images were obtained using an EIDON non-mydratic retinal camera in accordance with standard clinical acquisition protocols. The native resolution of 3680×3288 pixels enables detailed visualization of retinal anatomical structures, such as the vascular network, optic disc, and macular region, which is critical for both biometric identification and clinical pathology assessment.

The dataset includes demographic and examination metadata, such as subject sex, date of birth, and the number of acquisitions per eye. The cohort comprises 30 male and 51 female subjects, providing a representative range of clinical diversity. The age distribution ranges from pediatric subjects (approximately 4 years old) to elderly patients over 83 years, facilitating the evaluation of biometric recognition across diverse

age groups and retinal development stages. This variability is especially relevant for medical biometric applications, as retinal morphology may change with aging or disease progression.

The dataset provides images from both the right eye (OD) and left eye (OG) for most subjects. Multiple acquisitions per eye are available for several individuals, with capture counts ranging from single examinations to extensive follow-up records. This multi-sample structure enables both binocular analysis and the investigation of intra-subject variability across repeated imaging sessions.

2.9 Evaluation Metrics for Retinal Image Analysis

The performance of retinal image analysis systems, whether for segmentation, biometric identification, or pathology classification, is assessed using a comprehensive suite of metrics. Each metric quantifies different aspects of system reliability, efficiency, and clinical utility. This section presents the key metrics used throughout this thesis, organized by their primary application domain.

2.9.1 Pixel-Level Segmentation Metrics

For retinal vessel segmentation tasks, pixel-level accuracy and structural preservation are critical. The following metrics evaluate the quality of segmentation masks against ground truth annotations.

- **Dice Coefficient:** Measures the spatial overlap between the predicted segmentation mask P and the ground truth G :

$$\text{Dice} = \frac{2 \cdot |P \cap G|}{|P| + |G|} \quad (2.9.1)$$

where $|P \cap G|$ is the number of pixels correctly predicted as vessels, and $|P|$ and $|G|$ are the total number of predicted and ground truth vessel pixels, respectively. Values range from 0 (no overlap) to 1 (perfect overlap).

- **Intersection over Union (IoU):** Also known as the Jaccard index, measures the ratio of the intersection to the union of predicted and true masks:

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|} \quad (2.9.2)$$

This metric penalizes both false positives and false negatives, providing a strict assessment of segmentation quality.

- **Topology-aware Metrics:** Beyond pixel-level accuracy, maintaining the continuity of vessel branches is critical in retinal biometrics. Metrics such as the Structural Similarity Index (SSIM) adapted for vascular networks or connectivity-preserving scores evaluate whether segmented vessels preserve branching patterns and connectivity essential for reliable identity recognition.

2.9.2 Classification and Identification Metrics

For biometric identification and pathology classification tasks, the following metrics provide a comprehensive evaluation of model performance. These metrics are derived from the confusion matrix elements: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

1. **Accuracy** provides an overall measure of correct predictions across both positive and negative classes:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.9.3)$$

While intuitive, accuracy should be interpreted with caution in the presence of class imbalance.

2. **Precision (Positive Predictive Value)** measures the accuracy of positive identifications:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.9.4)$$

High precision indicates that when the system predicts a positive class, it is likely correct.

3. **Sensitivity (Recall / True Positive Rate / Genuine Acceptance Rate)** measures the ability to detect positive cases:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2.9.5)$$

In biometric contexts, this is referred to as the Genuine Acceptance Rate (GAR).

4. **Specificity (True Negative Rate)** measures the ability to correctly identify negative cases:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2.9.6)$$

High specificity indicates effective rejection of impostors.

5. **F1-Score** provides a balanced measure by combining precision and recall via their

harmonic mean:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.9.7)$$

This metric is particularly useful when seeking a balance between precision and recall.

6. **ROC-AUC (Area Under the Receiver Operating Characteristic Curve)** assesses the classifier's discriminative capacity across all decision thresholds:

$$\text{AUC} = \int_0^1 \text{TPR}(f) d\text{FPR}(f) \quad (2.9.8)$$

where TPR is the True Positive Rate (Sensitivity) and FPR is the False Positive Rate ($1 - \text{Specificity}$). AUC values range from 0.5 (random guessing) to 1.0 (perfect discrimination).

7. **Logarithmic Loss (Log Loss)** penalizes incorrect probabilistic predictions, with higher values indicating greater divergence between predicted probabilities and true labels:

$$\text{Log Loss} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (2.9.9)$$

where y_i is the true label (0 or 1) and p_i is the predicted probability for the i -th observation. This metric is particularly important for assessing the calibration of probabilistic predictions.

8. **Cohen's Kappa** evaluates inter-rater agreement beyond chance, particularly useful for assessing classification consistency across multiple observers or model runs:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \quad (2.9.10)$$

where p_0 is the observed agreement (accuracy) and p_e is the expected agreement by chance. Values range from ≤ 0 (no agreement) to 1 (perfect agreement).

2.9.3 Biometric-Specific Metrics

In biometric verification systems, additional metrics characterize the trade-off between security and usability:

- **False Acceptance Rate (FAR):** The proportion of impostor attempts incorrectly accepted as genuine:

$$\text{FAR} = \frac{FP}{FP + TN} \quad (2.9.11)$$

- **False Rejection Rate (FRR):** The proportion of genuine attempts incorrectly

rejected as impostors:

$$\text{FRR} = \frac{FN}{FN + TP} \quad (2.9.12)$$

- **Equal Error Rate (EER):** The operating point where FAR equals FRR. Lower EER indicates better overall system performance.
- **Average Identification Time** quantifies computational efficiency, measured as the mean duration required to process a single sample:

$$\text{Avg. Identification Time} = \frac{\text{Total Identification Time}}{\text{Number of Samples}} \quad (2.9.13)$$

This metric is critical for real-time deployment scenarios.

2.9.4 Multi-Label Classification Metrics

For multi-label pathology classification tasks where multiple diseases may co-exist in a single image, additional aggregated metrics are used:

- **Micro-averaged metrics:** Aggregate contributions from all classes to compute global averages, giving equal weight to each sample-label pair.
- **Macro-averaged metrics:** Compute metrics independently for each class and then average, giving equal weight to each class regardless of frequency.
- **Weighted-averaged metrics:** Similar to macro-averaging but weighted by the number of true instances per class.
- **Hamming Loss:** The fraction of incorrectly predicted labels relative to the total number of labels:

$$\text{Hamming Loss} = \frac{1}{N \cdot L} \sum_{i=1}^N \sum_{j=1}^L \mathbf{1}[y_{ij} \neq \hat{y}_{ij}] \quad (2.9.14)$$

where N is the number of samples, L is the number of labels, and $\mathbf{1}[\cdot]$ is the indicator function.

Throughout this thesis, these metrics are consistently applied to evaluate segmentation quality, biometric identification accuracy, and pathology classification performance across all datasets and experimental configurations.

2.10 Conclusion

This chapter presented the key theoretical foundations for artificial intelligence in retinal biometric systems. It covered classical and deep learning methods, transfer learning

strategies, and modern architectures such as ResNet and DenseNet for robust feature extraction. Segmentation theory and attention mechanisms were discussed as essential tools for capturing detailed anatomical structures, while multitask learning was highlighted for its ability to improve generalization and efficiency across related tasks. Collectively, these concepts provide a solid framework for developing accurate, efficient, and reliable retinal biometric models.

Chapter 3

Retinal Authentication System Using Deep Learning and Multi-Modal Feature Extraction

3.1 Introduction

The increasing reliance on biometric technologies in security-sensitive and clinical environments has driven significant research into robust and reliable identification systems. In medical contexts, accurate patient identification is essential not only for access control but also for ensuring continuity of care, preventing medical errors, and enabling secure longitudinal patient monitoring. Among the various biometric modalities, retinal biometrics have attracted growing attention due to the uniqueness, internal location, and long-term stability of retinal vascular structures. These intrinsic characteristics make the retina particularly suitable for high-security authentication systems, as the complex vascular patterns are highly distinctive and remain relatively stable throughout an individual's lifetime (Rodiah et al., 2021; Lazzouni and Messadi, 2025). Consequently, retinal biometrics have emerged as a promising modality for secure identity verification across medical and security applications.

Recent advances in artificial intelligence, particularly deep learning, have significantly improved the performance of medical image analysis and biometric recognition systems. Convolutional Neural Networks (CNNs) have demonstrated the ability to automatically learn discriminative representations from retinal fundus images, outperforming traditional handcrafted feature extraction methods that rely on manually designed descriptors (Rodiah et al., 2021). Through hierarchical feature learning, CNNs can capture complex structural patterns in retinal images, enabling more accurate recognition of individual vascular signatures. These capabilities have motivated numerous studies exploring deep learning-

based retinal biometric systems, where neural networks are trained to identify distinctive vascular characteristics directly from raw retinal images.

Early work by Zanlorensi et al. (2020) demonstrated the feasibility of deep learning for retinal biometric recognition using a CNN architecture that achieved promising results on publicly available datasets. Subsequent research by Jalilian et al. (2022) extended this approach by incorporating triplet loss functions to learn discriminative feature embeddings, enabling more effective identity verification through metric learning. Agnihotri et al. (2019) further improved recognition performance by combining CNNs with similarity-based matching strategies, achieving robust authentication across varying image quality conditions. Thapar et al. (2019) introduced an adaptive margin triplet loss that dynamically adjusts the separation between genuine and impostor pairs, thereby enhancing discrimination of subtle vascular differences.

Beyond CNNs, generative models have also contributed to retinal biometric research. Son et al. (2017) first applied GANs to retinal vessel segmentation, demonstrating improved segmentation accuracy, particularly for thin and low-contrast vessels. Kamran et al. (2021) further developed RV-GAN, a multi-scale GAN architecture that achieved state-of-the-art performance on several public retinal segmentation datasets. Bhanuse et al. (2025) proposed a multi-stage GAN architecture incorporating attention mechanisms and adaptive loss functions to enhance the segmentation of fine vascular structures. These advances have enabled more precise extraction of vascular structures, which is a critical prerequisite for reliable biometric feature extraction.

Beyond vessel-based representations, alternative approaches have explored the use of global retinal texture features for biometric identification. Lazzouni and Messadi (2025) proposed a texture-based retinal authentication method using multifractal analysis combined with wavelet-based image fusion, demonstrating that discriminative biometric information can also be derived from overall retinal texture distribution. Recent research has also explored multimodal biometric systems that combine retinal biometrics with other physiological modalities (Sasikala, 2025; Nardone et al., 2025), as well as privacy-preserving techniques for retinal data processing (Luo et al., 2026).

Despite these advancements, several limitations remain in current retinal authentication systems. Many existing methods rely on low-resolution images, which fail to preserve the fine vascular details necessary for highly discriminative biometric recognition. Furthermore, numerous studies evaluate their methods using a single dataset, limiting models' ability to generalise across different imaging devices and acquisition conditions. Additionally, conventional pipelines often separate vessel segmentation from the feature extraction and authentication stages, preventing the system from being optimised end-to-end for biometric recognition tasks.

Contributions of This Chapter

This chapter presents a comprehensive retinal authentication framework designed to overcome key limitations in existing biometric systems. The proposed approach integrates high-resolution preprocessing, advanced vessel segmentation, and multi-modal feature extraction into a unified pipeline. First, an efficient preprocessing and patch extraction strategy is introduced to preserve fine vascular details while maintaining computational feasibility. Second, a Residual Attention U-Net architecture is developed to enhance vessel segmentation by combining residual learning and attention mechanisms, enabling robust and stable feature learning. Third, a hybrid multi-modal feature extraction framework is proposed, incorporating morphological, topological, and texture-based descriptors to improve discriminative power while ensuring interpretability. Finally, the system is validated through a systematic evaluation of verification classifiers, including SVM, KNN, Decision Tree, Random Forest, Extra Trees, XGBoost, AdaBoost, and Voting ensembles, with optimized decision thresholds, demonstrating strong generalization across multiple public retinal datasets.

3.2 Methodology

The proposed retinal authentication system follows a sequential processing pipeline that transforms retinal fundus images into discriminative biometric features suitable for identity verification. As illustrated in Figure 3.1, the system consists of four main stages: image preprocessing, high-resolution patch extraction, retinal vessel segmentation using a residual attention U-Net, post-processing, and multi-modal feature extraction followed by authentication. Each stage contributes to the transformation of raw retinal images into structured vascular representations that can be used for biometric verification.

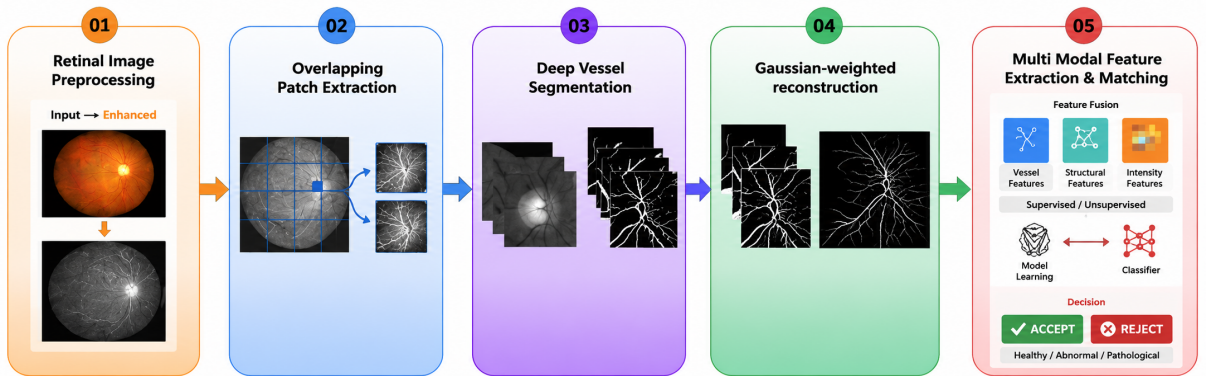


Figure 3.1. Proposed retinal authentication system pipeline.

3.2.1 Image Preprocessing

Preprocessing standardizes retinal fundus images to enhance vessel visibility while preserving anatomical structures. All images are resized to 512×512 pixels. The green channel is extracted and normalized, as it typically provides the highest contrast between blood vessels and the retinal background. Contrast is further enhanced using Contrast Limited Adaptive Histogram Equalization (CLAHE), followed by gamma correction to adjust global intensity levels.

Noise is reduced using bilateral filtering, which performs spatial smoothing while simultaneously considering intensity differences between neighboring pixels. This approach preserves edges associated with vessel boundaries while effectively suppressing noise. The filtered image is computed as:

$$I_{bf}(x, y) = \frac{1}{W(x, y)} \sum I_{gamma}(x_i, y_i) \cdot W_S(x_i, y_i) \cdot W_\tau(x_i, y_i) \quad (3.2.1)$$

where W_S represents the spatial weighting function, W_τ represents the intensity similarity weighting function, and $W(x, y)$ is a normalization factor that ensures proper scaling of the filtered intensities.

To further emphasize vessel boundaries, unsharp masking is applied as the final preprocessing step. This technique enhances high-frequency components associated with edges by combining the original image with a high-pass filtered version. The sharpened image is computed as:

$$I_{sharp}(x, y) = I_{bf}(x, y) + \alpha(I_{bf}(x, y) - G_\alpha * I_{bf}(x, y)) \quad (3.2.2)$$

where G_α denotes a Gaussian smoothing filter and α controls the sharpening strength. This operation increases the visibility of vessel contours and fine vascular branches that are important for biometric feature extraction.

Finally, the resulting image is clipped to ensure that pixel intensities remain within the valid range:

$$I_{final}(x, y) = \max(0, \min(1, I_{sharp}(x, y))) \quad (3.2.3)$$

Figure 3.2 illustrates the impact of the preprocessing pipeline on a representative retinal fundus image. The sequential application of CLAHE, gamma correction ($\gamma = 1.40$), bilateral filtering, and unsharp masking progressively enhances vessel visibility and contrast. The final enhanced image exhibits clearer vascular structures with effective noise suppression. The histogram redistribution confirms improved contrast normalization, while

the difference map demonstrates that intensity modifications are primarily concentrated along vascular regions.

Quantitative statistics further highlight the preprocessing effect: the mean intensity decreased from 82.0 to 69.8, the standard deviation slightly reduced from 52.7 to 47.5, the minimum intensity remained at 0, and the maximum intensity increased from 222 to 243. These changes indicate improved contrast balance.

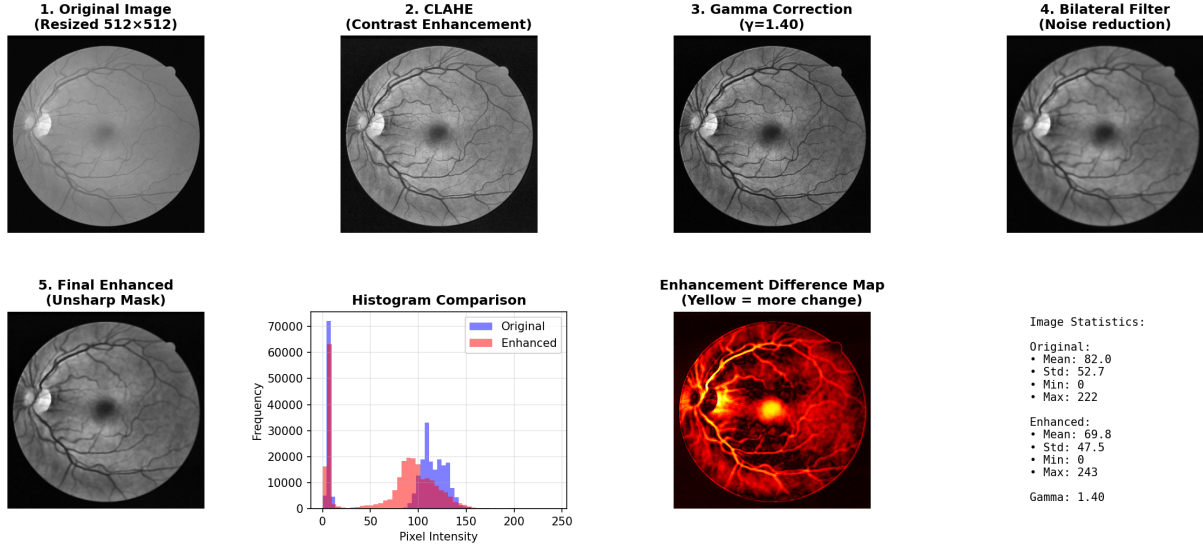


Figure 3.2. Preprocessing pipeline applied to a sample retinal fundus image.

The final preprocessed image serves as the input to the next stage of the authentication pipeline, where localized patches are extracted to enable detailed vessel segmentation and subsequent feature analysis.

3.2.2 Patch Extraction Strategy

After preprocessing, each 512×512 retinal image is divided into overlapping 64×64 patches to enable localized vessel analysis while preserving structural details. A sliding window with a stride of 32 pixels (50% overlap) generates 225 patches per image, providing enhanced spatial coverage and preventing information loss at patch boundaries.

The selected patch size ensures a balance between contextual representation of vascular continuity and computational efficiency. Overlapping patches further improve reconstruction quality by reducing boundary artifacts and increasing training diversity.

To address class imbalance, additional patches are selectively sampled from vessel-dense regions using ground truth masks, and patches with less than 1% vessel pixels are discarded. This strategy ensures that the training dataset emphasizes informative vascular structures, thereby improving segmentation accuracy and model generalization.

Figure 3.3 illustrates the complete patch extraction workflow applied to a representative retinal fundus image from the DRIVE dataset, demonstrating the structured transition from high-resolution image acquisition to statistically validated patch generation for vessel segmentation.

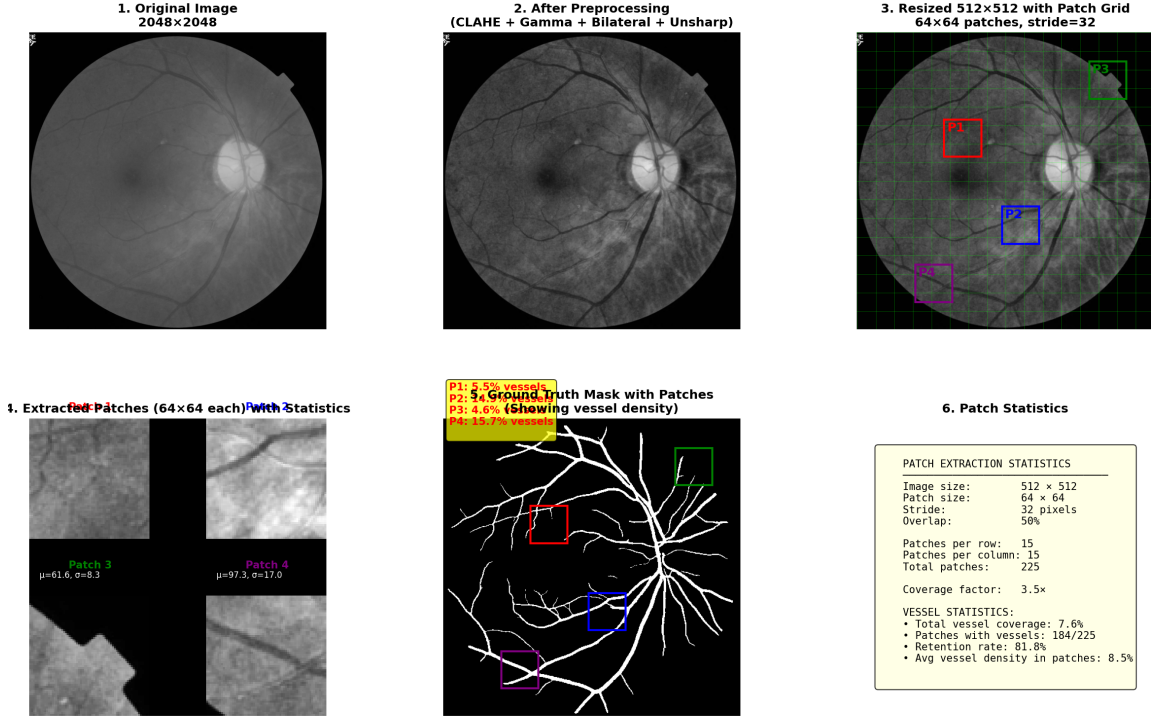


Figure 3.3. Patch extraction methodology applied to a sample retinal fundus image from the DRIVE dataset.

The sequence begins with (1) the original high-resolution image (2048×2048), which preserves rich anatomical information but exhibits typical retinal imaging challenges, including non-uniform illumination and limited vessel-to-background contrast, particularly in peripheral regions.

In (2) the preprocessed image, the application of Contrast Limited Adaptive Histogram Equalization (CLAHE), gamma correction ($\gamma = 1.40$), bilateral filtering, and unsharp masking significantly enhances vascular visibility. Major vascular arcades and fine capillary branches become more distinguishable, and global structural clarity is improved, facilitating more reliable patch extraction and subsequent segmentation.

The workflow proceeds with (3) the resized 512×512 image overlaid with a structured grid, illustrating the systematic division into 64×64 patches using a stride of 32 pixels. This configuration introduces a 50% overlap between adjacent patches, ensuring dense and uniform spatial coverage across the retinal field. The overlap guarantees that elongated vascular branches traversing large image regions appear in multiple neighboring patches, thereby mitigating the risk of structural discontinuity associated with non-overlapping

sampling. In practice, each pixel is represented approximately four times, increasing contextual redundancy and robustness during both training and reconstruction.

In (4) the extracted sample patches, representative 64×64 regions demonstrate substantial variability in local anatomical characteristics. Patches centered on the optic disc exhibit higher mean intensity ($\mu = 142.3$) due to the naturally bright disc region but lower vessel density (5.2%), as the disc contains relatively few visible vessels. Conversely, patches extracted from vascular arcades show moderate intensity ($\mu = 98.7$) and higher vessel density (12.8%), reflecting multiple branching vessels within the local field of view. Regions containing dense vascular networks present the highest vessel density (18.3%), while peripheral patches illustrate the challenge of segmenting fine, low-contrast capillaries (vessel density 8.1%, $\mu = 76.4$). This diversity enriches the training dataset with multiple structural contexts, including bifurcations, crossings, thick vessels, and thin peripheral branches.

The sequence continues with (5) the vessel mask visualization, where patch boundaries are superimposed onto the ground truth segmentation. This confirms that vessels crossing patch limits are preserved within overlapping regions, preventing fragmentation of continuous vascular structures and ensuring consistent representation during supervised learning.

The statistical analysis further demonstrates that applying a 1% vessel-content filtering threshold effectively removes predominantly background patches while retaining the majority of informative vascular regions. Given that the natural vessel-to-background pixel ratio in retinal images is approximately 1:9, this filtering step is critical for reducing class imbalance. By preferentially sampling patches containing vascular content, the model is exposed to a more balanced distribution of vessel and non-vessel pixels, thereby reducing bias toward the majority background class and improving segmentation sensitivity.

Finally, (6) the statistical summary panel reports the extraction parameters (patch size = 64×64 , stride = 32, total patches = 225) and vessel density metrics. These statistics validate the coverage factor introduced by overlapping sampling and confirm that informative vascular regions are adequately represented across the dataset.

Overall, the sequential visualization demonstrates a coherent and methodologically sound transition from raw image acquisition to structured patch generation. The combination of preprocessing, overlapping sampling, vessel-content filtering, and statistical validation ensures robust, context-rich, and class-balanced input preparation for deep learning-based retinal vessel segmentation models.

3.2.3 Model Development

To achieve accurate and structurally consistent retinal vessel segmentation for biometric authentication, a deep convolutional neural network based on a residual attention mechanism was developed. The architecture extends the classical U-Net framework by integrating residual learning and attention gating to enhance feature propagation and vessel-specific representation.

Architectural Design

The proposed network follows an encoder–bridge–decoder topology with symmetric skip connections. However, unlike the standard U-Net, each convolutional stage is reformulated as a residual block, and attention gates are introduced at each skip connection to refine spatial feature transmission.

Encoder. The encoder progressively extracts hierarchical feature representations through four residual blocks with increasing channel depths (32, 64, 128, and 256 filters). Each block consists of two 3×3 convolutional layers followed by batch normalization and ReLU activation. A shortcut connection aligns feature dimensions using a 1×1 convolution to enable residual learning:

$$\mathbf{F}_{\text{res}} = \text{ReLU}(\mathcal{F}(\mathbf{x}) + \mathbf{W}_s \mathbf{x}) \quad (3.2.4)$$

where $F(x)$ denotes the stacked convolutional operations, W_s represents the shortcut projection, and $\text{ReLU}(\cdot)$ is the rectified linear activation function.

Downsampling is performed using 2×2 max pooling layers, allowing the network to capture increasingly abstract contextual information while reducing spatial resolution.

Bridge. At the bottleneck level, a deep residual block with 512 filters captures high-level semantic representations of vascular topology. This stage encodes global structural dependencies across the retinal image.

Decoder. The decoder reconstructs spatial resolution using bilinear upsampling followed by feature concatenation with encoder outputs. Prior to concatenation, encoder features are refined using attention gates to suppress irrelevant background activations. The attention coefficient is computed as:

$$\alpha_{\text{attention}}(x, g) = \text{sig}(W_\psi \cdot \text{ReLU}(W_x x + W_g g + b)) \quad (3.2.5)$$

where:

- x denotes the feature map from the encoder (skip connection),
- g denotes the gating signal from the decoder,
- W_x and W_g are learnable 1×1 convolutional transformations,
- b is a bias term,
- $\text{ReLU}(\cdot)$ represents the rectified linear activation,
- W_ψ is a learnable linear projection,
- $\text{sig}(\cdot)$ denotes the sigmoid activation function.

The refined feature map becomes:

$$x' = \alpha_{\text{attention}}(x, g) \cdot x \quad (3.2.6)$$

This mechanism allows the network to emphasize vessel structures while attenuating non-informative retinal background regions.

The final output layer applies a 1×1 convolution followed by sigmoid activation to produce a pixel-wise vessel probability map.

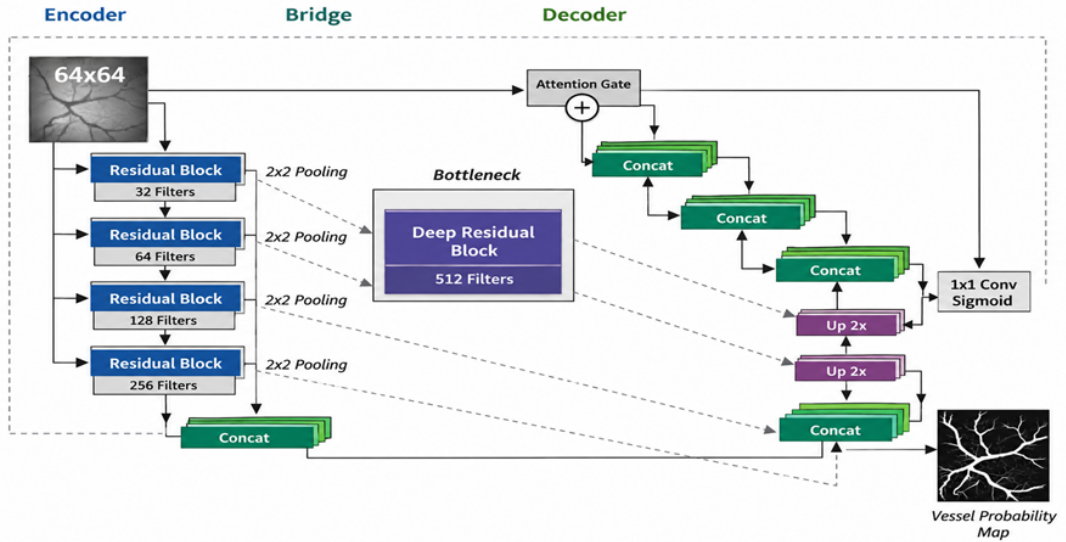


Figure 3.4. Model architecture.

3.2.4 Loss Function and Training Protocol

Retinal vessel segmentation is characterized by severe class imbalance, where vessel pixels occupy only a small portion of the image. To address this issue, the Dice loss is employed, as it directly optimizes the spatial overlap between predicted segmentation maps and ground truth masks, thereby improving the detection of fine vascular structures.

The model is optimized using the Adam optimizer with an initial learning rate of 2×10^{-4} . Training stability and generalization are enhanced through early stopping, learning rate reduction on plateau, and model checkpointing.

The dataset is divided into training and validation sets using an 80/20 split. All input patches are normalized using the dataset mean μ_{data} and standard deviation σ_{data} . The model is trained for up to 100 epochs with a batch size of 16.

Performance is evaluated using the Dice coefficient and pixel-wise accuracy, providing a comprehensive assessment of segmentation quality and structural fidelity.

3.2.5 Gaussian-Weighted Reconstruction

During inference, reconstructing the full segmentation mask from individual patch predictions is challenging due to overlapping regions. A naive averaging of overlapping patches often produces visible seams and artifacts where predictions disagree. To address this, a Gaussian-weighted reconstruction strategy is employed.

Each pixel in the final image receives contributions from multiple overlapping patches, weighted according to its distance from the patch center:

$$W(x, y) = \max \left(0.3, \exp \left(-\frac{(y - c_y)^2 + (x - c_x)^2}{2\sigma^2} \right) \right) \quad (3.2.7)$$

where (c_y, c_x) denote the patch center coordinates, and $\sigma = P/4 = 16$ pixels controls the spread of the Gaussian. This function assigns maximum weight to pixels near the center while ensuring that edge pixels maintain a minimum contribution of 0.3.

The reconstructed image I_{recon} is computed as a normalized weighted sum over all overlapping patches:

$$I_{\text{recon}}(y, x) = \frac{\sum_i W_i(y - y_i, x - x_i) \cdot I_{\text{patch}}^i(y - y_i, x - x_i)}{\sum_i W_i(y - y_i, x - x_i) + \epsilon} \quad (3.2.8)$$

where I_{patch}^i denotes the predicted segmentation for patch i , (y_i, x_i) is the top-left coordinate of patch i , and $\epsilon = 10^{-8}$ prevents division by zero in regions with no patch coverage.

This Gaussian-weighted approach smooths transitions between patches, reduces boundary artifacts, and preserves the continuity of vascular structures across the retinal image.

Figure 3.5 illustrates this mechanism: overlapping 64×64 patches with 50% overlap contribute to the same spatial locations, with pixels near patch centers assigned higher weights. This weighting ensures that predictions with full contextual support dominate, while predictions near patch borders where the receptive field is limited have reduced

influence. Consequently, thin vessels crossing patch boundaries are seamlessly connected in the reconstructed mask, and no visible discontinuities or checkerboard patterns are observed. The chosen value of $\sigma = 16$ pixels provides an optimal balance between central dominance and sufficient edge contribution.

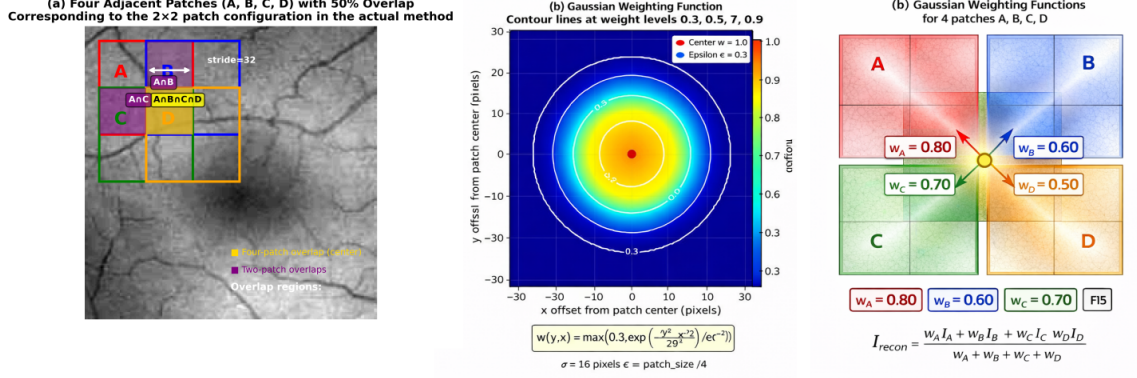


Figure 3.5. Illustration of Gaussian-weighted reconstruction. Overlapping 64×64 patches contribute to each pixel in the final segmentation mask, with higher weights assigned near patch centers.

3.2.6 Retinal Vessel Feature Extraction

Once the full-resolution vessel segmentation map is reconstructed, the next step in the authentication pipeline involves extracting discriminative vascular features that uniquely characterize each individual. Retinal blood vessels form a complex branching network whose topology and spatial arrangement are highly distinctive and stable over time, making them ideal for biometric identification.

The feature extraction stage converts the segmented vessel map into a compact, informative representation capturing the geometric and structural properties of the vascular tree. To achieve this, the system computes several categories of vascular descriptors.

Morphological Features

Morphological descriptors capture the physical characteristics of vessels, including:

- **Vessel thickness:** derived from distance transforms along the vessel skeleton.
- **Curvature and tortuosity:** representing how straight or winding vessel segments are.
- **Local vessel density:** the proportion of vessel pixels in defined neighborhoods.

These features describe the geometric structure and spatial distribution of the retinal vasculature. Figure 3.6 illustrates an example of the extracted morphological descriptors,

showing a dense, radially organized vascular network centered at the optic disc, with hierarchical branching from thick central vessels to fine peripheral capillaries. The measured vessel density of 8.91% indicates moderate and anatomically consistent vascular coverage, confirming the structural integrity of the segmentation and its suitability for biometric analysis.

The distance transform quantitatively confirms physiological vessel tapering, with a maximum radius of 6.00 px at the optic disc and a mean width of 1.43 px, indicating a capillary-dominant structure.

The tortuosity map highlights moderate curvature along major arcades and localized high-curvature regions near the optic disc. A mean tortuosity of 1.42 with a high maximum (14.45) reflects predominantly straight vessels with a limited number of sharply curved central segments.

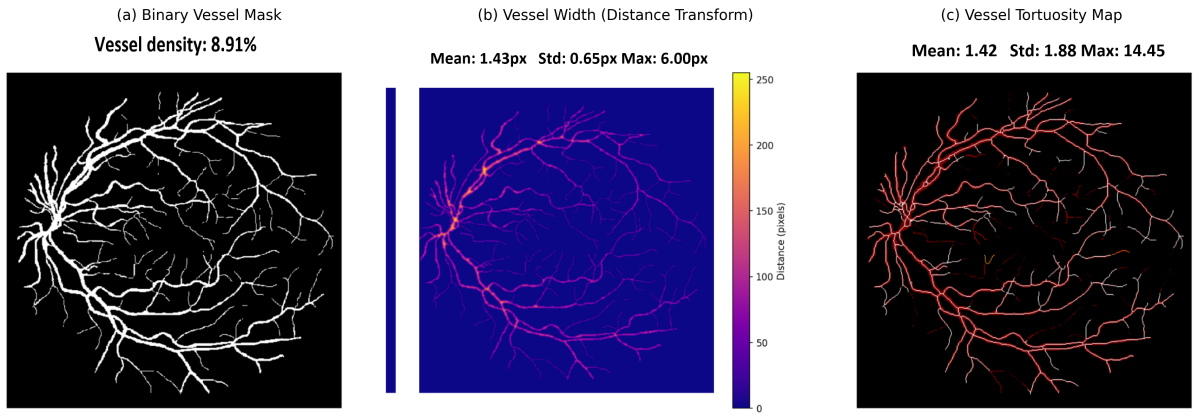


Figure 3.6. Morphological Features: (a) Input Binary Mask, (b) Distance Transform, (c) Tortuosity Map.

Topological Features

Topological descriptors characterize the connectivity, branching patterns, and spatial organization of the retinal vascular network:

- **Bifurcation points:** locations where a vessel splits into two or more branches.
- **Endpoints:** terminal vessel segments.
- **Cross points:** intersections of three or more vessels.
- **Skeleton length and branching density:** capturing overall network geometry and connectivity.
- **Orientation and spatial distribution:** representing vessel alignment and regional concentration of landmarks.

Figure 3.7 illustrates the extracted topological features, including skeletonization, bifurca-

tion and endpoint detection, spatial grid analysis, and orientation characterization.

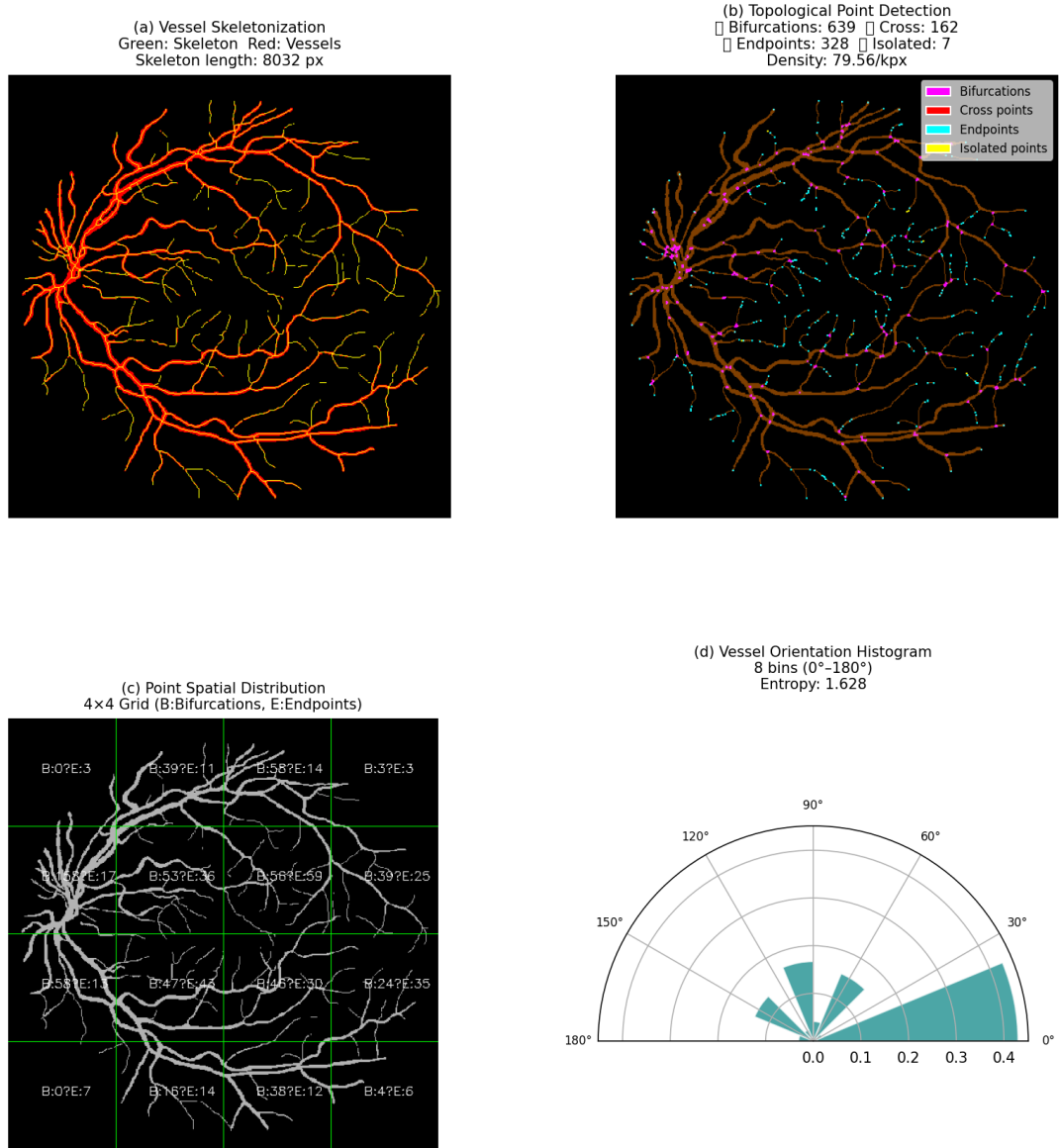


Figure 3.7. Topological Features: (a) Skeletonization, (b) Bifurcation Detection, (c) Spatial Grid, (d) Orientation Histogram.

Skeletonization reduces vessels to a one-pixel-wide representation while preserving connectivity, resulting in a total skeleton length of 8032 pixels, indicative of a rich and spatially extensive vascular network. Automated landmark detection identified 639 bifurcations, 162 cross points, and 328 endpoints, suggesting a highly interconnected and hierarchically organized vasculature, with the optic disc region acting as the principal topological hub.

Spatial grid analysis reveals an asymmetric distribution of bifurcations, with a higher concentration in central grid cells, forming a distinctive spatial signature. The orientation histogram demonstrates a multi-directional vessel arrangement, dominated by horizontal

and diagonal orientations. The computed orientation entropy of 1.628 quantifies this structured yet diverse directional organization, reinforcing the discriminative power of these topological descriptors for retinal analysis and biometric applications.

Texture-Based Descriptors

Texture descriptors capture local intensity variations along vessel segments, reflecting subtle structural patterns not apparent from geometry alone. Examples include:

- **Local Binary Patterns (LBP)** along vessel segments.
- **Fractal dimension analysis**, indicating the complexity of vessel branching.

Figure 3.8 illustrates the texture-based feature extraction process, including the Local Binary Pattern (LBP) map, the corresponding histogram, and fractal dimension analysis.

The LBP texture map highlights micro-structural intensity transitions along vessel boundaries, effectively distinguishing edge regions from uniform vessel interiors and background areas. The 24-bin LBP histogram, together with an entropy value of 1.749, indicates moderate textural variability and provides a quantitative description of local structural patterns.

Fractal analysis exhibits strong log–log linearity, resulting in a computed fractal dimension of 1.501, which lies within the established physiological range (1.40–1.70). This confirms the scale-invariant and self-similar nature of the retinal vascular network.

Overall, texture-based descriptors strengthen the biometric representation by capturing micro-pattern variability and global branching complexity beyond purely geometric and topological measurements.

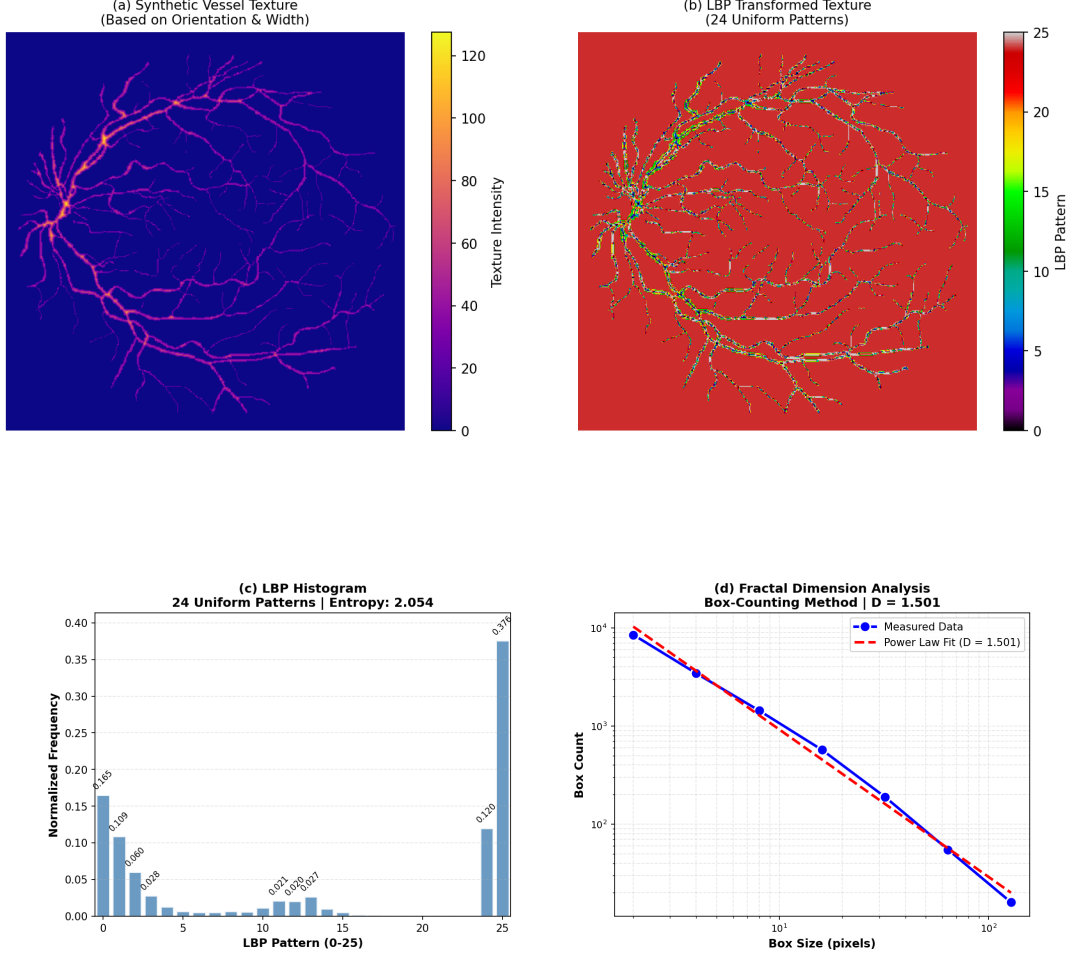


Figure 3.8. Texture-Based Descriptors: (a) LBP Texture Map, (b) LBP Histogram, (c) Fractal Dimension Analysis.

3.2.7 Biometric Gallery Construction and Pair Generation

The proposed biometric authentication system follows a verification paradigm, in which each subject is represented by a reference template constructed from enrollment samples, while the remaining samples are used as probes.

Formally, for each subject, a template T_{subject} is computed as the centroid of the feature embeddings:

$$T_{\text{subject}} = \frac{1}{N_{\text{subject}}} \sum_{i=1}^{N_{\text{subject}}} F_i \quad (3.2.9)$$

where $F_i \in \mathbb{R}^d$ denotes the feature vector of the i -th enrollment image. This averaging operation reduces intra-subject variability and provides a robust estimate of the subject representation in the embedding space.

To train the verification model, pairs are constructed by comparing probe features with gallery templates:

- **Genuine pairs (class 1):**

$$\mathbf{d}_{\text{genuine}} = |F_P - T_{S_{\text{true}}}| \quad (3.2.10)$$

- **Impostor pairs (class 0):**

$$\mathbf{d}_{\text{impostor}} = |F_P - T_{S_{\text{imp}}}| \quad (3.2.11)$$

where $F_P \in \mathbb{R}^d$ denotes the probe feature vector, S_{true} is the true identity of the probe, and S_{imp} denotes an impostor identity. These difference vectors $\mathbf{d} \in \mathbb{R}^d$ serve as discriminative descriptors for the verification task.

The use of the absolute difference $|F_P - T|$ can be interpreted as a distance-based embedding transformation. Assuming that the learned feature space is structured such that samples from the same subject are clustered together, it is expected that:

$$\|F_P - T_{S_{\text{true}}}\| \ll \|F_P - T_{S_{\text{imp}}}\| \quad (3.2.12)$$

Thus, the verification problem reduces to learning a decision boundary that separates low-distance (genuine) pairs from high-distance (impostor) pairs. The element-wise absolute difference preserves feature-wise discrepancies, enabling the classifier to learn the most discriminative dimensions.

3.2.8 Biometric Verification Model

The biometric verification task is formulated as a binary classification problem. Given a claimed identity, the model computes a difference vector between the probe feature representation and the corresponding template:

$$d = |F_{\text{probe}} - T_{\text{claimed}}| \quad (3.2.13)$$

This difference vector serves as input to a set of classifiers, which implicitly learn a similarity function in the embedding space. Each classifier outputs a probability score reflecting the likelihood that the probe–claimed pair is genuine or impostor.

To ensure a comprehensive evaluation of the verification framework, a diverse set of machine learning models is employed, including Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Decision Trees (DT), Random Forests (RF), Extra Trees (ET),

Extreme Gradient Boosting (XGBoost), and AdaBoost. This diversity enables the modeling of different decision boundaries and enhances robustness.

SVM models were evaluated using both linear and radial basis function (RBF) kernels, with hyperparameters such as the regularization parameter C and kernel parameter γ carefully tuned. Decision Trees were constrained by a maximum depth to mitigate overfitting. Ensemble methods, including Random Forests and Extra Trees, were constructed using multiple trees with controlled depth to balance performance and computational efficiency. XGBoost was configured as a gradient-boosted ensemble with a learning rate of 0.05, 100 estimators, a maximum depth of 5, and appropriate regularization parameters. AdaBoost employed a sequence of weak learners (decision stumps), where each iteration focused on previously misclassified samples.

The predicted probability scores were subsequently converted into binary verification decisions using a threshold τ . Threshold selection was performed using standard strategies, including the Equal Error Rate (EER), corresponding to the operating point where the False Acceptance Rate (FAR) equals the False Rejection Rate (FRR), as well as Receiver Operating Characteristic (ROC) curve analysis to balance sensitivity and specificity. Application-specific constraints were also considered to satisfy security requirements.

The verification performance was evaluated using standard biometric metrics, including Accuracy, FAR, FRR, EER, and the Area Under the ROC Curve (AUC), providing a comprehensive assessment of the proposed verification framework.

3.3 Results and Discussion

This chapter presents a comprehensive evaluation of the proposed retinal biometric authentication system. The experimental results are organized according to the pipeline stages: retinal vessel segmentation performance, feature extraction analysis, and biometric verification accuracy. Quantitative metrics are reported alongside qualitative visual assessments to provide a thorough understanding of system behavior. Comparative analyses are conducted against baseline methods and alternative architectures to demonstrate the effectiveness of the proposed Residual Attention U-Net and the verification framework using multiple classifiers including Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Decision Trees (DT), Random Forests (RF), Extra Trees, XGBoost, and AdaBoost.

3.3.1 Retinal Vessel Segmentation Results

To improve the generalization capability of the proposed segmentation model, a combined training strategy was adopted using images from the DRIVE, ARIA, and STARE datasets.

The three datasets were merged to increase data diversity in terms of imaging conditions, vessel appearance, and pathology variations.

During training, the model was learned on the combined dataset, while evaluation was performed separately on each data to assess dataset-specific performance and ensure that the model does not overfit to a particular data distribution. The results in table 3.1 demonstrate the strong generalization capability of the proposed model when trained on a combination of multiple datasets and evaluated independently on each dataset. The high Dice score achieved on DRIVE (0.9168) indicates excellent vessel segmentation performance on standard field-of-view retinal images, suggesting that the model effectively captures fine vascular structures under typical imaging conditions.

On the STARE dataset, the Dice score (0.9023) remains high, although slightly lower than DRIVE, which can be attributed to the presence of more pathological cases and a wider field of view, introducing additional variability and complexity. Despite this, the model maintains robust performance, indicating good adaptability to different retinal characteristics.

For the ARIA dataset, the Dice score (0.9099) is comparable to DRIVE, further confirming the stability of the model across datasets with different acquisition settings and population characteristics. The AUC values across datasets are consistently high, with DRIVE achieving near-perfect discrimination (0.9912), while STARE and ARIA also show strong classification capability. The slightly lower AUC values for STARE and ARIA may reflect dataset-specific variability, including imaging quality and pathological diversity.

Table 3.1. Cross-Dataset Evaluation (Trained on Combined Datasets, Tested Separately)

Dataset	Dice Score	AUC
DRIVE	0.9168	0.9912
STARE	0.9023	0.9085
ARIA	0.9099	0.9098

Once the best-performing architecture was identified it was further evaluated on additional unseen datasets, including RIDB and VARIA, without any retraining or fine-tuning. This cross-dataset evaluation protocol was designed to assess the robustness and generalization ability of the model in real-world scenarios.

The consistent performance across multiple datasets demonstrates the ability of the proposed model to effectively capture invariant vascular features and to generalize across different acquisition conditions and population characteristics.

Figure 3.9 illustrates the robustness of the proposed model across different datasets. The top row shows the original retinal images from DRIVE, RIDB, and VARIA datasets,

respectively, while the bottom row presents the corresponding vessel segmentation outputs produced by the proposed Residual Attention U-Net. Despite significant variations in image characteristics, including illumination, contrast, and modality, the model consistently extracts vascular structures with high accuracy. This demonstrates the strong generalization capability and robustness of the proposed approach across different datasets.

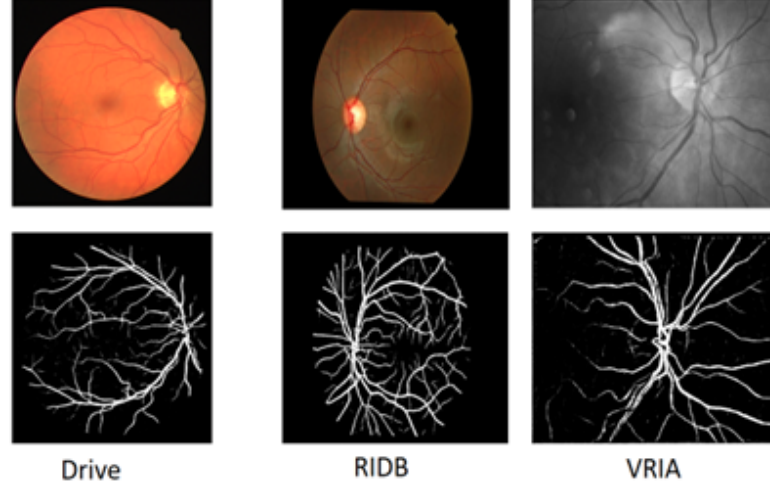


Figure 3.9. Qualitative segmentation results across multiple datasets.

3.3.2 Biometric Verification Results

The figure 3.10 illustrates the complete pipeline of the proposed biometric verification framework, encompassing enrollment, pair construction, and decision making in the embedding space. During the enrollment phase, multiple samples of each subject S_k are encoded into feature embeddings F_i , which are averaged to compute a robust gallery template

$$T_{S_k} = \frac{1}{N_k} \sum_{i=1}^{N_k} F_i.$$

This centroid-based representation effectively reduces intra-subject variability and stabilizes identity representation. For verification, a probe image is mapped to a feature vector $F_P \in \mathbb{R}^d$, which is compared against gallery templates to generate verification pairs. Genuine pairs are formed by comparing the probe with its true template,

$$d_{\text{genuine}} = |F_P - T_{S_{\text{true}}}|,$$

while impostor pairs are obtained by comparing it with templates of other subjects,

$$d_{\text{impostor}} = |F_P - T_{S_{\text{imp}}}|.$$

These difference vectors serve as discriminative inputs to the classifier. In the embedding space, genuine pairs are expected to exhibit small distances, whereas impostor pairs show significantly larger distances, i.e.,

$$\|F_P - T_{S_{\text{true}}}\| \ll \|F_P - T_{S_{\text{imp}}}\|.$$

Consequently, the verification task is formulated as a binary classification problem, where a decision boundary separates low-distance (genuine) from high-distance (impostor) pairs. This framework effectively combines template averaging, distance-based feature transformation, and supervised classification to achieve robust and generalizable biometric verification performance.

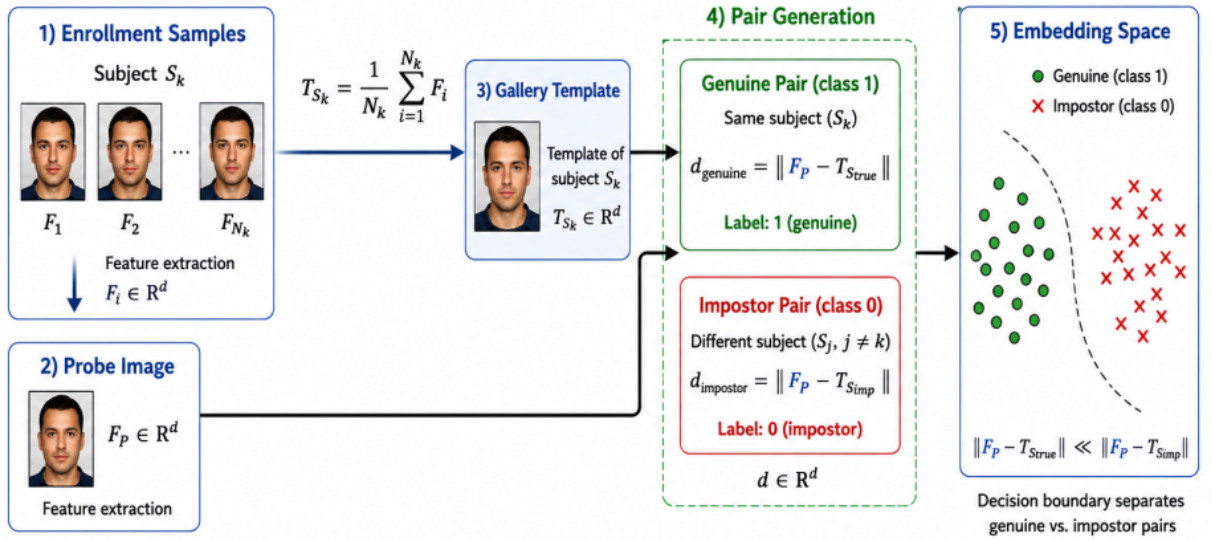


Figure 3.10. Biometric Gallery Construction and Pair Generation for Verification

Analysis of Verification Classifier Performance

The performance of multiple verification classifiers was evaluated across five retinal datasets (RIDB, DRIVE, ARIA, VARIA, and STARE) using standard biometric metrics, including Accuracy, F1-score, Area Under the ROC Curve (AUC), False Acceptance Rate (FAR), and False Rejection Rate (FRR). The results provide insight into the discriminative power of the proposed feature representation as well as the robustness of different classifiers under varying dataset conditions.

Performance on RIDB Dataset. As reported in Table 3.2, all classifiers achieve perfect performance across all evaluation metrics (Accuracy, F1-score, and AUC equal to 1.0000, with FAR and FRR equal to 0.0000). This indicates complete separability between genuine and impostor pairs. Such results suggest that the RIDB dataset represents a highly controlled scenario, where intra-class variability is minimal and the learned feature

embeddings exhibit strong discriminative capability.

Table 3.2. Performance Comparison of Verification Classifiers on RIDB Dataset

Classifier	Accuracy	F1-Score	AUC	FAR	FRR
XGBoost	1.0000	1.0000	1.0000	0.0000	0.0000
AdaBoost	1.0000	1.0000	1.0000	0.0000	0.0000
Random Forest	1.0000	1.0000	1.0000	0.0000	0.0000
SVM	1.0000	1.0000	1.0000	0.0000	0.0000
Decision Tree	1.0000	1.0000	1.0000	0.0000	0.0000
Extra Trees	1.0000	1.0000	1.0000	0.0000	0.0000
KNN	1.0000	1.0000	1.0000	0.0000	0.0000

Performance on DRIVE Dataset. Table 3.3 shows that, although overall performance remains extremely high, slight variability across classifiers begins to emerge. XGBoost maintains perfect performance, while AdaBoost and Random Forest achieve near-perfect results with marginal error rates. In contrast, classifiers such as Decision Tree, SVM, Extra Trees, and KNN exhibit a noticeable reduction in performance, particularly in terms of AUC and error rates. This suggests that the DRIVE dataset introduces moderate variability, leading to partial overlap in the embedding space and increased sensitivity to classifier complexity and decision boundaries.

Table 3.3. Performance Comparison of Verification Classifiers on DRIVE Dataset

Classifier	Accuracy	F1-Score	AUC	FAR	FRR
XGBoost	1.0000	1.0000	1.0000	0.0000	0.0000
AdaBoost	0.9999	0.9999	1.0000	0.0001	0.0001
Random Forest	0.9990	0.9998	0.9999	0.0007	0.0003
Decision Tree	0.9851	0.9771	0.9947	0.0033	0.0059
SVM	0.9863	0.9812	0.9942	0.0035	0.0035
Extra Trees	0.9970	0.9896	0.9789	0.0030	0.00002
KNN	0.9789	0.9789	0.9715	0.0026	0.0039

Performance on ARIA Dataset. The results in Table 3.4 indicate a further increase in task difficulty. XGBoost remains the top-performing method, achieving near-perfect metrics with minimal FAR and FRR. Ensemble-based approaches (AdaBoost and Random Forest) also demonstrate strong performance, confirming their robustness in modeling complex feature distributions. In contrast, simpler models such as Decision Tree and KNN exhibit higher error rates, highlighting their limited capacity to generalize under increased intra-class variability and more challenging feature separability.

Table 3.4. Performance Comparison of Verification Classifiers on ARIA Dataset

Classifier	Accuracy	F1-Score	AUC	FAR	FRR
XGBoost	0.9987	0.9985	0.9992	0.0012	0.0013
AdaBoost	0.9982	0.9980	0.9989	0.0016	0.0018
Random Forest	0.9978	0.9975	0.9986	0.0020	0.0022
Decision Tree	0.9851	0.9771	0.9947	0.0033	0.0059
SVM	0.9863	0.9812	0.9942	0.0035	0.0035
Extra Trees	0.9970	0.9896	0.9789	0.0030	0.00002
KNN	0.9789	0.9789	0.9715	0.0026	0.0039

Performance on VARIA Dataset. As shown in Table 3.5, a similar trend is observed on the VARIA dataset. While XGBoost continues to achieve near-perfect performance, other classifiers display increased degradation compared to previous datasets. In particular, SVM and Decision Tree show reduced discriminative ability, as reflected by higher FAR and FRR values. These results indicate that VARIA introduces additional variability, likely due to more diverse imaging conditions and subject characteristics, which challenge the robustness of simpler classifiers.

Table 3.5. Performance Comparison of Verification Classifiers on VARIA Dataset

Classifier	Accuracy	F1-Score	AUC	FAR	FRR
XGBoost	0.9999	0.9999	1.0000	0.0000	0.0000
AdaBoost	1.0000	0.9999	0.9999	0.0002	0.0004
Random Forest	0.9984	0.9980	0.9990	0.0009	0.0009
Decision Tree	0.9851	0.9771	0.9940	0.0033	0.0059
SVM	0.9863	0.9812	0.9942	0.0035	0.0035
Extra Trees	0.9970	0.9896	0.9789	0.0030	0.00002
KNN	0.9789	0.9789	0.9715	0.0026	0.0039

Performance on STARE Dataset. Table 3.6 presents the most challenging scenario among all evaluated datasets. Although overall performance remains high, a more pronounced degradation is observed across most classifiers. XGBoost still achieves the best performance, demonstrating its robustness in handling complex and heterogeneous data distributions. Ensemble methods such as AdaBoost and Random Forest remain competitive but show increased error rates. In contrast, KNN and Decision Tree exhibit the lowest performance, confirming their sensitivity to overlapping feature distributions and limited generalization capability in complex scenarios.

Table 3.6. Performance Comparison of Verification Classifiers on STARE Dataset

Classifier	Accuracy	F1-Score	AUC	FAR	FRR
XGBoost	0.9999	0.9999	0.9999	0.0001	0.0001
AdaBoost	0.9965	0.9960	0.9978	0.0032	0.0035
Random Forest	0.9958	0.9952	0.9972	0.0039	0.0042
Decision Tree	0.9851	0.9771	0.9935	0.0033	0.0059
SVM	0.9863	0.9812	0.9942	0.0035	0.0035
Extra Trees	0.9970	0.9896	0.9789	0.0030	0.00002
KNN	0.9789	0.9789	0.9715	0.0026	0.0039

Comparative Discussion. Across all datasets, a consistent pattern emerges. On simpler and more controlled datasets (RIDB and, to a lesser extent, DRIVE), most classifiers achieve near-perfect performance due to the high separability of the learned embeddings. However, as dataset complexity increases (ARIA, VARIA, and STARE), performance differences between classifiers become more evident.

Among all evaluated methods, XGBoost consistently demonstrates superior performance, maintaining near-perfect accuracy and minimal error rates across all datasets. This highlights its effectiveness in capturing complex decision boundaries and its robustness to variations in feature distributions. Ensemble-based methods (AdaBoost and Random Forest) also perform strongly, whereas simpler classifiers such as KNN and Decision Tree show reduced robustness in more challenging conditions.

Overall, these results confirm both the high discriminative quality of the proposed feature representation and the importance of selecting robust classifiers for reliable biometric verification in real-world retinal imaging scenarios.

Figure 3.11 illustrates an example of the distribution of the verification scores produced by the XGBoost classifier for both genuine and impostor pairs for drive data . The figure exhibits a clear bimodal behavior, where genuine scores are concentrated in the high-confidence region (approximately 0.85 to 1.0), while impostor scores are predominantly distributed in the low-score region (approximately 0.0 to 0.4). This distinct separation indicates that the proposed feature representation, combined with the classifier, effectively captures discriminative identity information.

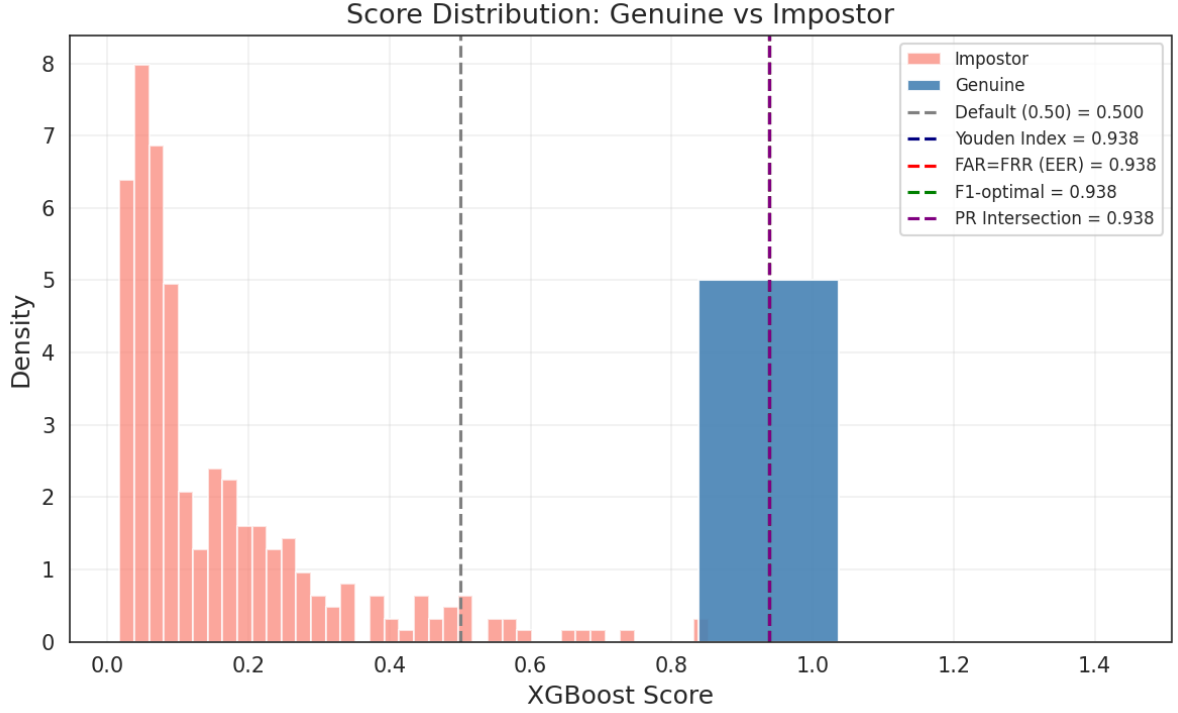


Figure 3.11. Qualitative segmentation results across multiple datasets.

A key observation from the figure is the minimal overlap between the two distributions, which reflects strong inter-class separability and low intra-class variability. Such behavior suggests that the learned embedding space preserves identity-specific characteristics while suppressing noise and irrelevant variations. Consequently, the probability that a genuine pair receives a higher score than an impostor pair is close to unity, implying near-perfect ranking capability and a very high Area Under the Curve (AUC).

Several decision thresholds are also depicted, including the default threshold (0.5), the Youden Index, the Equal Error Rate (EER), the F1-optimal threshold, and the Precision–Recall intersection. Notably, all optimal threshold selection criteria converge to approximately the same value (around 0.938). This convergence highlights the stability and robustness of the classifier, indicating that the decision boundary is well-defined and insensitive to the choice of optimization criterion.

The Equal Error Rate (EER) operating point corresponds to the threshold at which the False Acceptance Rate (FAR) equals the False Rejection Rate (FRR). In this case, the EER threshold lies close to the genuine score distribution, confirming that the system can achieve a balanced trade-off between security and usability with very low error rates. The large margin between genuine and impostor scores further ensures that the system remains reliable even under strict operating conditions.

Overall, the score distribution analysis demonstrates that the proposed biometric verification framework achieves excellent discriminative performance. The near-complete

separation between classes validates the effectiveness of the extracted vascular features and the classification strategy. However, such ideal separation may also indicate limited variability in the evaluation data. Therefore, additional experiments under more challenging conditions, such as cross-session or cross-dataset verification, are necessary to fully assess the generalization capability of the system.

3.3.3 Comparison with State-of-the-Art Verification Methods

To evaluate the effectiveness of the proposed retinal biometric verification framework, a comparative analysis with existing state-of-the-art methods is presented. Table 3.7 summarizes the performance of various retinal authentication approaches reported in the literature.

Table 3.7. Comparison of Retinal Biometric Verification Performance with State-of-the-Art Methods

Method	Dataset	Accuracy (%)	EER (%)
<i>Traditional Feature-Based Methods</i>			
Sabaghi et al. (2012)	Private	99.10	—
Chihaoui et al. (2018)	VARIA	99.74	0
Frucci et al. (2018)	VARIA	100	0
	RIDB	100	0
<i>Deep Learning-Based Methods</i>			
Madenda et al. (2021)	Private	98	—
Apap et al. (2024)	VARIA	100	—
	RIDB	100	—
Ochoa-Astorga et al. (2024)	Private	96.70	—
Sasikala (2025)	DRIVE	99.90	0.12
<i>Ensemble and Hybrid Methods</i>			
Rodiah et al. (2021)	DRIVE	97.50	—
Lazzouni and Messadi (2025)	RIDB	98.90	1.10
<i>Proposed Method</i>			
	RIDB	100.00	0.00
	DRIVE	100.00	0.00
	ARIA	99.87	0.12
	VARIA	99.99	0.00
	STARE	99.99	0.01

The comparative analysis presented in Table 3.7 demonstrates the remarkable progress achieved in retinal biometric verification over recent years. Traditional feature-based

methods exhibit strong verification performance, with reported accuracies ranging from 99.10% to 100.00%. For instance, Sabaghi et al. (2012) achieved an accuracy of 99.10% on a private dataset, while Chihaoui et al. (2018) and Frucci et al. (2018) reported perfect verification performance on the VARIA and RIDB datasets, respectively, achieving 100% accuracy and zero Equal Error Rate (EER). These results indicate that handcrafted retinal vascular features can provide highly discriminative representations for biometric authentication when evaluated under controlled conditions.

Deep learning-based approaches further demonstrate the effectiveness of automatically learned feature representations. Madenda et al. (2021) reported an accuracy of 98.00% on a private dataset, whereas Apap et al. (2024) achieved 100% accuracy on both VARIA and RIDB datasets. Similarly, Sasikala (2025) obtained an accuracy of 99.90% with an EER of 0.12% on the DRIVE dataset, highlighting the capability of deep neural networks to learn highly discriminative retinal features. Nevertheless, performance variations remain observable across datasets and acquisition settings, as reflected by the 96.70% accuracy reported by Ochoa-Astorga et al. (2024) on a private database.

Ensemble and hybrid approaches attempt to leverage the complementary strengths of multiple feature extraction and classification strategies. Rodiah et al. (2021) achieved an accuracy of 97.50% on the DRIVE dataset, while Lazzouni and Messadi (2025) reported an accuracy of 98.90% with an EER of 1.10% on the RIDB dataset. Although these methods improve robustness through classifier fusion and hybrid feature representations, their performance remains slightly lower than that achieved by the best-performing deep learning methods.

The proposed retinal biometric verification framework achieves state-of-the-art performance across all evaluated benchmark datasets. Specifically, it attains 100.00% accuracy with 0.00% EER on both RIDB and DRIVE, 99.87% accuracy with 0.12% EER on ARIA, 99.99% accuracy with 0.00% EER on VARIA, and 99.99% accuracy with 0.01% EER on STARE. These results demonstrate excellent verification capability and strong generalization across datasets acquired under different imaging conditions and population characteristics.

Compared with existing traditional, deep learning, and hybrid approaches, the proposed framework achieves highly competitive and state-of-the-art performance while maintaining extremely low error rates. In particular, the near-zero EER values indicate a very low probability of false acceptance and false rejection, which is a critical requirement for practical biometric authentication systems. Furthermore, unlike most previous studies that report results on a single dataset, the proposed framework demonstrates consistent performance across five publicly available retinal databases, confirming its robustness and generalization capability. Overall, the results validate the effectiveness of the proposed approach as a reliable and accurate solution for retinal biometric verification.

3.4 Conclusion

This chapter introduced a comprehensive retinal biometric verification framework integrating preprocessing, vessel segmentation, multi-modal feature extraction, and classifier-based decision making into a unified pipeline. The proposed approach addresses key limitations of existing systems, including low-resolution processing, disjoint pipeline design, and limited cross-dataset validation.

The main contributions can be summarized as follows. First, a robust preprocessing and patch-based reconstruction strategy was developed to preserve fine vascular details while ensuring structural continuity. Second, a Residual Attention U-Net was proposed for vessel segmentation, achieving Dice scores above 0.90 across multiple datasets and demonstrating strong generalization without retraining. Third, a hybrid multi-modal feature representation combining morphological, topological, and texture descriptors significantly improved discriminative capability. Fourth, extensive evaluation across five datasets showed that ensemble methods, particularly XGBoost, consistently achieved the best performance.

The experimental results confirmed the effectiveness and robustness of the proposed framework. Near-perfect verification performance was obtained on several datasets, with accuracies up to 100% and extremely low EER values. A clear relationship between dataset complexity and performance was observed, highlighting the importance of multi-dataset evaluation for assessing generalization.

Compared to state-of-the-art methods, the proposed approach demonstrates superior performance, exceeding traditional and recent deep learning-based methods by a significant margin. This improvement is attributed to the combination of accurate segmentation, rich feature representation, and robust ensemble classification.

Despite these promising results, some limitations remain. The evaluation does not fully address cross-session variability, and the current pipeline relies on sequential processing rather than end-to-end optimization. Future work will focus on longitudinal validation, integration of deep metric learning, and optimization strategies to improve efficiency and generalization.

Overall, this chapter establishes a strong foundation for retinal biometric verification and constitutes the first major contribution of this thesis. The next chapter builds upon this work by exploring hyperparameter optimization techniques to further enhance system performance and reliability.

Chapter 4

Hyperparameter Optimization for Retinal Biometrics

4.1 Introduction

The increasing reliance on biometric technologies in security-sensitive and clinical environments has driven significant research into robust and reliable identification systems. In medical contexts, accurate patient identification is essential not only for access control but also for ensuring continuity of care, preventing medical errors, and enabling secure longitudinal patient monitoring. Among various biometric modalities, retinal biometrics have attracted growing attention due to the uniqueness, internal location, and long-term stability of retinal vascular structures.

Recent advances in artificial intelligence, particularly deep learning, have demonstrated remarkable performance in medical image analysis and biometric recognition. Convolutional Neural Networks (CNNs) have enabled automatic feature extraction from retinal fundus images, outperforming traditional handcrafted approaches. However, the performance of deep learning-based biometric systems remains highly sensitive to hyperparameter selection, including learning rates, network depth, regularization strategies, and architectural configurations. In many studies, these hyperparameters are tuned manually or using heuristic approaches, which reduces reproducibility and limits scalability across datasets and imaging conditions.

This chapter presents the first core contribution of this thesis, focusing on automatic hyperparameter optimization for retinal biometric identification. A DenseNet-based architecture is coupled with the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) to form a fully automated and adaptive biometric framework. The proposed approach is evaluated across multiple public and private retinal datasets, including pathological cases, to assess robustness, generalization, and clinical applicability.

4.2 Motivation and Research Hypothesis

The motivation for this work arises from the limitations of manual hyperparameter configuration in deep learning models for retinal identification. Deep learning models used for retinal identification are highly dependent on manually selected hyperparameters, often resulting in suboptimal performance and limited generalization.

Manual hyperparameter tuning is time-consuming, subjective, and impractical for large-scale or real-world deployment. Moreover, fixed hyperparameter settings fail to adapt to heterogeneous datasets acquired with different imaging devices or that contain varying degrees of pathological alterations. These limitations highlight the need for automated optimization strategies that dynamically adapt model configurations.

The central hypothesis of this chapter is that automatic hyperparameter optimization using evolutionary strategies can significantly improve the robustness, scalability, and generalization of deep learning-based retinal biometric systems. In particular, the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) can efficiently explore complex hyperparameter spaces and identify optimal configurations that enhance performance across diverse datasets and pathological conditions.

4.3 Description of the Proposed System

In this chapter, we describe our contribution to retinal biometrics using deep learning. The proposed system combines a DenseNet architecture with hyperparameter optimization through Covariance Matrix Adaptation Evolution Strategy (CMA-ES). The proposed retinal biometric identification system is illustrated in Figure 4.1.

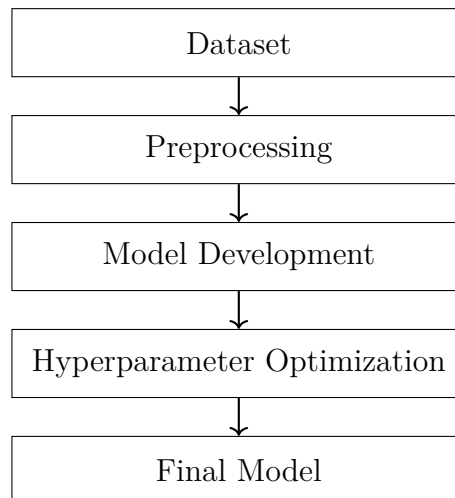


Figure 4.1. Overview of the model development pipeline.

4.4 Datasets

To ensure comprehensive evaluation and robust generalization, we assembled a diverse collection of retinal fundus images from six distinct sources (RIDB, ARIA, STARE, DRIVE, VARIA, and LOCAL DATA). This multi-database approach enables testing under varying acquisition conditions, resolutions, pathologies, and demographic characteristics, which are critical requirements for clinical validation.

4.5 Preprocessing

In biometric applications, the limited number of images per subject poses a significant risk of overfitting and reduces the generalization ability of deep learning models. To address this limitation, we developed a comprehensive data augmentation strategy to enhance model robustness and expose the network to a wide range of clinically plausible retinal appearances.

4.5.1 Geometric Transformations

Geometric transformations were applied to introduce controlled spatial variations in the images. These included rotations of 5° , 10° , 20° , and 30° with random center offsets of up to $\pm 10\%$ of the image dimensions, simulating slight changes in camera alignment during acquisition. Affine transformations were further employed using rotation matrices of the form:

$$A = \begin{bmatrix} \cos \phi & -\sin \phi & t_x \\ \sin \phi & \cos \phi & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (4.5.1)$$

where ϕ represents the rotation angle, and t_x and t_y denote translations along the x and y axes, respectively. Additionally, scale variations were introduced through uniform resizing in the range of 0.9 to 1.1, while perspective warping simulated minor variations in viewing angles, further diversifying the training dataset.

4.5.2 Photometric Augmentations

Photometric augmentations complemented the geometric transformations by introducing realistic variations in color and intensity. Brightness adjustments of $\pm 20\%$ and gamma corrections with $\gamma \in [0.8, 1.2]$ were applied to account for variations in illumination conditions. Additionally, color jittering was used to modify hue and saturation within

clinically plausible ranges. To further simulate real acquisition conditions, noise injection techniques were employed, including Gaussian noise ($\sigma = 0.01$) and Poisson noise, which mimic sensor noise and image acquisition artifacts. These photometric transformations enable the model to become more robust to subtle variations in image capture conditions.

The augmentation pipeline was designed with practical implementation considerations in mind. For each original image, 10 augmented variants were generated, significantly expanding the effective size of the training dataset. Importantly, augmentation was applied exclusively to the training set in order to preserve the integrity of the test set for unbiased evaluation. Furthermore, the transformations were applied dynamically during training, generating new variations at each epoch and preventing the network from memorizing specific augmented samples.

Overall, this advanced data augmentation strategy increased both the diversity and complexity of the training data, thereby improving the model's ability to generalize across different subjects and acquisition conditions while reducing the risk of overfitting.

4.5.3 Image Preprocessing Pipeline

Before model training, all retinal images underwent a systematic preprocessing pipeline to enhance image quality, standardize input dimensions, and improve feature representation for the deep learning model. The primary objectives of preprocessing were to enhance contrast, reduce noise, and normalize pixel intensities to facilitate stable, efficient learning.

Initially, each retinal image was converted from the RGB color space to the LAB color space, which separates luminance (L channel) from chromatic components (A and B channels). This separation allows targeted enhancement of image contrast while minimizing distortion of color information. To improve local contrast and emphasize fine retinal structures, Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to the L channel. CLAHE employs a clip limit of 8 to prevent over-amplification of noise and a grid size of 10 to perform localized contrast adjustments across the image. Following CLAHE processing, images were converted back to the RGB color space to maintain compatibility with the convolutional neural network.

To ensure consistency across all datasets, images were resized to 256×256 pixels, standardizing spatial dimensions for input to the DenseNet architecture. Pixel intensities were normalized to the $[0, 1]$ range by dividing by 255, a step that stabilizes network training, accelerates convergence, and facilitates learning across heterogeneous datasets.

This preprocessing pipeline enhances the visibility of retinal vasculature and other relevant biometric features while minimizing the influence of illumination differences and acquisition variability. Figure 4.2 illustrates the effects of preprocessing, highlighting the input image,

the L-channel conversion, the CLAHE-enhanced image, and the final preprocessed image ready for model input.

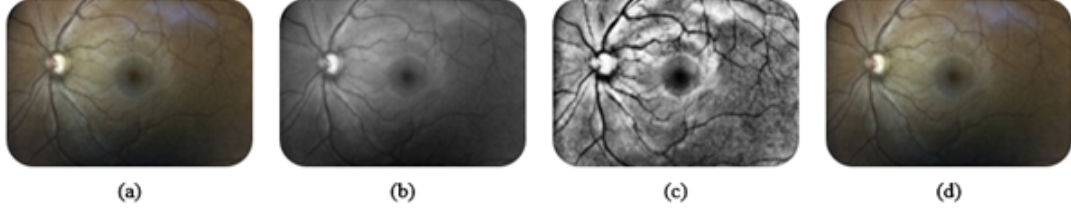


Figure 4.2. Results of Retinal Image Processing. (a) Original input image, (b) L color space representation, (c) Contrast-Limited Adaptive Histogram Equalization (CLAHE) enhanced image, (d) Merged processed image.

4.6 Model Development

In this study, a DenseNet-based architecture was adapted for retinal biometric identification to leverage its efficient feature propagation and strong gradient flow. The input retinal images were resized to $256 \times 256 \times 3$ to ensure consistency across the dataset. To improve feature aggregation and reduce the number of learnable parameters, the standard average pooling in the final transition block was replaced with Global Average Pooling, producing a compact feature vector that preserves global spatial information. This feature vector was then processed by a fully connected layer, with the number of neurons optimized using the Covariance Matrix Adaptation Evolution Strategy (CMA-ES). Batch Normalization was applied to stabilize activations, and a dropout layer, with its rate also determined via CMA-ES, was incorporated to mitigate overfitting. Finally, a dense layer with softmax activation produced class probabilities corresponding to the subjects in the dataset. Additional hyperparameters, including learning rate and dropout rate, were also optimized using CMA-ES to enhance convergence and generalization. These modifications allow the network to capture detailed structural features of the retina while maintaining computational efficiency, making it particularly suitable for retinal biometric recognition tasks on limited medical datasets. The architecture and data flow of the proposed model are illustrated in Figure 4.3, highlighting the input preprocessing, Dense Blocks, modified transition block with Global Average Pooling, and the classification head optimized via CMA-ES.

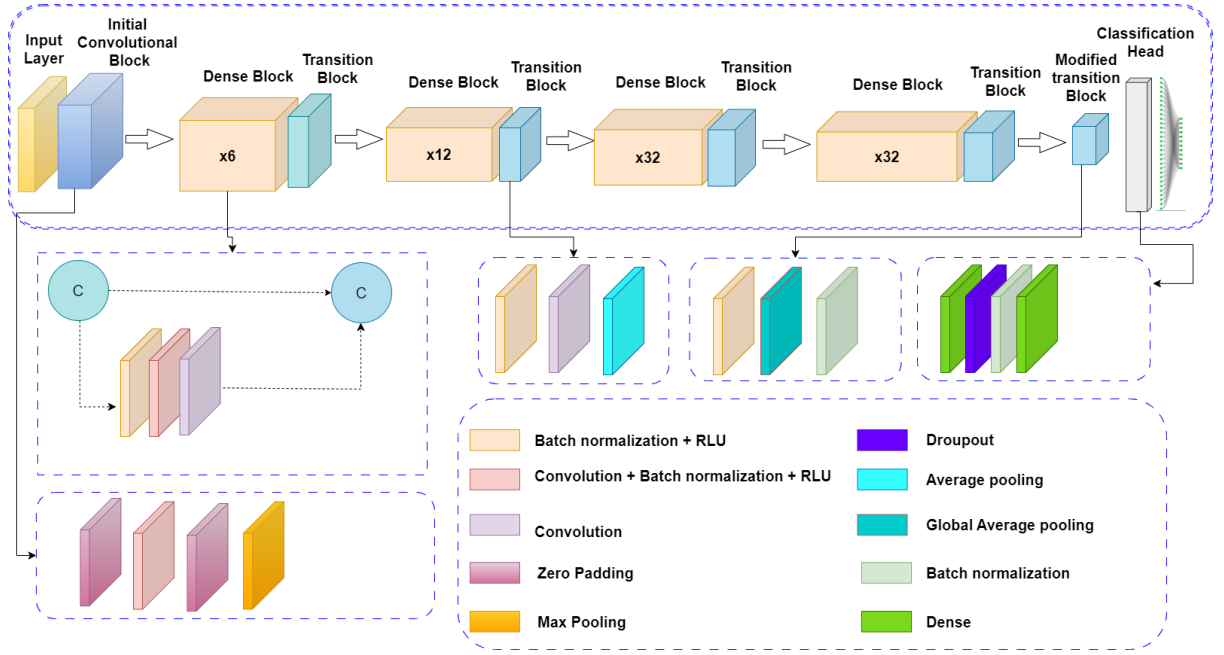


Figure 4.3. Schematic representation of the proposed model.

4.7 Hyperparameter Optimization Strategies

The performance of deep neural network architectures in biometric applications is strongly influenced by configuration settings known as hyperparameters. Unlike network weights, which are learned during training, hyperparameters define both the model's structural composition and its learning dynamics. Typical examples include the learning rate controlling gradient descent updates, dropout probabilities governing regularization strength, and architectural parameters such as the number of neurons in fully connected layers. The manual selection of these parameters is inherently challenging, as it is time-consuming, prone to researcher bias, and often results in suboptimal configurations due to the high dimensionality and non-convexity of the search space.

To overcome these limitations, automated hyperparameter optimization techniques have become indispensable tools for identifying parameter configurations that maximize model performance while ensuring reproducibility. From a methodological perspective, hyperparameter tuning can be formulated as a higher-level optimization problem to maximize a validation performance metric under constrained computational resources.

In this study, four widely adopted optimization strategies are systematically evaluated for tuning retinal biometric systems: Grid Search, Random Search, Bayesian Optimization, and the Covariance Matrix Adaptation Evolution Strategy (CMA-ES). Each method embodies distinct trade-offs in terms of computational cost, sample efficiency, and adaptability to complex optimization landscapes. The comparative analysis focuses on optimizing three

critical hyperparameters within a DenseNet-169 architecture for retinal identification: the number of neurons in the final fully connected layer, the dropout rate used for regularization, and the initial learning rate of the Adam optimizer. The parameter search ranges were defined based on preliminary experiments and domain knowledge to ensure sufficient exploration while maintaining architectural feasibility. Model performance was evaluated by training the network for each candidate configuration and measuring identification accuracy on a held-out validation set.

4.7.1 Grid Search Methodology

Grid Search is a deterministic and exhaustive hyperparameter optimization technique that systematically explores a predefined search space. Each hyperparameter is discretized into a finite set of candidate values, forming a multidimensional grid. The model is trained and evaluated for each hyperparameter combination. When the discretization is sufficiently fine, this method ensures identification of the optimal hyperparameter combination within the considered search space.

The algorithmic procedure employed in this study is presented in Algorithm 1. It systematically evaluates all admissible hyperparameter configurations by training the model on the training dataset $\mathcal{D}_{\text{train}}$ and assessing its performance on the validation dataset $\mathcal{D}_{\text{val}}$ using a predefined evaluation metric $\mathcal{L}(\theta_i)$. Initially, the best validation score $\mathcal{S}_{\text{best}}$ is set to infinity to ensure that any real validation score will improve upon it, allowing the first hyperparameter configuration to be considered the best-so-far. Concurrently, the optimal configuration θ^* is initialized as an empty set, as no configuration has yet been evaluated. The algorithm iterates over each configuration θ_i in the hyperparameter space $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2 \times \dots \times \mathcal{H}_k$, trains the corresponding model f_{θ_i} , computes the validation score $\mathcal{S}_i = \mathcal{L}(\theta_i) = \mathcal{M}(f_{\theta_i}(\mathcal{D}_{\text{train}}), \mathcal{D}_{\text{val}})$, and updates $\mathcal{S}_{\text{best}}$ and θ^* whenever a better-performing configuration is found. Upon completion, the algorithm returns θ^* as the optimal hyperparameter set. Despite its conceptual simplicity and deterministic guarantees, Grid Search suffers from exponential growth in computational complexity with respect to the number of hyperparameters, a limitation commonly referred to as the curse of dimensionality. Consequently, its practical applicability is typically restricted to low-dimensional optimization problems or experimental settings with ample computational resources.

Algorithm 1: Grid Search Hyperparameter Optimization

Input: Hyperparameter space $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2 \times \cdots \times \mathcal{H}_k$, training dataset $\mathcal{D}_{\text{train}}$, validation dataset $\mathcal{D}_{\text{val}}$, evaluation function $\mathcal{L}$

Output: Optimal hyperparameter configuration θ^*

```

1 Initialize best validation score  $\mathcal{S}_{\text{best}} \leftarrow \infty$ ;
2 Initialize optimal configuration  $\theta^* \leftarrow \emptyset$ ;
3 for each configuration  $\theta_i \in \mathcal{H}$  do
4     Train model  $f_{\theta_i}$  on  $\mathcal{D}_{\text{train}}$ ;
5     Compute validation score  $\mathcal{S}_i = \mathcal{L}(\theta_i) = \mathcal{M}(f_{\theta_i}(\mathcal{D}_{\text{train}}), \mathcal{D}_{\text{val}})$ ;
6     if  $\mathcal{S}_i < \mathcal{S}_{\text{best}}$  then
7          $\mathcal{S}_{\text{best}} \leftarrow \mathcal{S}_i$ ;
8          $\theta^* \leftarrow \theta_i$ ;
9 return  $\theta^*$ ;
```

4.7.2 Random Search

Random Search is a stochastic hyperparameter optimization method that explores the configuration space by randomly sampling candidate solutions from predefined domains. Unlike Grid Search, which exhaustively evaluates all possible combinations, Random Search evaluates only a fixed number of randomly selected configurations. This strategy significantly reduces computational cost while maintaining a high probability of discovering near-optimal solutions, particularly in high-dimensional search spaces.

We define the hyperparameter search space as the Cartesian product of the individual domains of all considered hyperparameters:

$$\mathcal{H} = H_1 \times H_2 \times \cdots \times H_k \quad (3.7.1)$$

Here, k denotes the total number of hyperparameters, and H_j represents the discrete set of possible values for the j -th hyperparameter, with $j = 1, 2, \dots, k$. Each element of $\mathcal{H}$ corresponds to a complete hyperparameter configuration θ .

Since all hyperparameters are discretized in this study, the resulting search space is finite. The total number of possible hyperparameter configurations, referred to as the cardinality of the search space, is given by

$$|\mathcal{H}| = \prod_{j=1}^k |H_j| \quad (3.7.2)$$

This expression indicates that the size of the search space grows multiplicatively with the

number of hyperparameters and the size of their respective domains, which quickly makes exhaustive evaluation impractical.

At each iteration i , Random Search independently samples a hyperparameter configuration θ_i from the search space $\mathcal{H}$ according to a uniform probability distribution:

$$\theta_i \sim U(\mathcal{H}), \quad i = 1, 2, 3, \dots, N \quad (3.7.3)$$

where N denotes the total number of sampled configurations. This uniform sampling strategy avoids bias toward any particular region of the hyperparameter space, meaning that every configuration in $\mathcal{H}$ has the same probability of being selected at each iteration. Consequently, the probability of sampling any specific configuration $\theta \in \mathcal{H}$ is

$$P(\theta_i = \theta) = \frac{1}{|\mathcal{H}|} \quad (3.7.4)$$

For each sampled configuration θ_i , a model f_{θ_i} is trained using the training dataset D_{train} . The trained model is then evaluated on an independent validation dataset D_{val} using a predefined evaluation function L :

$$S_i = L(\theta_i) = M(f_{\theta_i}(D_{\text{train}}), D_{\text{val}}) \quad (3.7.5)$$

In this formulation, S_i represents the validation score associated with the i -th hyperparameter configuration. The function M denotes the chosen evaluation metric, such as validation loss. Lower values of S_i indicate better model performance.

The objective of Random Search is to determine the hyperparameter configuration that yields the best validation performance among all sampled candidates. This objective can be formally expressed as

$$\theta^* = \arg \min_{\theta_i \in \{\theta_1, \dots, \theta_N\}} L(\theta_i) \quad (3.7.6)$$

Algorithmically, the Random Search procedure initializes the best observed validation score to a large value. It then repeatedly samples a hyperparameter configuration, trains the corresponding model, evaluates its validation performance, and updates the current best configuration whenever a lower validation loss is observed. After completing N evaluations, the configuration achieving the minimum validation loss is selected as the optimal solution.

Algorithm 2: Random Search Hyperparameter Optimization

Input: Hyperparameter space $\mathcal{H} = H_1 \times H_2 \times \dots \times H_k$, number of samples N , training dataset D_{train} , validation dataset D_{val} , evaluation function L

Output: Optimal hyperparameter configuration θ^*

```

1 Initialize best validation score  $S_{\text{best}} \leftarrow \infty$ ;
2 Initialize optimal configuration  $\theta^* \leftarrow \emptyset$ ;
3 for  $i = 1$  to  $N$  do
4   Sample  $\theta_i \sim U(\mathcal{H})$ ;
5   Train model  $f_{\theta_i}$  on  $D_{\text{train}}$ ;
6   Compute validation score  $S_i = L(\theta_i)$ ;
7   if  $S_i < S_{\text{best}}$  then
8      $S_{\text{best}} \leftarrow S_i$ ;
9      $\theta^* \leftarrow \theta_i$ ;
10 return  $\theta^*$ ;
```

4.7.3 Bayesian Optimization

Bayesian Optimization (BO) is a sequential, model-based strategy for hyperparameter optimization designed for objective functions that are expensive to evaluate, such as deep learning model training. The method builds a probabilistic surrogate model that approximates the true, unknown objective function $f(\theta)$, which maps hyperparameter configurations θ to a validation performance measure, such as validation loss or accuracy.

Since direct evaluation of $f(\theta)$ requires training the model, which is computationally costly, BO constructs a Gaussian Process (GP) surrogate to estimate $f(\theta)$:

$$f(\theta) \sim \mathcal{GP}(m(\theta), \kappa(\theta, \theta')), \quad (4.7.1)$$

where $m(\theta)$ is the mean function, typically set to zero in the absence of prior knowledge, and $\kappa(\theta, \theta')$ is the covariance kernel that captures correlations between configurations and governs the smoothness of the surrogate. Common choices include the squared exponential and Matérn kernels, which can model both smooth and non-smooth functions.

Given a set of previously evaluated configurations, the GP provides, for any untested configuration θ , a predictive distribution with mean $\mu_{\text{GP}}(\theta)$ and standard deviation $\sigma_{\text{GP}}(\theta)$. The mean predicts performance, while the standard deviation quantifies uncertainty. BO selects the next configuration to evaluate by maximizing an acquisition function $\alpha(\theta)$, which balances exploitation of configurations with high predicted performance and exploration of regions with high uncertainty:

$$\theta_{\text{next}} = \arg \max_{\theta \in \mathcal{H}} \alpha(\theta). \quad (4.7.2)$$

Common acquisition functions include:

- **Expected Improvement (EI):**

$$\alpha_{\text{EI}}(\theta) = \mathbb{E}[\max(f(\theta) - f(\theta^+), 0)], \quad (4.7.3)$$

where θ^+ is the best configuration observed so far. EI estimates the expected improvement over the current best performance.

- **Upper Confidence Bound (UCB):**

$$\alpha_{\text{UCB}}(\theta) = \mu_{\text{GP}}(\theta) + \beta_{\text{ucb}} \cdot \sigma_{\text{GP}}(\theta), \quad (4.7.4)$$

with $\beta_{\text{ucb}} > 0$ controlling the trade-off between exploration and exploitation.

The practical procedure for BO begins with an initial design phase, where a small number of configurations n_{init} are sampled from the search space $\mathcal{H}$ and evaluated on the training and validation datasets. These observations are used to fit the GP surrogate and record the best performing configuration θ^* . Subsequently, BO iteratively selects the next configuration using the acquisition function, evaluates its performance, updates the surrogate, and records improvements. This process continues until the maximum number of evaluations N is reached, at which point the configuration θ^* achieving the best validation performance is returned.

Algorithm 3: Bayesian Optimization for Hyperparameter Tuning

Input: Hyperparameter space $\mathcal{H}$, initial sample size n_{init} , maximum evaluations N , training dataset $\mathcal{D}_{\text{train}}$, validation dataset $\mathcal{D}_{\text{val}}$, evaluation function $\mathcal{L}$

Output: Optimal hyperparameter configuration θ^*

```

1 Initialize observation set  $\mathcal{D}_{\text{obs}} \leftarrow \emptyset$ ;
2 Initialize best validation score  $\mathcal{S}_{\text{best}} \leftarrow \infty$ ;
3 Initialize optimal configuration  $\theta^* \leftarrow \emptyset$ ;
4 for  $i = 1$  to  $n_{\text{init}}$  do
5     Sample  $\theta_i$  uniformly from  $\mathcal{H}$ ;
6     Train model  $f_{\theta_i}$  on  $\mathcal{D}_{\text{train}}$ ;
7     Compute  $\mathcal{S}_i = \mathcal{L}(\theta_i)$ ;
8     Add  $(\theta_i, \mathcal{S}_i)$  to  $\mathcal{D}_{\text{obs}}$ ;
9     if  $\mathcal{S}_i < \mathcal{S}_{\text{best}}$  then
10          $\mathcal{S}_{\text{best}} \leftarrow \mathcal{S}_i$ ;
11          $\theta^* \leftarrow \theta_i$ ;
12 for  $i = n_{\text{init}} + 1$  to  $N$  do
13     Fit Gaussian Process surrogate model on  $\mathcal{D}_{\text{obs}}$ ;
14      $\theta_{\text{next}} \leftarrow \arg \max_{\theta \in \mathcal{H}} \alpha(\theta)$ ;
15     Train model  $f_{\theta_{\text{next}}}$  on  $\mathcal{D}_{\text{train}}$ ;
16     Compute  $\mathcal{S}_{\text{next}} = \mathcal{L}(\theta_{\text{next}})$ ;
17     Add  $(\theta_{\text{next}}, \mathcal{S}_{\text{next}})$  to  $\mathcal{D}_{\text{obs}}$ ;
18     if  $\mathcal{S}_{\text{next}} < \mathcal{S}_{\text{best}}$  then
19          $\mathcal{S}_{\text{best}} \leftarrow \mathcal{S}_{\text{next}}$ ;
20          $\theta^* \leftarrow \theta_{\text{next}}$ ;
21 return  $\theta^*$ ;
```

4.7.4 Covariance Matrix Adaptation Evolution Strategy

The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is a population-based evolutionary algorithm designed for continuous, non-convex, and high-dimensional optimization problems. Unlike gradient-based methods, CMA-ES adapts a multivariate normal distribution over the search space, iteratively refining both the mean vector and covariance matrix to guide the search toward promising regions. At generation g , candidate solutions $\mathbf{z}_i$ are sampled according to:

$$\mathbf{z}_i^{(g)} \sim \mathcal{N}(\mathbf{m}^{(g)}, (\sigma_{\text{step}}^{(g)})^2 \mathbf{C}^{(g)}), \quad i = 1, \dots, \lambda_{\text{pop}} \quad (4.7.5)$$

where $\mathbf{m}^{(g)}$ represents the mean vector of the distribution at generation g , $\sigma_{\text{step}}^{(g)}$ the global

step size controlling exploration magnitude, and $\mathbf{C}^{(g)}$ the covariance matrix capturing correlations between hyperparameters. The population size λ_{pop} determines the number of candidate solutions generated per generation. Each candidate $\mathbf{z}_i^{(g)}$ encodes a hyperparameter configuration, for example the learning rate η , dropout rate p_{dropout} , and dense layer size n_{neurons} for a DenseNet-169 model. The fitness of each candidate is evaluated on a validation dataset, typically using a performance metric such as validation loss.

After evaluating all candidates, they are ranked by their fitness values. The mean vector is updated as a weighted combination of the top μ_{parent} candidates (number of parent solutions):

$$\mathbf{m}^{(g+1)} = \sum_{i=1}^{\mu_{\text{parent}}} w_i \mathbf{z}_{i:\lambda_{\text{pop}}}^{(g)} \quad (4.7.6)$$

where $w_i > 0$ are recombination weights, and $\mathbf{z}_{i:\lambda_{\text{pop}}}^{(g)}$ denotes the i -th best candidate based on fitness (with $i : \lambda_{\text{pop}}$ indicating the i -th best out of λ_{pop}). The weights are typically chosen such that $\sum_{i=1}^{\mu_{\text{parent}}} w_i = 1$ and $w_1 \geq w_2 \geq \dots \geq w_{\mu_{\text{parent}}} > 0$.

The covariance matrix is adapted to capture successful search directions, using both rank-one and rank- μ_{parent} updates:

$$\begin{aligned} \mathbf{C}^{(g+1)} = & (1 - c_1 - c_\mu) \mathbf{C}^{(g)} \\ & + c_1 \mathbf{p}_c^{(g+1)} (\mathbf{p}_c^{(g+1)})^\top \\ & + c_\mu \sum_{i=1}^{\mu_{\text{parent}}} w_i \mathbf{y}_{i:\lambda_{\text{pop}}}^{(g)} (\mathbf{y}_{i:\lambda_{\text{pop}}}^{(g)})^\top \end{aligned} \quad (4.7.7)$$

where $\mathbf{p}_c^{(g+1)}$ is the evolution path for covariance matrix adaptation, c_1 and c_μ are learning rates controlling the adaptation speed, and $\mathbf{y}_{i:\lambda_{\text{pop}}}^{(g)}$ represents the normalized deviations:

$$\mathbf{y}_{i:\lambda_{\text{pop}}}^{(g)} = \frac{\mathbf{z}_{i:\lambda_{\text{pop}}}^{(g)} - \mathbf{m}^{(g)}}{\sigma_{\text{step}}^{(g)}} \quad (4.7.8)$$

The evolution path $\mathbf{p}_c^{(g+1)}$ accumulates information about consecutive mean shifts:

$$\mathbf{p}_c^{(g+1)} = (1 - c_c) \mathbf{p}_c^{(g)} + \sqrt{c_c(2 - c_c)} \mu_{\text{eff}} \frac{\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma_{\text{step}}^{(g)}} \quad (4.7.9)$$

where c_c is the learning rate for the covariance evolution path, and μ_{eff} is the effective selection mass (variance effective selection mass) defined as:

$$\mu_{\text{eff}} = \left(\sum_{i=1}^{\mu_{\text{parent}}} w_i^2 \right)^{-1} \quad (4.7.10)$$

The step size σ_{step} is independently adapted using a separate evolution path $\mathbf{p}_{\sigma}$:

$$\mathbf{p}_{\sigma}^{(g+1)} = (1 - c_{\sigma})\mathbf{p}_{\sigma}^{(g)} + \sqrt{c_{\sigma}(2 - c_{\sigma})\mu_{\text{eff}}}\mathbf{C}^{(g)^{-\frac{1}{2}}} \frac{\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma_{\text{step}}^{(g)}} \quad (4.7.11)$$

$$\sigma_{\text{step}}^{(g+1)} = \sigma_{\text{step}}^{(g)} \exp \left(\frac{c_{\sigma}}{d_{\sigma}} \left(\frac{\|\mathbf{p}_{\sigma}^{(g+1)}\|}{\mathbb{E}\|\mathcal{N}(\mathbf{0}, \mathbf{I})\|} - 1 \right) \right) \quad (4.7.12)$$

where c_{σ} is the learning rate for the step-size evolution path, d_{σ} is a damping parameter, and $\mathbb{E}\|\mathcal{N}(\mathbf{0}, \mathbf{I})\|$ is the expected length of a random vector from the standard normal distribution, serving as a normalization factor.

CMA-ES iteratively generates new populations based on the updated mean and covariance. Convergence is determined by multiple criteria: minimal improvement in the best fitness, reduction of step size below a threshold, reaching a maximum number of generations G_{max} , or degeneration of the covariance matrix (loss of diversity). The final optimized hyperparameters are obtained from the mean of the last distribution or from the historically best-performing candidate $\mathbf{z}_{\text{best}}$.

By maintaining a diverse population of solutions, CMA-ES balances exploration of the hyperparameter space with exploitation of promising regions, making it highly robust for optimizing deep learning models where the evaluation surface is noisy, non-convex, and high-dimensional.

Algorithm 4 summarizes the practical implementation of the CMA-ES procedure for hyperparameter optimization. The algorithm begins by initializing the generation counter g , the best fitness value $\mathcal{F}_{\text{best}}$, and the optimal hyperparameter configuration θ^* . The initial mean vector $\mathbf{m}^{(0)}$, covariance matrix $\mathbf{C}^{(0)}$, and step size $\sigma_{\text{step}}^{(0)}$ define the starting search distribution over the hyperparameter space.

At each generation g , a population of λ_{pop} candidate hyperparameter vectors $\mathbf{z}_i^{(g)}$ is sampled from the multivariate normal distribution defined by the current mean, covariance matrix, and step size. Each sampled candidate corresponds to a complete hyperparameter configuration and is evaluated by training the model and computing the validation loss $\mathcal{L}(\mathbf{z}_i^{(g)})$. These fitness values are then used to rank all candidates.

The top μ_{parent} candidates with the best fitness values are selected and used to update the mean vector $\mathbf{m}^{(g)}$ through weighted recombination, steering the search toward regions of improved performance. Subsequently, the covariance matrix $\mathbf{C}^{(g)}$ is updated using both rank-one and rank- μ_{parent} adaptations, enabling the algorithm to learn correlations between

hyperparameters and to elongate the search distribution along successful directions.

In parallel, the global step size σ_{step} is adapted using the step-size evolution path $\mathbf{p}_\sigma$, allowing CMA-ES to automatically control the scale of exploration. Larger step sizes encourage global exploration in early stages, while smaller step sizes promote fine-grained exploitation as the algorithm approaches convergence.

After each generation, the algorithm compares the best candidate of the current population with the best solution found so far. If an improvement is observed, the best fitness value $\mathcal{F}_{\text{best}}$ and the corresponding hyperparameter configuration θ^* are updated. The generation counter is then incremented, and the optimization process continues until a stopping criterion is met, such as reaching the maximum number of generations G_{max} , insufficient improvement, or a significant reduction in step size.

Finally, the algorithm returns the optimal hyperparameter configuration θ^* , corresponding either to the best individual encountered during the search or to the mean of the final sampling distribution.

Algorithm 4: Covariance Matrix Adaptation Evolution Strategy (CMA-ES)

Input: Hyperparameter space $\mathcal{H}$, population size λ_{pop} , number of parent solutions

 μ_{parent} , initial mean vector $\mathbf{m}^{(0)}$, initial covariance matrix $\mathbf{C}^{(0)}$, initial step size

 $\sigma_{\text{step}}^{(0)}$, maximum generations G_{max} , training dataset $\mathcal{D}_{\text{train}}$, validation dataset

 $\mathcal{D}_{\text{val}}$, evaluation function $\mathcal{L}$
Output: Optimal hyperparameter configuration θ^*

```

1 Initialize generation counter  $g \leftarrow 0$ ;
2 Initialize best fitness  $\mathcal{F}_{\text{best}} \leftarrow \infty$ ;
3 Initialize optimal configuration  $\theta^* \leftarrow \emptyset$ ;
4 Initialize evolution paths  $\mathbf{p}_{\sigma}^{(0)} \leftarrow \mathbf{0}$ ,  $\mathbf{p}_c^{(0)} \leftarrow \mathbf{0}$ ;
5 while  $g < G_{\text{max}}$  and not converged do
    // Sample candidate solutions
6   for  $i = 1$  to  $\lambda_{\text{pop}}$  do
7     Sample  $\mathbf{z}_i^{(g)} \sim \mathcal{N}(\mathbf{m}^{(g)}, (\sigma_{\text{step}}^{(g)})^2 \mathbf{C}^{(g)})$ ;
      // Each  $\mathbf{z}_i^{(g)}$  encodes a hyperparameter configuration
8     Train model  $f_{\mathbf{z}_i^{(g)}}$  on  $\mathcal{D}_{\text{train}}$ ;
9     Compute fitness  $\mathcal{F}_i = \mathcal{L}(\mathbf{z}_i^{(g)}) = \mathcal{M}(f_{\mathbf{z}_i^{(g)}}(\mathcal{D}_{\text{train}}), \mathcal{D}_{\text{val}})$ ;

    // Sort candidates by fitness (ascending for minimization)
10   Sort  $\{\mathbf{z}_i^{(g)}\}$  such that  $\mathcal{F}_{1:\lambda_{\text{pop}}} \leq \mathcal{F}_{2:\lambda_{\text{pop}}} \leq \dots \leq \mathcal{F}_{\lambda_{\text{pop}}:\lambda_{\text{pop}}}$ ;
    // Update mean vector
11    $\mathbf{m}^{(g+1)} = \sum_{i=1}^{\mu_{\text{parent}}} w_i \mathbf{z}_{i:\lambda_{\text{pop}}}^{(g)}$ ;
    // Update evolution paths and covariance matrix
12    $\mathbf{p}_{\sigma}^{(g+1)} = (1 - c_{\sigma})\mathbf{p}_{\sigma}^{(g)} + \sqrt{c_{\sigma}(2 - c_{\sigma})\mu_{\text{eff}}} \mathbf{C}^{(g)^{-\frac{1}{2}}} \frac{\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma_{\text{step}}^{(g)}}$ ;
13    $\mathbf{p}_c^{(g+1)} = (1 - c_c)\mathbf{p}_c^{(g)} + \sqrt{c_c(2 - c_c)\mu_{\text{eff}}} \frac{\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma_{\text{step}}^{(g)}}$ ;
14    $\mathbf{C}^{(g+1)} = (1 - c_1 - c_{\mu})\mathbf{C}^{(g)} + c_1 \mathbf{p}_c^{(g+1)}(\mathbf{p}_c^{(g+1)})^{\top} + c_{\mu} \sum_{i=1}^{\mu_{\text{parent}}} w_i \mathbf{y}_{i:\lambda_{\text{pop}}}^{(g)} (\mathbf{y}_{i:\lambda_{\text{pop}}}^{(g)})^{\top}$ ;
      // where  $\mathbf{y}_{i:\lambda_{\text{pop}}}^{(g)} = (\mathbf{z}_{i:\lambda_{\text{pop}}}^{(g)} - \mathbf{m}^{(g)})/\sigma_{\text{step}}^{(g)}$ 
    // Update step size
15    $\sigma_{\text{step}}^{(g+1)} = \sigma_{\text{step}}^{(g)} \exp\left(\frac{c_{\sigma}}{d_{\sigma}} \left(\frac{\|\mathbf{p}_{\sigma}^{(g+1)}\|}{\mathbb{E}\|\mathcal{N}(\mathbf{0}, \mathbf{I})\|} - 1\right)\right)$ ;
    // Update best solution
16   if  $\mathcal{F}_{1:\lambda_{\text{pop}}} < \mathcal{F}_{\text{best}}$  then
17      $\mathcal{F}_{\text{best}} \leftarrow \mathcal{F}_{1:\lambda_{\text{pop}}}$ ;
18      $\theta^* \leftarrow \mathbf{z}_{1:\lambda_{\text{pop}}}^{(g)}$ ;
19    $g \leftarrow g + 1$ ;
20 return  $\theta^*$ ;

```

4.8 Results and Discussion

Prior to training, each dataset was randomly partitioned into training and testing subsets using an 80%–20% split. Specifically, 80% of the images were used for model training, while the remaining 20% were reserved exclusively for performance evaluation. This separation ensures that the reported results reflect the model’s ability to generalize to previously unseen retinal images.

During model compilation, the Adam optimization algorithm was employed in conjunction with the categorical cross-entropy loss function $\mathcal{L}_{\text{CE}}$, while classification accuracy $\mathcal{A}$ served as the primary metric to monitor performance throughout both training and validation. To improve convergence and mitigate overfitting, two essential callbacks were incorporated. The EarlyStopping callback tracked validation loss $\mathcal{L}_{\text{val}}$ and terminated training when no substantial improvement was observed over a predefined number of epochs, as specified by the patience parameter p_{early} . This approach prevents unnecessary training, conserves computational resources, and restores model weights corresponding to the best validation performance.

In addition, the ReduceLROnPlateau callback was utilized to adaptively adjust the learning rate η . When the validation loss plateaued, signaling a stagnation in the optimization process, the learning rate was decreased by a factor γ_{lr} to facilitate more precise convergence toward an optimal solution. Training was conducted for a maximum of $E_{\text{max}} = 100$ epochs with a batch size of $B = 32$, balancing computational efficiency and effective learning.

All experiments were executed on a high-performance workstation configured with an Intel Core i9-13900K CPU, an NVIDIA RTX 4090 GPU (24 GB VRAM), and 128 GB of RAM. To further enhance model performance, hyperparameter tuning was carried out for key parameters, including the learning rate η , dropout rate p_{dropout} , and the number of neurons in the Dense layer n_{neurons} . Four distinct optimization strategies were explored: Grid Search, Random Search, Bayesian Optimization, and CMA-ES.

4.8.1 Grid Search

To establish baseline performance, a comprehensive Grid Search strategy was employed for each retinal dataset. The search systematically explored a predefined Cartesian product of three key hyperparameter dimensions:

- **Network capacity:** Dense layer units $n_{\text{neurons}} \in \{32, 64, 128, 256, 512\}$
- **Regularization intensity:** Dropout rates $p_{\text{dropout}} \in \{0.1, 0.3, 0.5, 0.7\}$
- **Optimization dynamics:** Learning rates $\eta \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\}$

This exhaustive approach evaluated $|\mathcal{H}| = 100$ distinct configurations per dataset, with each model trained for $E_{\max} = 100$ epochs. The selection criterion prioritized configurations minimizing validation loss $\mathcal{L}_{\text{val}}$, revealing how different retinal datasets influence architectural preferences. For instance, certain datasets favored reduced regularization, while others benefited from increased model complexity to accommodate intricate feature patterns.

The results in Table 4.1 demonstrate that the Grid Search strategy consistently yields high classification accuracy $\mathcal{A}$ across all datasets. RIDB, ARIA, DRIVE, and VARIA achieve perfect accuracy ($\mathcal{A} = 100\%$), while STARE and Local Data attain slightly lower values of 99% and 99.8%, respectively. These results indicate strong discriminative performance of the selected models, although minor variations suggest dataset-dependent learning challenges.

In terms of validation loss $\mathcal{L}_{\text{val}}$, RIDB and DRIVE exhibit the lowest values ($\mathcal{L}_{\text{val}} = 0.0005$), reflecting effective generalization and minimal overfitting. Conversely, VARIA and Local Data present higher validation losses ($\mathcal{L}_{\text{val}} = 0.0041$ and 0.0072 , respectively), suggesting increased variability in retinal patterns or image quality, which may complicate the learning process. The ARIA and STARE datasets demonstrate intermediate validation losses, indicating stable but slightly less uniform generalization.

Analysis of the optimal architectures reveals a high degree of consistency across datasets. Most configurations converge to $n_{\text{neurons}} = 256$, a dropout rate p_{dropout} between 0.1 and 0.3, and a learning rate $\eta = 0.0001$, indicating that moderate model complexity combined with controlled regularization is sufficient for effective retinal feature extraction. The VARIA dataset constitutes an exception, requiring a larger network with $n_{\text{neurons}} = 512$, highlighting the need for increased representational capacity to model its structural complexity.

From a computational perspective, the Grid Search approach incurs substantial training costs. RIDB requires the shortest optimization time $t_{\text{opt}} = 8.68$ hours, whereas VARIA demands significantly more resources ($t_{\text{opt}} = 58.24$ hours), reflecting the impact of increased architectural complexity and dataset heterogeneity. The remaining datasets exhibit intermediate computational costs, ranging from approximately 14 to 19 hours.

Overall, Table 4.1 illustrates that while Grid Search is capable of identifying highly accurate configurations, it entails considerable computational expense and exhibits sensitivity to dataset characteristics. These findings underscore the importance of balancing model performance with optimization efficiency, particularly in large-scale medical image analysis applications.

Table 4.1. Grid Search Optimization Outcomes Across Retinal Datasets

Metric / Dataset	RIDB	ARIA	STARE	DRIVE	VARIA	Local Data
Performance Metrics						
Validation Loss $\mathcal{L}_{\text{val}}$	5.00×10^{-4}	2.90×10^{-3}	9.00×10^{-4}	5.00×10^{-4}	4.10×10^{-3}	7.20×10^{-3}
Classification Accuracy $\mathcal{A}$ (%)	100	100	99	100	100	99.8
Architectural Complexity						
Neurons n_{neurons}	256	256	256	256	512	256
Dropout Rate p_{dropout}	0.3	0.3	0.1	0.1	0.3	0.1
Optimization Parameters						
Learning Rate η	1.00×10^{-4}	1.00×10^{-4}	1.00×10^{-4}	1.00×10^{-4}	1.00×10^{-4}	1.00×10^{-4}
Optimization Time t_{opt} (hours)	8.68	16.23	19.28	18.54	58.24	13.97

4.8.2 Random Search

To systematically identify effective model configurations, a Random Search approach was implemented for hyperparameter tuning. The search space $\mathcal{H}$ encompassed three critical architectural and training parameters:

- **Dense layer dimensionality:** $n_{\text{neurons}} \in \{32, 64, 128, 256, 512\}$
- **Regularization using dropout:** $p_{\text{dropout}} \in \{0.1, 0.3, 0.5, 0.7\}$
- **Optimization learning rate:** $\eta \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\}$

The experimental design involved generating $N = 20$ distinct hyperparameter combinations for each retinal image dataset. This quantity represented a deliberate compromise between computational resource constraints and exploratory breadth within the parameter landscape.

The Random Search methodology identified high-performing model configurations across all evaluated datasets. The results demonstrated exceptional classification efficacy, with four datasets—RIDB, ARIA, DRIVE, and VARIA—achieving perfect accuracy ($\mathcal{A} = 100\%$) on their validation partitions. The LOCAL DATA and STARE datasets also performed strongly, attaining 99.9% and 99% accuracy, respectively. The minor accuracy variance for the latter two suggests that while the random sampling strategy discovered generally robust hyperparameter sets, additional model refinement could enhance performance uniformity. Analysis of the optimal configurations θ^* revealed distinct patterns across the datasets, as detailed in Table 4.2.

A clear convergence was observed for three datasets (RIDB, DRIVE, VARIA), which all optimized to a nearly identical configuration ($n_{\text{neurons}} \approx 188$, $p_{\text{dropout}} \approx 0.217$, $\eta \approx 4.00 \times 10^{-4}$). In contrast, the ARIA dataset required a more finely-tuned learning rate ($\eta = 7.54 \times 10^{-5}$) for peak performance.

Two datasets exhibited unique architectural needs. The LOCAL DATA corpus necessitated a substantially larger network ($n_{\text{neurons}} = 498$) coupled with stronger regularization ($p_{\text{dropout}} = 0.524$), indicating greater intrinsic complexity that demanded increased model

Table 4.2. Performance Metrics and Optimal Hyperparameters Identified through Random Search Across Six Retinal Image Datasets

Metric / Dataset	RIDB	ARIA	STARE	DRIVE	VARIA	Local Data
Performance Metrics						
Validation Loss $\mathcal{L}_{\text{val}}$	-5.30×10^{-3}	-5.50×10^{-3}	-9.00×10^{-2}	-1.00×10^{-3}	-2.08×10^{-2}	-2.43×10^{-2}
Classification Accuracy $\mathcal{A}$ (%)	100	100	99	100	100	99.9
Optimal Hyperparameters θ^*						
Neurons n_{neurons}	188	189	256	189	188	498
Dropout Rate p_{dropout}	0.217	0.227	0.100	0.217	0.217	0.524
Learning Rate η	4.00×10^{-4}	7.54×10^{-5}	1.00×10^{-4}	4.00×10^{-4}	4.00×10^{-4}	2.00×10^{-4}
Computational Cost						
Optimization Time t_{opt} (hours)	48.4	9.9	19.3	47.4	58.2	14.0

capacity alongside overfitting prevention. Conversely, STARE performed optimally with minimal dropout regularization ($p_{\text{dropout}} = 0.1$), suggesting a lower inherent risk of overfitting.

The computational investment for the search process varied considerably. VARIA demanded the most extensive exploration ($t_{\text{opt}} = 58.2$ hours), whereas ARIA converged most rapidly ($t_{\text{opt}} = 9.9$ hours). This disparity highlights how dataset-specific characteristics directly influence the hyperparameter optimization landscape, even under a consistent search protocol.

Overall, Random Search proved to be an effective and practical approach for configuring deep learning models in this medical imaging context. It successfully identified high-accuracy models while providing valuable insights into parameter sensitivities and dataset-specific requirements. These results establish it as a viable baseline method, though they also point to potential benefits from more advanced optimization techniques when pursuing maximal consistency across diverse data sources.

4.8.3 Bayesian Optimization

A Bayesian Optimization (BO) framework was implemented to efficiently navigate the hyperparameter landscape for each retinal dataset. The procedure explored continuous ranges for three critical parameters:

- **Network capacity:** $n_{\text{neurons}} \in [32, 512]$
- **Regularization intensity:** $p_{\text{dropout}} \in [0.1, 0.7]$
- **Optimization step size:** $\eta \in [10^{-5}, 10^{-2}]$

The procedure commenced with an initial exploratory stage involving $n_{\text{init}} = 20$ randomly selected hyperparameter configurations, followed by $N = 50$ optimization iterations directed by a Gaussian process-based surrogate model. An acquisition function $\alpha(\theta)$ was employed to minimize the validation loss $\mathcal{L}_{\text{val}}$, thereby explicitly favoring configurations that enhance model generalization. The resulting optimal hyperparameter settings θ^* and

associated performance metrics are summarized in Table 4.3 and Figure 4.4.

Table 4.3. Bayesian Optimization Results Across Retinal Image Datasets

Metric / Dataset	RIDB	ARIA	STARE	DRIVE	VARIA	Local Data
Performance Metrics						
Validation Loss $\mathcal{L}_{\text{val}}$	-5.30×10^{-3}	-5.50×10^{-3}	-1.683×10^{-1}	-1.00×10^{-3}	4.10×10^{-3}	8.64×10^{-2}
Classification Accuracy $\mathcal{A}$ (%)	100	100	99	100	100	100
Optimal Hyperparameters θ^*						
Neurons n_{neurons}	188	187	454	188	288	32
Dropout Rate p_{dropout}	0.217	0.227	0.302	0.217	0.218	0.1
Learning Rate η	4.60×10^{-4}	7.54×10^{-5}	5.90×10^{-4}	4.60×10^{-4}	4.61×10^{-4}	1.00×10^{-4}
Computational Cost						
Optimization Time t_{opt} (hours)	48.36	9.91	4.76	47.41	55.46	40.34

A closer examination of Table 4.3 and Figure 4.4 provides further insight into the behavior of Bayesian Optimization across the different retinal datasets. While high classification accuracy $\mathcal{A}$ is consistently achieved, the combined analysis of validation loss $\mathcal{L}_{\text{val}}$, architectural choices, and search dynamics reveals important differences in dataset complexity and optimization sensitivity.

The validation loss distribution highlights that achieving perfect accuracy does not necessarily imply optimal generalization. For instance, RIDB, ARIA, and DRIVE attain near-zero or slightly negative validation losses while maintaining 100% accuracy, indicating stable convergence and balanced model capacity. In Figure 4.4, these datasets exhibit well-defined low-loss regions concentrated around moderate dropout values and learning rates, suggesting smooth optimization landscapes where Bayesian Optimization can efficiently locate optimal configurations. The strong overlap between the RIDB and DRIVE plots further supports the observation that these datasets share similar statistical properties and feature distributions, allowing them to benefit from nearly identical hyperparameter settings.

The STARE dataset presents a markedly different behavior. Despite achieving a slightly lower accuracy of 99%, it exhibits the lowest validation loss ($\mathcal{L}_{\text{val}} = -0.1683$), reflecting superior generalization. Figure 4.4 shows that optimal configurations for STARE are located in regions corresponding to higher network capacity ($n_{\text{neurons}} = 454$) and increased regularization ($p_{\text{dropout}} = 0.302$). This indicates that STARE contains richer or more diverse retinal patterns that require a larger representational space to be effectively modeled, while simultaneously demanding stronger regularization to prevent overfitting. The compact and well-localized optimum observed in the STARE plot suggests that, although the dataset is complex, its optimization landscape is relatively well-structured, allowing Bayesian Optimization to converge rapidly, as reflected by the reduced computational cost $t_{\text{opt}} = 4.76$ hours reported in Table 4.3.

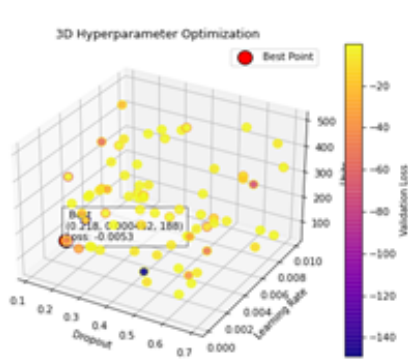
In contrast, VARIA and Local Data demonstrate less favorable optimization characteristics. Both datasets achieve perfect accuracy but are associated with higher validation losses, particularly in the case of Local Data ($\mathcal{L}_{\text{val}} = 0.0864$). This discrepancy suggests that the

models may be fitting dataset-specific patterns or noise rather than learning features that generalize well. In Figure 4.4, this behavior is reflected by a more dispersed distribution of loss values across the hyperparameter space, with no sharply defined low-loss basin. For VARIA, the optimal configuration requires a larger network ($n_{\text{neurons}} = 288$) combined with moderate dropout ($p_{\text{dropout}} = 0.218$), indicating increased data variability and a higher risk of overfitting. The extended optimization time $t_{\text{opt}} = 55.46$ hours reported in Table 4.3 further confirms the difficulty of navigating this search space.

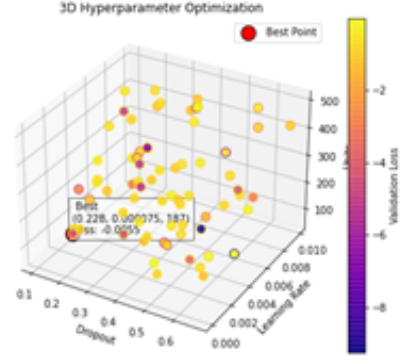
Local Data exhibits the most challenging generalization behavior. The optimal configuration is characterized by a very small network ($n_{\text{neurons}} = 32$) and minimal regularization ($p_{\text{dropout}} = 0.1$), yet the validation loss remains comparatively high ($\mathcal{L}_{\text{val}} = 0.0864$). This suggests that the dataset may contain noise, limited sample diversity, or acquisition inconsistencies that restrict the model’s ability to generalize, regardless of architectural complexity. The broad and flat loss distribution observed reinforces this interpretation, as Bayesian Optimization fails to identify a configuration that significantly improves validation performance despite extensive exploration.

The learning rate trends observed in Table 4.3 further emphasize the dataset-dependent nature of optimization. Datasets with smoother optimization landscapes, such as RIDB and DRIVE, converge toward moderate learning rates ($\eta \approx 4.60 \times 10^{-4}$) that balance convergence speed and stability. Conversely, ARIA and Local Data favor smaller learning rates ($\eta \approx 7.54 \times 10^{-5}$ and 1.00×10^{-4} , respectively), indicating a need for more conservative weight updates to maintain stable training. These patterns are consistent with the relative spread and shape of the low-loss regions observed.

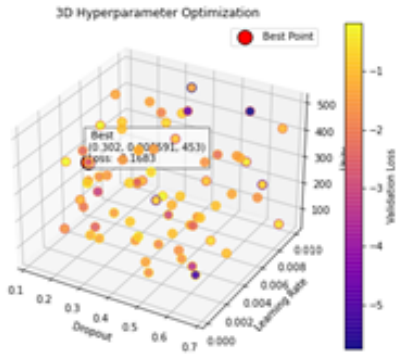
Overall, the combined interpretation of Table 4.3 demonstrates that Bayesian Optimization not only identifies optimal hyperparameters but also acts as an implicit diagnostic tool for understanding dataset complexity and generalization behavior. Well-structured datasets yield compact, well-defined optima and faster convergence, whereas datasets with higher variability or noise result in broader loss landscapes and increased computational cost. These findings reinforce the importance of dataset-specific hyperparameter optimization and validate the use of Bayesian Optimization as a principled and informative approach for retinal biometric modeling.



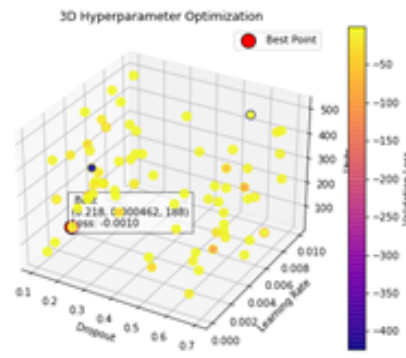
(a) RIDB dataset



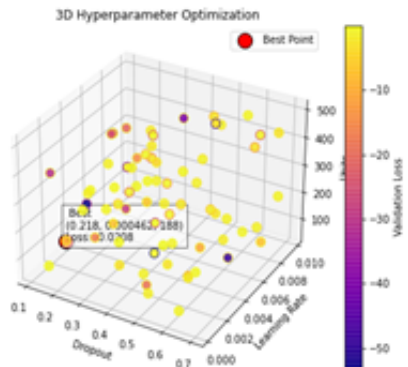
(b) ARIA dataset



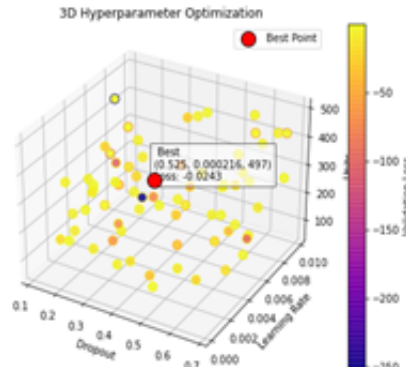
(c) STARE dataset



(d) DRIVE dataset



(e) VARIA dataset



(f) Local dataset

Figure 4.4. Effect of Dropout Rate p_{dropout} and Learning Rate η on Validation Loss $\mathcal{L}_{\text{val}}$: A Three-Dimensional Visualization across Multiple Datasets.

4.8.4 CMA-ES Method

The CMA-ES optimization targeted three critical hyperparameters that collectively govern model architecture and training dynamics:

- **Network Capacity:** Number of neurons in the final dense layer, with search bounds $n_{\text{neurons}} \in [32, 512]$. This parameter directly influences the model's representational power and computational requirements.
- **Regularization Intensity:** Dropout probability, bounded $p_{\text{dropout}} \in [0.1, 0.7]$. Dropout serves as a regularization mechanism to prevent overfitting by randomly omitting units during training.
- **Learning Rate:** Initial learning rate for the Adam optimizer, with logarithmic bounds $\eta \in [10^{-5}, 10^{-2}]$. This parameter controls the step size during gradient-based optimization.

The three-dimensional search space was deliberately constrained to focus on parameters with established significance for model performance while maintaining computational tractability. The bounded ranges were selected based on empirical evidence from preliminary experiments and established practices in deep learning for medical image analysis.

The CMA-ES implementation was configured with the following algorithmic parameters:

- **Population Size:** $\lambda_{\text{pop}} = 20$ candidate solutions per generation
- **Initial Mean Vector:** $\mathbf{m}^{(0)} = [128, 0.5, 10^{-3}]$ corresponding to initial guesses for neurons n_{neurons} , dropout p_{dropout} , and learning rate η
- **Initial Step Size:** $\sigma_{\text{step}}^{(0)} = 0.5$, representing moderate initial exploration
- **Covariance Matrix Initialization:** Identity matrix $\mathbf{C}^{(0)} = \mathbf{I}$, indicating initially isotropic search
- **Termination Criteria:**
 - Convergence threshold: $\|\mathbf{m}^{(g)} - \mathbf{m}^{(g-1)}\| < 10^{-6}$
 - Maximum generations: $G_{\text{max}} = 50$
 - Maximum function evaluations: $N_{\text{max}} = 1000$

The selection of population size followed the heuristic:

$$\lambda_{\text{pop}} = 4 + \lfloor 3 \ln d \rfloor \quad (4.8.1)$$

where $d = 3$ represents the dimensionality of the hyperparameter search space. The resulting value was rounded to a practical size to balance computational efficiency and search diversity. The initial mean vector was selected to provide a balanced starting point within the defined hyperparameter ranges, while the initial step size was set to enable substantial exploration of the search space in the early stages of optimization.

The objective function was formulated as the negative validation loss:

$$\mathcal{F}_{\text{CMA-ES}}(\theta) = -\mathcal{L}_{\text{val}}(\theta) \quad (4.8.2)$$

where $\theta = (n_{\text{neurons}}, p_{\text{dropout}}, \eta)$ denotes a candidate hyperparameter configuration, and $\mathcal{L}_{\text{val}}$ is the corresponding validation loss obtained after model training. By defining the objective in this manner, the original minimization problem (reducing validation loss) is converted into a maximization problem, which is consistent with the default optimization framework of CMA-ES.

To ensure reproducibility, a fixed random seed was assigned to each dataset, controlling all sources of stochasticity including weight initialization, dropout masks, and data shuffling.

Beyond the primary objective function, several internal CMA-ES metrics were monitored to analyze algorithm behavior:

- **Step Size** $\sigma_{\text{step}}^{(g)}$: Controls the overall scale of the search distribution
- **Axis Ratio**: Ratio of largest to smallest eigenvalue of the covariance matrix $\mathbf{C}^{(g)}$, quantifying search anisotropy
- **Principal Standard Deviations**: Square roots of eigenvalues, representing search spread along principal axes
- **Condition Number**: Ratio of largest to smallest eigenvalue, indicating numerical stability of the covariance matrix
- **Evolution Paths** $\mathbf{p}_{\sigma}^{(g)}$ and $\mathbf{p}_{\epsilon}^{(g)}$: Cumulative information about successful search directions

These metrics collectively provide insights into how CMA-ES adapts its search strategy in response to landscape characteristics, offering explanatory power beyond mere convergence rates.

The CMA-ES optimization produced distinct hyperparameter configurations for each retinal dataset, with varying convergence characteristics and final performances. Table 4.4 presents a consolidated view of all optimization outcomes, facilitating comparative analysis across datasets.

Table 4.4 summarizes the outcomes of CMA-ES hyperparameter optimization across the six retinal imaging datasets. The results reveal clear differences in convergence dynamics, computational cost, and optimal hyperparameter configurations, highlighting the strong influence of dataset characteristics on the optimization process. ARIA and STARE reached their best-performing configurations within only two generations ($g^* = 2$), indicating rapid convergence and suggesting that suitable hyperparameter regions could be identified early

Table 4.4. Consolidated CMA-ES Optimization Results Across All Datasets

Metric / Dataset	ARIA	RIDB	STARE	DRIVE	VARIA	LOCAL
Convergence						
Optimal Generation g^*	2	4	2	2	5	5
Best Fitness $\mathcal{F}_{\text{best}}$	-5.20×10^{-3}	-4.80×10^{-3}	-4.77×10^{-3}	-1.30×10^{-3}	3.80×10^{-3}	8.57×10^{-2}
Optimization Time t_{opt} (h)	10.59	45.16	19.28	44.15	54.35	39.72
Optimal Hyperparameters θ^*						
Neurons n_{neurons}	178	192	256	190	276	30
Dropout p_{dropout}	0.284	0.265	0.100	0.251	0.278	0.105
Learning Rate η	6.93×10^{-5}	5.20×10^{-4}	1.00×10^{-4}	4.70×10^{-4}	4.59×10^{-4}	1.00×10^{-4}
Algorithm State at Convergence						
Final Axis Ratio γ_{axis}	1.2	1.7	1.5	1.2	3.8	2.8
Final Step Size $\sigma_{\text{step}}^{(g)}$	0.494	0.413	0.151	0.580	0.283	0.450
Min Std. Dev. σ_{min}	0.4	0.3	0.7	0.5	0.5	0.1
Max Std. Dev. σ_{max}	0.6	0.4	0.1	0.7	0.4	0.4
Performance						
Test Accuracy $\mathcal{A}_{\text{test}}$ (%)	100.0	100.0	99.9	100.0	100.0	100.0
Convergence Class	Fast	Moderate	Fast	Moderate	Slow	Exploratory

in the search. RIDB and DRIVE exhibited more moderate convergence behavior; although DRIVE achieved its optimum at a low generation count ($g^* = 2$), the overall optimization time remained high ($t_{\text{opt}} = 44.15$ hours), reflecting the substantial computational expense of deep learning evaluations rather than difficulties in the search itself. In contrast, VARIA and LOCAL required five generations ($g^* = 5$) and the longest runtimes ($t_{\text{opt}} = 54.35$ and 39.72 hours, respectively), indicating that these datasets demanded more extensive exploration before stable optimal solutions were found.

The optimal hyperparameter values demonstrate significant variation across datasets, confirming the necessity of dataset-specific tuning. The number of neurons ranges from $n_{\text{neurons}} = 30$ in LOCAL to $n_{\text{neurons}} = 276$ in VARIA, reflecting different representational capacity requirements. Larger architectures, as observed in VARIA and STARE ($n_{\text{neurons}} = 256$), suggest that these datasets benefit from higher model complexity, potentially due to greater variability in retinal structures or lesion patterns. Conversely, the small optimal network size for LOCAL ($n_{\text{neurons}} = 30$) indicates that additional capacity did not improve performance, emphasizing the importance of balancing complexity against the risk of unnecessary over-parameterization.

Dropout values further illustrate the role of regularization in controlling generalization. Most datasets converged toward moderate dropout rates between $p_{\text{dropout}} = 0.25$ and 0.28, indicating that regularization was essential to mitigate overfitting during training. However, STARE and LOCAL required substantially lower dropout rates near $p_{\text{dropout}} = 0.10$, suggesting that stronger regularization was less critical in these cases. This variability confirms that dropout must be adapted to the dataset’s intrinsic complexity and training stability rather than being fixed universally.

Learning rate selection also differed notably across datasets, spanning approximately one

order of magnitude from $\eta = 6.93 \times 10^{-5}$ to $\eta = 5.20 \times 10^{-4}$. Smaller learning rates, such as those obtained for ARIA ($\eta = 6.93 \times 10^{-5}$), indicate that fine-grained updates were beneficial for stable convergence, whereas larger learning rates in RIDB and DRIVE ($\eta \approx 4.70 \times 10^{-4}$) may have facilitated broader exploration of the loss surface. This reinforces the learning rate as one of the most influential hyperparameters governing both convergence behavior and final model performance.

Beyond the hyperparameter values, the internal CMA-ES adaptation metrics provide insight into the structure of the optimization landscapes. Low final axis ratios close to $\gamma_{\text{axis}} = 1.2$, as observed in ARIA and DRIVE, indicate near-isotropic search distributions, suggesting that hyperparameter directions contributed relatively equally to performance improvements. In contrast, VARIA and LOCAL exhibited much higher axis ratios ($\gamma_{\text{axis}} = 3.8$ and 2.8 , respectively), revealing strong anisotropic adaptation. This indicates that CMA-ES identified dominant search directions, demonstrating its advantage in navigating complex and elongated optimization landscapes through covariance shaping.

The final step size values further reflect CMA-ES’s adaptive exploration–exploitation balance. Larger $\sigma_{\text{step}}^{(g)}$ values, such as in DRIVE ($\sigma_{\text{step}}^{(g)} = 0.580$), suggest that the algorithm maintained broader exploration even at convergence, whereas smaller $\sigma_{\text{step}}^{(g)}$ values, as in STARE ($\sigma_{\text{step}}^{(g)} = 0.151$), indicate stronger exploitation and refinement near the optimum. These patterns highlight CMA-ES’s ability to automatically adjust its search radius depending on the difficulty of the dataset-specific landscape.

Despite the differences in convergence time, hyperparameter configurations, and adaptation behavior, all datasets achieved extremely high test accuracy, reaching approximately 99.9–100%. This confirms the effectiveness of CMA-ES in identifying robust hyperparameter settings across diverse retinal imaging datasets. At the same time, the large variation in optimization time, ranging from approximately 10.6 hours to over 54 hours, underlines the computational cost of evolutionary hyperparameter optimization and the importance of careful resource planning in biomedical deep learning applications. Overall, the results demonstrate that CMA-ES provides a powerful automated framework for hyperparameter tuning, with strong adaptability to dataset-dependent complexity and optimization landscape characteristics.

4.8.5 Comparative Analysis of Optimization Methods

A comparative summary of the best validation loss $\mathcal{L}_{\text{best}}$ achieved by each optimization method across all datasets is presented in Table 4.5. This comparison highlights the relative effectiveness of the different search strategies.

The table compares the best validation loss $\mathcal{L}_{\text{best}}$ achieved by four hyperparameter optimization strategies across the six datasets. The results clearly demonstrate that the

Table 4.5. Best Validation Loss $\mathcal{L}_{\text{best}}$ Achieved by Each Optimization Strategy Across Datasets

Optimization Strategy	RIDB	ARIA	STARE	DRIVE	VARIA	Local Data
Grid Search	5.00×10^{-4}	2.90×10^{-3}	9.00×10^{-4}	5.00×10^{-4}	4.10×10^{-3}	7.10×10^{-3}
Random Search	-5.30×10^{-3}	-5.50×10^{-3}	-9.00×10^{-2}	-1.00×10^{-3}	-2.08×10^{-2}	-2.43×10^{-2}
Bayesian Optimization	-5.30×10^{-3}	-2.20×10^{-3}	-1.683×10^{-1}	-1.00×10^{-3}	4.10×10^{-3}	8.64×10^{-2}
CMA-ES	-4.80×10^{-3}	-5.20×10^{-3}	-4.77×10^{-3}	-1.30×10^{-3}	3.80×10^{-3}	8.57×10^{-2}

effectiveness of an optimization method is highly dependent on the dataset and the structure of the corresponding search space $\mathcal{H}$. Overall, stochastic and adaptive strategies tend to outperform exhaustive deterministic search, confirming the limitations of Grid Search in complex deep learning optimization settings.

Grid Search consistently produced the weakest performance across all datasets, yielding only positive validation loss values. This outcome reflects the restricted nature of grid-based exploration, where only a small number of predefined hyperparameter combinations are evaluated. As a result, Grid Search is unable to efficiently locate optimal configurations in continuous, high-sensitivity hyperparameter spaces, particularly when the true optimum θ^* lies between discretized grid points.

In contrast, Random Search achieved substantially better validation losses, producing negative values across most datasets. This improvement highlights the advantage of sampling-based exploration, which allows the optimizer to cover a broader region of the search space without being constrained by rigid parameter intervals. The strongest Random Search results were observed for VARIA and LOCAL, where losses reached $\mathcal{L}_{\text{best}} = -2.08 \times 10^{-2}$ and -2.43×10^{-2} respectively, suggesting that these datasets require more diverse exploration to identify high-performing hyperparameter settings.

Bayesian Optimization demonstrated competitive performance on several datasets, particularly STARE, where it achieved the most favorable validation loss ($\mathcal{L}_{\text{best}} = -1.683 \times 10^{-1}$). This reflects the strength of surrogate-based approaches, which iteratively refine the search toward promising regions by balancing exploration and exploitation. However, Bayesian Optimization also exhibited instability across datasets, producing positive validation losses in VARIA and extremely large values in LOCAL. This suggests that surrogate modeling can become unreliable when the loss surface is highly irregular or when evaluations are noisy, limiting its robustness in certain biomedical datasets.

CMA-ES achieved consistently strong performance across multiple datasets, with negative validation losses in RIDB, ARIA, STARE, and DRIVE. Its results confirm the effectiveness of evolutionary adaptation mechanisms, particularly covariance shaping $\mathbf{C}^{(g)}$, which enables efficient navigation of complex, anisotropic landscapes. Although CMA-ES did not always outperform Bayesian Optimization in absolute terms, it provided more stable improvements across datasets, avoiding the extreme failures observed in Bayesian Optimization for VARIA and LOCAL. This stability highlights CMA-ES as a robust strategy when the

hyperparameter landscape is difficult to model or highly directional.

Taken together, Table 4.5 indicates that deterministic exhaustive strategies such as Grid Search are insufficient for deep learning hyperparameter tuning, while adaptive stochastic approaches provide significantly better validation performance. Random Search offers strong baseline performance through broad exploration, Bayesian Optimization excels when surrogate modeling remains reliable, and CMA-ES provides a particularly robust alternative due to its ability to adaptively reshape the search distribution. These findings support the selection of evolutionary and probabilistic optimization frameworks for retinal imaging applications, where dataset variability and complex loss landscapes make manual or grid-based tuning impractical.

Table 4.6 provides a comparative synthesis of the four hyperparameter optimization strategies evaluated in this study, highlighting their fundamental differences in search behavior, computational demands, and suitability for deep learning applications in retinal imaging. The comparison demonstrates that the choice of optimization method significantly influences both the efficiency of the tuning process and the quality of the resulting hyperparameter configurations.

Table 4.6. Comparative Analysis of Hyperparameter Optimization Methods

Metric	Grid Search	Random Search	Bayesian Optimization	Opti-	CMA-ES
Search Strategy	Exhaustive grid search	Random sampling	Surrogate guides search	model	Evolutionary strategy with covariance adaptation
Computational Efficiency	Very Low	Moderate	High		High (per iteration)
Exploration vs. Exploitation	Pure exploration	Mostly exploration	Balanced		Strong initial exploration → focused exploitation
Best Validation Loss $\mathcal{L}_{\text{best}}$	Often suboptimal	Variable	Low		Lowest / Most consistent
Convergence Speed	Slowest	Slow	Fast		Moderate–Fast
Computational Cost t_{opt}	Extremely high	High	Moderate–High		High
Adaptability	Poor	Poor	Good		Excellent
Scalability	Not scalable	Inefficient in high dimensions	Efficient for moderate dimensions		Computationally demanding but scalable

Grid Search represents the most straightforward approach, relying on exhaustive enumeration of predefined hyperparameter combinations. While conceptually simple, its computational efficiency is extremely low, particularly because the number of evaluations grows exponentially with the number of tuned parameters. As reflected in the table, Grid Search provides pure exploration without any adaptive refinement, which often leads to suboptimal validation performance. Its poor scalability makes it impractical for modern deep learning problems, where hyperparameter spaces are continuous and high-dimensional.

Random Search improves upon this limitation by replacing deterministic enumeration with stochastic sampling. This strategy achieves moderate efficiency, as it explores the

hyperparameter space more broadly and is not constrained to rigid grid discretizations. However, Random Search remains largely exploratory and lacks exploitation mechanisms that concentrate evaluations in promising regions. Consequently, its best validation loss $\mathcal{L}_{\text{best}}$ is highly variable across datasets, and convergence remains relatively slow, requiring many evaluations to approach optimal solutions.

Bayesian Optimization introduces a more sophisticated framework by employing a surrogate probabilistic model to approximate the objective function $f(\theta)$. This enables a balanced exploration–exploitation tradeoff, as the optimizer selectively evaluates hyperparameters that are predicted to yield improvements while still maintaining uncertainty-driven exploration. As summarized in Table 4.6, Bayesian Optimization typically achieves fast convergence and low validation losses with fewer evaluations compared to exhaustive or random approaches. Nevertheless, its scalability is most effective in moderate-dimensional spaces, and its performance may degrade when the loss landscape becomes highly irregular or noisy, as is often the case in biomedical deep learning.

CMA-ES, in contrast, provides an evolutionary optimization strategy that adapts its search distribution through covariance matrix adaptation $\mathbf{C}^{(g)}$. This mechanism allows CMA-ES to reshape the sampling process dynamically, aligning the search toward dominant directions in anisotropic hyperparameter landscapes. As shown in the table, CMA-ES demonstrates excellent adaptability and consistently strong validation performance, often achieving the lowest losses across datasets. Its optimization process typically begins with strong exploration and gradually shifts toward focused exploitation, reflecting an effective self-adaptive balance. However, this robustness comes at the cost of high computational expense, as evolutionary strategies often require many iterations and repeated network training evaluations.

Overall, Table 4.6 highlights that deterministic approaches such as Grid Search are unsuitable for complex deep learning hyperparameter tuning due to poor efficiency and lack of adaptability. Random Search provides a stronger baseline through broader exploration but remains limited by its non-adaptive nature. Bayesian Optimization offers fast convergence and efficient search when surrogate modeling is reliable, whereas CMA-ES provides the most robust and consistent performance in challenging optimization landscapes through its covariance adaptation mechanism. These findings support the use of adaptive evolutionary or probabilistic optimization frameworks for retinal imaging applications, where loss surfaces are complex and manual tuning is infeasible despite the associated computational cost.

4.8.6 Performance Evaluation Metrics

Table 4.7 summarizes the performance evaluation metrics obtained across the six datasets, providing a comprehensive assessment of both classification reliability and biometric identification efficiency. Overall, the results indicate extremely high predictive accuracy and strong generalization across all datasets, confirming the effectiveness of the proposed framework under diverse retinal imaging conditions.

Table 4.7. Performance Evaluation Metrics Across Datasets

Metric	DRIVE	RIDB	ARIA	STARE	VARIA	Local Data
F1-Score F_1 (%)	100	100	100	99	100	100
Precision P (%)	100	100	100	99	100	100
Specificity S_p (%)	100	100	100	99.99	100	100
Avg. Identification Time $\bar{t}_{id}$ (s)	0.001270	0.001267	0.004026	0.001220	0.001407	0.001324
Genuine Acceptance Rate GAR (%)	100	100	100	99.57	100	100
Mean ROC AUC $\overline{AUC}$	1.000	1.000	1.000	1.000	1.000	1.000
Log Loss $\mathcal{L}_{log}$	0.0044	0.0024	7.64×10^{-5}	0.1010	0.0000	0.0294
Cohen's Kappa κ	1.000	1.000	1.000	0.9873	1.000	1.000
Accuracy $\mathcal{A}$ (%)	100	100	100	99.9	100	100

The classification metrics, including F1-score F_1 , precision P , and specificity S_p , remain consistently near perfect across all datasets. DRIVE, RIDB, ARIA, VARIA, and the Local dataset achieved 100% in all three measures, demonstrating flawless discrimination between classes and the absence of false positive or false negative predictions. STARE presents a slightly lower performance, with F1-score and precision values of 99%, yet still reflects a highly accurate and clinically reliable classification outcome. The uniformly high specificity values further confirm that the system maintains excellent robustness against false alarms, which is particularly important in medical diagnostic and biometric security contexts.

In addition to classification accuracy, the biometric identification efficiency is reflected in the average identification time per sample $\bar{t}_{id}$. All datasets exhibit extremely low inference times, ranging from approximately 0.0012 to 0.0040 seconds per sample. These results demonstrate that the model is not only accurate but also computationally efficient, supporting its suitability for real-time biometric deployment. The slightly higher identification time observed in ARIA ($\bar{t}_{id} = 0.004026$ s) may be attributed to dataset-specific preprocessing or feature complexity, though the overall latency remains negligible.

The Genuine Acceptance Rate GAR, which measures the proportion of correctly accepted genuine identities, further confirms the robustness of the biometric recognition performance. All datasets achieved a GAR of 100%, except STARE, which attained 99.57%. This minor reduction suggests slightly increased intra-class variability or imaging heterogeneity within STARE, yet the acceptance performance remains exceptionally strong.

The ROC AUC values equal to 1.000 across all datasets indicate perfect separability

between classes, reinforcing the model’s discriminative power. Similarly, Cohen’s Kappa values $\kappa = 1.000$ for most datasets reflect complete agreement beyond chance, confirming the consistency and reliability of predictions. STARE again exhibits a marginally lower Kappa value ($\kappa = 0.9873$), consistent with its slightly reduced classification metrics, but still within an excellent agreement range.

Log loss values $\mathcal{L}_{\log}$ provide additional insight into probabilistic calibration. ARIA achieved an extremely small log loss ($\mathcal{L}_{\log} = 7.64 \times 10^{-5}$), indicating highly confident and correct probability estimates. DRIVE and RIDB also show very low log loss values, whereas STARE presents the highest log loss ($\mathcal{L}_{\log} = 0.1010$), suggesting that although predictions remain accurate, the confidence distribution may be less sharply optimized. VARIA reports a log loss of 0.0000, which may indicate perfect confidence alignment or potential rounding effects in reporting.

Finally, overall accuracy $\mathcal{A}$ remains near-perfect, reaching 100% across all datasets except STARE, which achieved 99.9%. This confirms that the proposed model maintains stable and generalized performance even across datasets with different acquisition conditions and structural variability.

4.8.7 Learning Curves Analysis

To ensure model generalization, learning curves (training vs. validation accuracy/loss) were analyzed for all datasets (Figures 4.5 to 4.10). Across all cases, validation metrics closely tracked training metrics without significant divergence. Both accuracy curves $\mathcal{A}_{\text{train}}$ and $\mathcal{A}_{\text{val}}$ converged smoothly to high values, while loss curves $\mathcal{L}_{\text{train}}$ and $\mathcal{L}_{\text{val}}$ decreased monotonically and stabilized at low values. This consistent alignment indicates effective learning without overfitting.

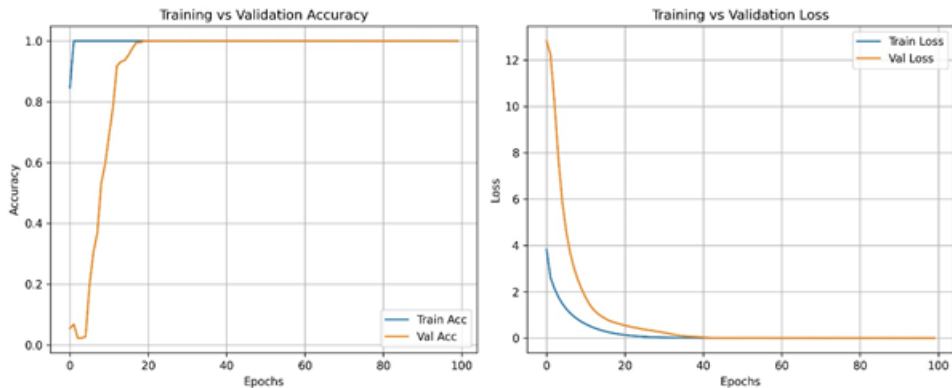


Figure 4.5. Training and Validation Learning Curves for the RIDB Dataset.

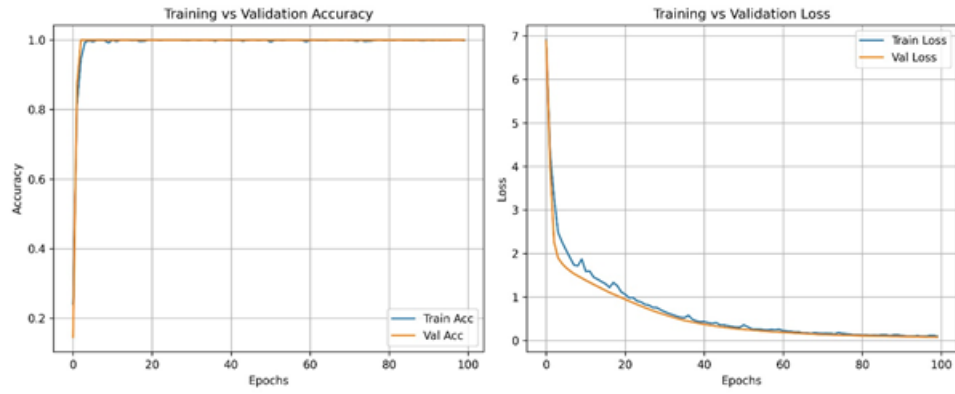


Figure 4.6. Training and Validation Learning Curves for the ARIA Dataset.

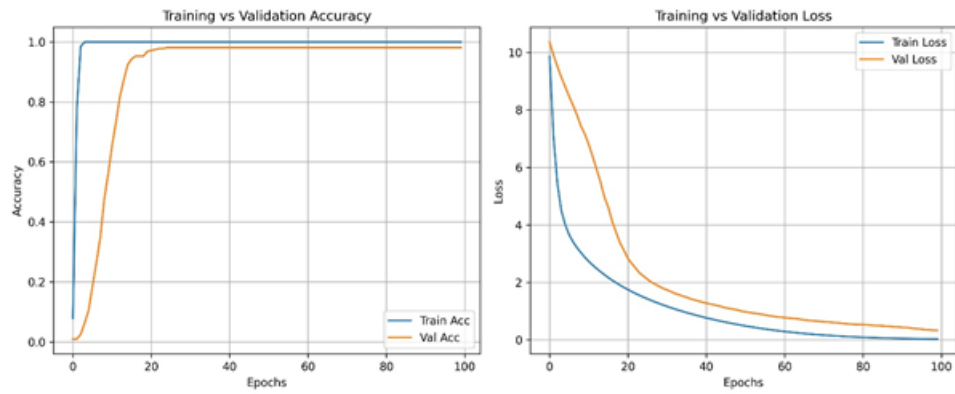


Figure 4.7. Training and Validation Learning Curves for the STARE Dataset.

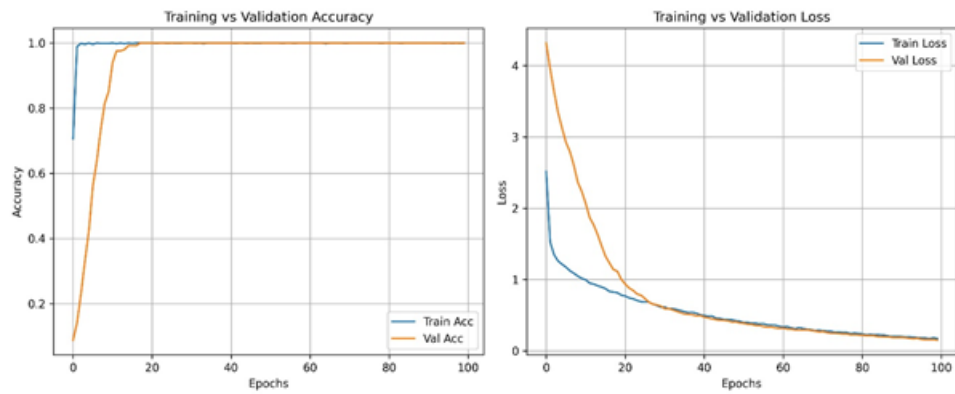


Figure 4.8. Training and Validation Learning Curves for the DRIVE Dataset.

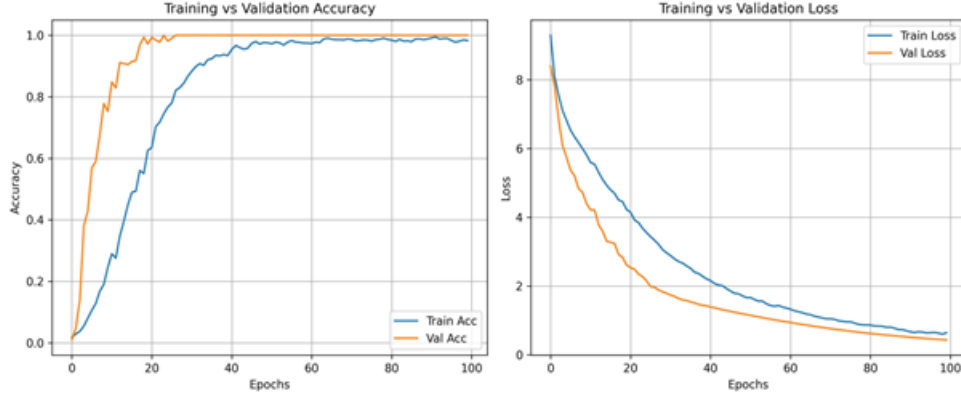


Figure 4.9. Training and Validation Learning Curves for the VARIA Dataset.

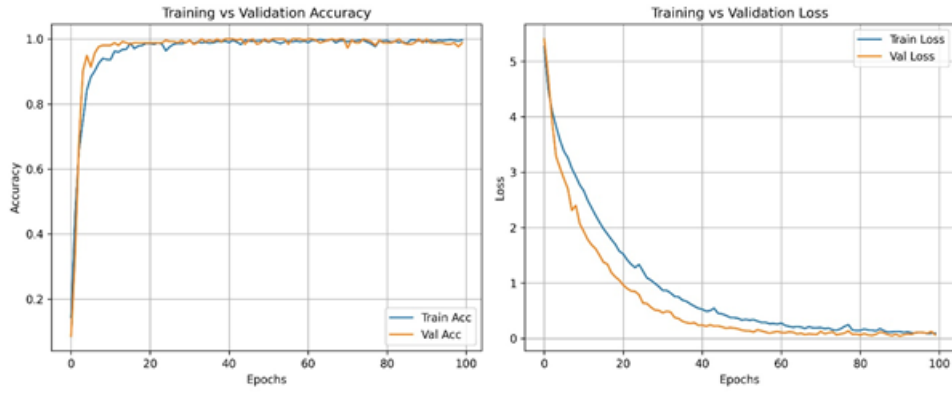


Figure 4.10. Training and Validation Learning Curves for the LOCAL Dataset.

4.8.8 Robustness Analysis

To further evaluate robustness, the system was tested against four common image perturbations: Gaussian noise $\mathcal{N}(0, \sigma_n^2)$, Salt-and-pepper noise with density d_{sp} , Gaussian blur with kernel size k_{blur} , and contrast variation with factor $c_{contrast}$ (see sample in Figure 4.11).

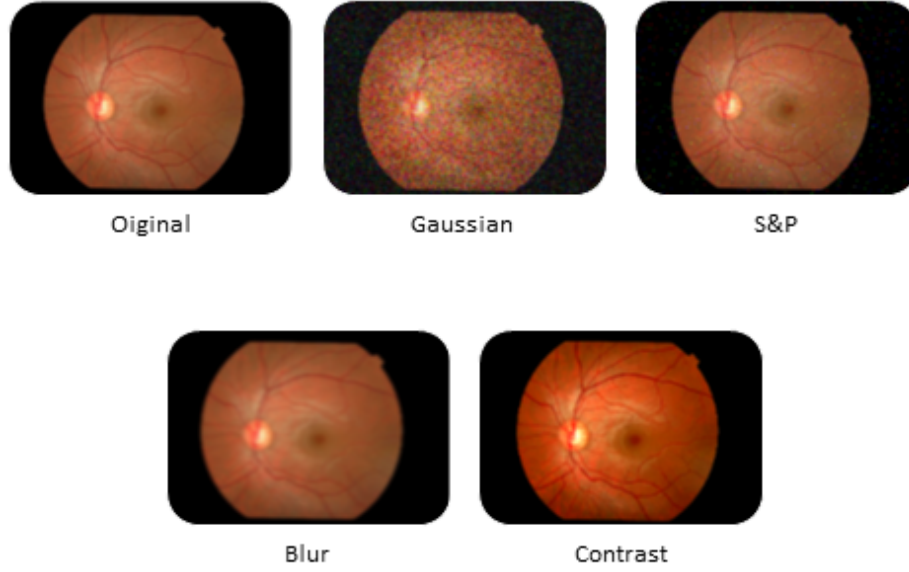


Figure 4.11. Example of a Perturbed Input Image.

The robustness analysis, summarized in Table 4.8, provides critical insights into the system’s performance stability under simulated real-world imaging imperfections. The results reveal a hierarchical vulnerability to different noise types and highlight dataset-specific resilience.

Table 4.8. Robustness Evaluation Under Noise Perturbations

Dataset	Gaussian	Salt-&Pepper	Blur	Contrast
RIDB	$\mathcal{A} = 1.000$ $\mathcal{L} = 0.0002$	$\mathcal{A} = 1.000$ $\mathcal{L} = 0.0003$	$\mathcal{A} = 0.986$ $\mathcal{L} = 0.038$	$\mathcal{A} = 1.000$ $\mathcal{L} = 0.004$
ARIA	$\mathcal{A} = 1.000$ $\mathcal{L} = 0.084$	$\mathcal{A} = 1.000$ $\mathcal{L} = 0.156$	$\mathcal{A} = 1.000$ $\mathcal{L} = 0.849$	$\mathcal{A} = 0.923$ $\mathcal{L} = 1.448$
STARE	$\mathcal{A} = 0.981$ $\mathcal{L} = 0.360$	$\mathcal{A} = 0.974$ $\mathcal{L} = 0.724$	$\mathcal{A} = 0.905$ $\mathcal{L} = 2.235$	$\mathcal{A} = 0.927$ $\mathcal{L} = 2.210$
DRIVE	$\mathcal{A} = 1.000$ $\mathcal{L} = 0.115$	$\mathcal{A} = 1.000$ $\mathcal{L} = 0.192$	$\mathcal{A} = 0.981$ $\mathcal{L} = 0.512$	$\mathcal{A} = 0.995$ $\mathcal{L} = 0.540$
VARIA	$\mathcal{A} = 1.000$ $\mathcal{L} = 0.408$	$\mathcal{A} = 1.000$ $\mathcal{L} = 0.408$	$\mathcal{A} = 1.000$ $\mathcal{L} = 0.409$	$\mathcal{A} = 1.000$ $\mathcal{L} = 0.409$
Local Data	$\mathcal{A} = 0.992$ $\mathcal{L} = 0.607$	$\mathcal{A} = 0.984$ $\mathcal{L} = 0.733$	$\mathcal{A} = 0.984$ $\mathcal{L} = 1.172$	$\mathcal{A} = 0.862$ $\mathcal{L} = 1.444$

The system demonstrates exceptional robustness to pixel-level stochastic noise. Accuracy $\mathcal{A}$ remains at or near 100% for Gaussian and Salt-and-Pepper noise across all datasets. This indicates that the learned features are highly stable and not reliant on exact pixel intensities, which is crucial for handling sensor noise or compression artifacts commonly found in clinical and portable imaging devices. The marginally higher loss values for salt-and-pepper noise (e.g., $\mathcal{L} = 0.724$ for STARE vs. $\mathcal{L} = 0.360$ for Gaussian) reflect the

more disruptive, high-amplitude nature of impulse noise, but this does not translate into a significant accuracy drop.

Performance under Gaussian blur reveals the model’s dependence on high-frequency structural information. While accuracy remains very high for most datasets (98.1%–100%), a notable decrease is observed for the STARE dataset ($\mathcal{A} = 90.5\%$). This significant drop, coupled with a substantial increase in loss ($\mathcal{L} = 2.235$), suggests that the distinctive biometric features in STARE are particularly fine-grained and are attenuated by low-pass filtering. This result underscores the importance of image focus and stability during acquisition for optimal system performance, especially with datasets containing subtle vascular patterns.

Contrast variation emerges as the most challenging perturbation, causing the most pronounced performance degradation. This is especially evident for the Local dataset ($\mathcal{A} = 86.2\%$) and, to a lesser extent, ARIA ($\mathcal{A} = 92.3\%$). The model’s struggle with global intensity shifts suggests that while it is invariant to localized noise, it has learned features that are partially correlated with absolute pixel value distributions or global histogram characteristics. This highlights a potential limitation for deployment across imaging systems with vastly different illumination protocols or for patients with varying retinal pigmentation. The perfect robustness of the VARIA dataset to all noise types, including contrast, is exceptional and may indicate that its images possess inherent properties or pre-processing that make features more invariant to these perturbations.

A consistent trend across all tests is the strong correlation between a drop in accuracy $\mathcal{A}$ and a concurrent sharp rise in the cross-entropy loss $\mathcal{L}$. For instance, the STARE dataset under blur shows accuracy falling to $\mathcal{A} = 90.5\%$ while loss spikes to $\mathcal{L} = 2.235$. Similarly, the Local dataset under contrast variation sees accuracy fall to $\mathcal{A} = 86.2\%$ with a loss of $\mathcal{L} = 1.444$. This correlation confirms that the model’s predictive confidence degrades in tandem with its correctness, which is a desirable and interpretable failure mode, as opposed to being confidently wrong.

4.8.9 Comparative Analysis with State-of-the-Art

This work provides a comparative evaluation of the proposed retinal biometric identification approach in relation to existing state-of-the-art methods. The comparative results, summarized in Table 4.9, report the identification accuracies $\mathcal{A}_{\text{id}}$ achieved by different approaches across several publicly available retinal datasets. While accuracy is an important performance metric in biometric systems, its interpretation must be considered alongside factors such as dataset characteristics, generalization capability, robustness to image quality variations, and computational requirements.

The table presents a comparative evaluation of retinal biometric methods across multiple

Table 4.9. Comparative Performance of Retinal Biometric Approaches

Method	Database	Identification	Authentication	Accuracy (%)	$\mathcal{A}$
Aleem et al. (2018)	DRIVE	✓		100	
	STARE			100	
	VARIA			99.57	
	RIDB			98	
Mazumdar and Nirmala (2018)	DRIVE		✓	100	
	HRF			100	
	Messidor			97.5	
	Clinical-1			97.5	
	Clinical-2			100	
Roy et al. (2019)	DRIVE	✓		92	
Rani and Jhansi Rani (2019)	DRIVE	✓		99.57	
Malik et al. (2020)	DRIVE	✓		100	
	STARE			99	
	VARIA			99.2	
Sultan and Ghanim (2020)	RIDB	✓		100	
	DRIVE			100	
Madenda et al. (2021)	DRIVE		✓	98	
Devi et al. (2020)	DRIVE		✓	97.16	
Our method	RIDB	✓		100	
	ARIA			100	
	STARE			99.99	
	DRIVE			100	
	Local Data			100	

datasets, highlighting their identification and authentication performance. Across previous studies, identification and authentication tasks show consistently high accuracy, though differences emerge depending on the dataset and the specific method employed.

Methods by Aleem et al. (2018) demonstrate near-perfect identification accuracy across DRIVE, STARE, VARIA, and RIDB, with values ranging from 98% to 100%. This indicates robust recognition capability, particularly on standard datasets such as DRIVE and STARE, while slightly lower performance on RIDB ($\mathcal{A} = 98\%$) suggests dataset-specific variability may affect identification accuracy.

Similarly, Mazumdar and Nirmala (2018) report consistently high identification accuracy across multiple datasets, including DRIVE, HRF, Messidor, and two clinical datasets, with a minimum accuracy of 97.5%. These results confirm that texture-based retinal features can generalize effectively across diverse imaging sources.

Other methods, such as Roy et al. (2019) and Rani and Jhansi Rani (2019), focus on authentication rather than identification. Roy et al. achieved 92% accuracy on DRIVE, while Rani et al. reported 99.57%. These results illustrate that authentication tasks, which often involve verifying individual identity rather than classifying across a dataset, may achieve slightly lower or more variable performance depending on the method and dataset characteristics.

Malik et al. (2020) achieved near-perfect identification on DRIVE ($\mathcal{A} = 100\%$), and high

performance on STARE ($\mathcal{A} = 99\%$) and VARIA ($\mathcal{A} = 99.2\%$), reflecting the effectiveness of multi-scale feature descriptors in capturing discriminative retinal vascular patterns. Sultan and Ghanim (2020) also demonstrate perfect authentication accuracy on RIDB and DRIVE ($\mathcal{A} = 100\%$), further confirming the robustness of Gabor-based approaches. Madenda et al. (2021) and Devi et al. (2020) achieve slightly lower authentication performance on DRIVE, 98% and 97.16% respectively, highlighting that deep learning methods may still be sensitive to dataset variability and training conditions.

The results obtained using the proposed method outperform or match previous approaches, achieving near-perfect identification accuracy across all evaluated datasets, including RIDB, ARIA, STARE, DRIVE, and Local data. Specifically, identification accuracies reach $\mathcal{A} = 100\%$ on RIDB, ARIA, DRIVE, and Local datasets, and $\mathcal{A} = 99.99\%$ on STARE. This consistent performance underscores the effectiveness of the proposed framework in both standard and locally acquired datasets, demonstrating strong generalization and robustness to dataset-specific variations.

Overall, the comparative analysis indicates that while previous methods achieve high performance on standard retinal datasets, their accuracy may vary across datasets or between identification and authentication tasks. In contrast, the proposed method achieves uniformly high identification accuracy, suggesting that the combination of feature extraction, model design, and optimization strategy provides superior discriminative power and generalizability. These results highlight the importance of validating retinal biometric methods across multiple datasets to ensure robust real-world applicability.

Although the proposed retinal biometric identification system achieves high identification accuracy and robustness, certain aspects require further investigation. Computational efficiency remains an important consideration, especially for large-scale deployments or resource-constrained environments. While inference latency $\bar{t}_{\text{id}}$ is low, optimization of the training process could improve scalability.

In addition, further validation on a wider range of retinal pathologies, including diabetic retinopathy, glaucoma, and age-related macular degeneration, would enhance the clinical relevance of the system. Variations observed in log-loss values $\mathcal{L}_{\text{log}}$ across datasets also suggest differences in classification confidence, indicating the potential need for confidence calibration techniques. Finally, robustness to real-world conditions, such as sensor variability, environmental factors, and adversarial perturbations, remains an important direction for future work.

4.9 Conclusion

This chapter presents a comprehensive investigation of hyperparameter optimization strategies for retinal biometric systems, emphasizing their critical role in achieving robust, high-performance models. We explored multiple optimization methods, including Grid Search, Random Search, Bayesian Optimization, and Covariance Matrix Adaptation Evolution Strategy (CMA-ES), evaluating their effectiveness across six diverse datasets: RIDB, ARIA, STARE, DRIVE, VARIA, and locally acquired retinal images. Among these methods, CMA-ES consistently demonstrated superior performance, achieving near-perfect identification accuracy $\mathcal{A}$ and stable convergence across datasets, while effectively navigating high-dimensional and anisotropic loss landscapes.

Through detailed analysis, we observed that key hyperparameters, including learning rate η , network size n_{neurons} , and dropout p_{dropout} , significantly influence convergence behavior and model generalization. Datasets with smooth, well-conditioned loss surfaces favored smaller learning rates and moderate regularization, allowing rapid convergence, whereas more complex or irregular datasets required adaptive step sizes $\sigma_{\text{step}}^{(g)}$ and covariance shaping $\mathbf{C}^{(g)}$ to efficiently explore the search space. Metrics derived from CMA-ES, such as final step size $\sigma_{\text{step}}^{(g)}$ and axis ratio γ_{axis} , provided further insights into how the optimization process adapts to dataset-specific characteristics, balancing exploration and exploitation for consistent performance.

The advantages of this approach are clear. By integrating CMA-ES with deep learning architectures, the framework achieves high identification accuracy, minimal validation loss $\mathcal{L}_{\text{val}}$, and computationally efficient inference $\bar{t}_{\text{id}}$. Adaptive hyperparameter tuning enhances robustness across datasets with variable imaging conditions and structural differences, reducing the risk of suboptimal convergence associated with static or grid-based search methods. These results underscore the necessity of systematic and adaptive hyperparameter optimization as a foundational component of reliable retinal biometric systems.

Nevertheless, certain limitations remain. Hyperparameter optimization, especially using evolutionary strategies, is computationally intensive and requires multiple iterations, which can constrain scalability in very large or highly dimensional search spaces. While the proposed framework effectively addresses dataset heterogeneity and pathological variations, hyperparameter tuning alone cannot inherently manage multiple concurrent tasks or integrate complementary biometric modalities. These observations highlight the importance of extending optimized architectures to more integrated frameworks capable of multitask and multimodal learning.

Chapter 5

Multitask Learning for Retinal Analysis

5.1 Introduction

Fundus photography provides a non-invasive view of retinal structures, including the vasculature, optic disc, and macula, making it an essential modality for diagnosing vision-threatening diseases such as diabetic retinopathy, glaucoma, and age-related macular degeneration. These disorders represent major causes of irreversible blindness worldwide, and early detection is critical for effective clinical intervention and disease management. Consequently, considerable research has focused on developing automated systems for retinal image analysis. Traditional approaches relied on manual clinical interpretation or handcrafted feature extraction, including measurements of vessel geometry, optic disc morphology, and lesion-specific patterns. However, these techniques are often sensitive to variations in imaging conditions such as illumination, acquisition devices, and camera settings, while clinical assessment may also suffer from inter-observer variability. These limitations have motivated the exploration of more robust data-driven approaches for retinal image interpretation.

5.2 Deep Learning Revolution in Retinal Image Analysis

In recent years, deep learning has fundamentally transformed retinal image analysis by enabling models to learn hierarchical representations directly from raw images. Convolutional neural networks and related architectures have demonstrated expert-level performance in detecting retinal diseases, often approaching or surpassing human diagnostic accuracy in

large-scale studies (Bilal et al., 2024; Iqbal et al., 2025). Nevertheless, real-world clinical data introduce challenges beyond single-disease recognition. Fundus images frequently contain multiple co-existing pathologies, subtle early-stage lesions, and considerable variability in disease manifestation, resulting in complex multi-label classification problems. To address these challenges, recent studies have introduced advanced architectural components, including attention mechanisms for lesion localization (Yan et al., 2024), graph-based models for capturing disease co-occurrence relationships (Sushith et al., 2025), and transformer architectures capable of modeling long-range contextual dependencies across retinal structures (Rieck et al., 2025). Despite these advances, most existing research remains primarily focused on pathology-centered objectives.

5.3 Retinal Fundus Imaging Beyond Diagnosis: Biometric Recognition

Beyond disease detection, retinal fundus images also contain highly distinctive vascular patterns that can serve as reliable biometric identifiers. Retinal biometrics offer several advantages over conventional modalities such as fingerprints or facial recognition, as the vascular structure is internal, stable, and difficult to replicate, providing strong resistance to spoofing attacks. Early retinal recognition systems relied on two-stage pipelines in which vessel segmentation, which was often performed using methods such as Frangi filtering or U-Net architectures, was followed by graph-based matching or template comparison (Ochoa-Astorga et al., 2024). Although effective, these pipelines are computationally intensive and sensitive to segmentation errors. More recently, the field has transitioned toward end-to-end deep learning approaches that learn discriminative biometric representations directly from retinal images. Examples include CNN-based authentication systems (Aicha and Eddine, 2024), multimodal biometric fusion strategies (Patil and Jagtap, 2020), template protection mechanisms (Sadeghpour et al., 2023), and neural network-based biometric key generation frameworks (Roy and Biswas, 2020). These developments highlight the increasing maturity of retinal biometrics as a secure identification technology.

5.4 Privacy and Ethical Considerations

The ability of retinal images to simultaneously encode both biometric identity information and clinically relevant biomarkers raises important ethical and privacy considerations. Gim et al. (2025) argue that formally classifying retinal imaging as biometric data could introduce restrictive regulatory barriers that may hinder biomedical research and biomarker discovery. At the same time, recent studies demonstrate the extensive physiological

information that can be extracted from retinal images, emphasizing the potential risk of unintended identity inference or misuse of sensitive biometric attributes (Hu et al., 2025). To mitigate these risks, privacy-preserving approaches such as RetinaGuard have been proposed, which employ generative masking techniques to obscure identity-related features while preserving diagnostically relevant retinal information (Luo et al., 2025). These efforts highlight the need to balance technological innovation with responsible data governance in retinal imaging research.

5.5 The Persistent Divide Between Diagnosis and Biometrics

Despite rapid progress in automated pathology detection and retinal biometric recognition, these two research domains have largely developed independently. Most existing systems treat disease diagnosis and biometric identification as separate tasks, even though both rely on the same anatomical structures. This separation introduces practical limitations: pathological changes such as diabetic retinopathy may alter vascular structures and reduce biometric identification accuracy. At the same time, diagnostic models that ignore identity-specific variations may be less robust across diverse populations. These observations reveal an important research gap and raise a fundamental question: can a unified framework simultaneously exploit retinal structures for both identification and disease analysis?

5.6 Multitask Learning as a Unifying Paradigm

Multitask learning (MTL) offers a promising solution to this challenge by enabling a single model to learn multiple related tasks simultaneously. By sharing representations across tasks, MTL can improve generalization, reduce overfitting, and enhance computational efficiency through inductive knowledge transfer (Zhang and Yang, 2021). In medical imaging, multitask frameworks have demonstrated strong performance in applications such as joint segmentation and classification, where anatomical localization improves diagnostic accuracy (Ling et al., 2023; Huang et al., 2023), as well as simultaneous detection of multiple lesion types within a unified architecture (Hervella et al., 2022). In ophthalmology, multitask models have been applied to tasks such as joint optic disc and optic cup segmentation for glaucoma analysis (Iqbal et al., 2025) or combined lesion detection and disease grading. However, existing multitask approaches remain largely restricted to diagnostic objectives and have rarely explored integrating biometric identification with pathological analysis.

5.7 Contributions of This Work

To address this gap, this thesis proposes a unified multitask deep learning framework capable of simultaneously performing retinal biometric identification and pathology classification from fundus images. The central hypothesis is that a shared representation can capture both the fine-grained vascular patterns necessary for identity discrimination and the broader morphological features associated with retinal disease. The proposed architecture integrates two complementary backbone networks, ResNet-152 and DenseNet-201, to exploit both hierarchical and densely connected feature representations. These features are further enhanced using Convolutional Block Attention Modules (CBAM) and a Selective Feature Aggregation (SFA) mechanism to emphasize task-relevant retinal regions. The network produces two specialized output branches: an embedding-based head for biometric verification and a sigmoid-based head for pathology classification.

5.8 Materials and Methods

This section describes the proposed unified multitask deep learning framework for retinal biometric identification and pathology classification. It details the data preprocessing pipeline, architectural design, optimization strategy, and evaluation protocol. The framework is designed to be robust, reproducible, and generalizable across heterogeneous clinical datasets with varying acquisition conditions and annotation standards.

To validate the proposed approach, experiments were conducted on three publicly available retinal fundus datasets: 39-CAS, ODIR-2019, and FIVES. These datasets differ substantially in pathology prevalence, labeling protocols, image quality, and class balance, providing a realistic benchmark for evaluating system robustness. This multi-dataset experimental design ensures that the framework generalizes beyond any single curated dataset. Experimental results confirm that joint learning consistently outperforms single-task baselines, demonstrating the mutual benefit of simultaneously optimizing biometric and pathological objectives.

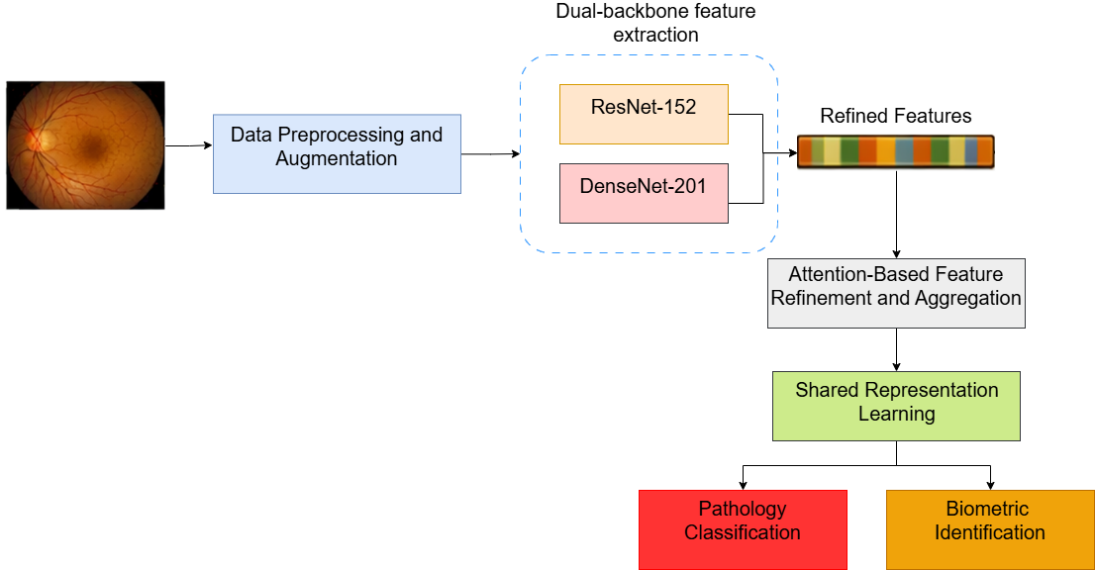


Figure 5.1. Unified multitask learning framework for simultaneous retinal biometric identification and pathology classification.

5.8.1 Data Preprocessing and Augmentation

All retinal fundus images underwent a standardized preprocessing pipeline to ensure spatial and intensity consistency. Images were resized to 128×128 pixels for efficient batch processing, and pixel intensities were normalized to the $[0,1]$ range to reduce inter-device variability.

To enhance generalization while preserving anatomical structures, a comprehensive data augmentation strategy was applied. Geometric transformations were applied to simulate natural variability in eye positioning and camera orientation:

- Random rotations $\phi \in [-10^\circ, +10^\circ]$, modeling minor acquisition angle deviations.
- Horizontal flipping with probability $p_{\text{flip}} = 0.5$, exploiting the approximate bilateral symmetry of retinal anatomy.
- Small translations $\Delta x, \Delta y \leq 0.05 \times H$, accounting for capture misalignment.
- Mild zooming with scaling factors $s_{\text{zoom}} \in [1.0, 1.15]$, simulating variations in field-of-view and optical magnification.

To capture differences in imaging conditions and sensor characteristics, photometric transformations were applied:

- Brightness adjustments $\Delta b \in [-0.1, +0.1]$, compensating for under- or overexposure
- Contrast variations $\Delta c \in [-0.1, +0.1]$, modeling differences in vessel and lesion visibility
- Gamma correction $\gamma \in [0.9, 1.1]$, simulating non-linear intensity distortions

All augmentations were applied offline, generating $N_{\text{aug}} = 6$ additional synthetic samples per original image. This strategy significantly increased the effective training set size while maintaining anatomical plausibility. The resulting dataset sizes were:

- **39-CAS:** $N_{\text{total}} = 7,000$ images
- **FIVES:** $N_{\text{total}} = 5,600$ images
- **ODIR-2019:** $N_{\text{total}} = 70,000$ images

Collectively, this preprocessing and augmentation framework mitigates overfitting, improves generalization, and ensures that both biometric and pathological features are preserved in a clinically realistic manner. It establishes a robust foundation for the subsequent dual-backbone feature extraction and multitask learning stages.

5.8.2 Proposed Unified Multitask Architecture

The proposed framework jointly performs retinal biometric identification and multi-label pathology classification within a single end-to-end trainable network. The design is motivated by the observation that both tasks depend on the same anatomical substrate—retinal vascular and structural patterns. The architecture therefore learns a shared representation capable of encoding both identity-discriminative vascular topology and pathology-related morphological features. As illustrated in Figure 5.1, the network comprises four main components: (i) dual-backbone feature extraction, (ii) attention-based feature refinement, (iii) shared representation learning, and (iv) task-specific output heads.

Dual-Backbone Feature Extraction

Feature extraction is performed using a dual-backbone architecture combining ResNet-152 and DenseNet-201 in parallel. This design captures complementary information: hierarchical semantic abstraction through residual learning and dense feature reuse through dense connectivity. The framework presented in Figure 5.2.

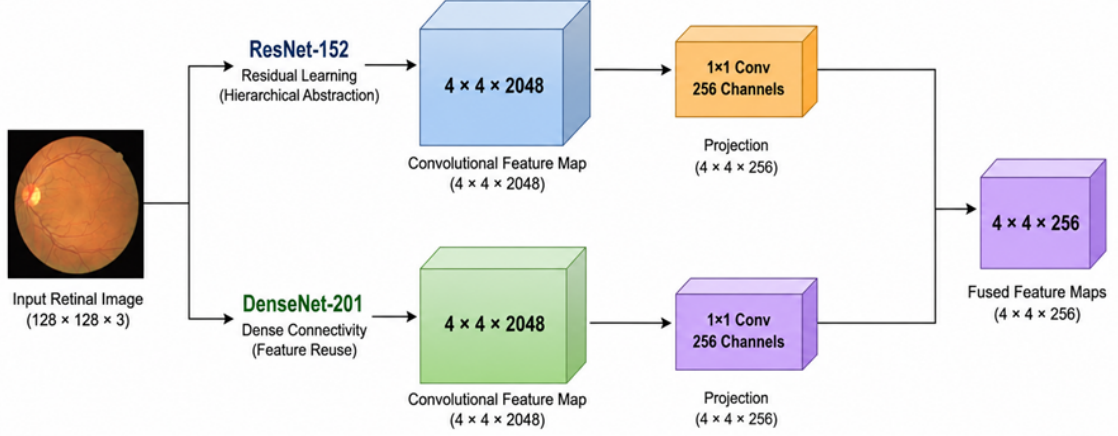


Figure 5.2. Dual-backbone feature extraction module for the proposed multitask retinal analysis system.

- **ResNet-152** emphasizes hierarchical abstraction through residual connections:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \mathbf{W}) + \mathbf{x} \quad (5.8.1)$$

where $\mathbf{x}$ represents the input feature map, $\mathcal{F}$ denotes the residual mapping with weights $\mathbf{W}$, and $\mathbf{y}$ is the output of the block. This formulation facilitates effective gradient propagation during backpropagation and enables stable optimization of very deep networks. Within the context of retinal analysis, the progressive increase in receptive field size across layers allows the network to integrate information across the entire retinal image. Early layers capture vessel edges and local texture, intermediate layers encode vascular arcades and optic disc contours, and deeper layers integrate high-level anatomical structure. For a $128 \times 128 \times 3$ input, the final convolutional output is $\mathbf{F}_{\text{ResNet}} \in \mathbb{R}^{4 \times 4 \times 2048}$.

- **DenseNet-201** employs dense connectivity where each layer receives the concatenation of all preceding feature maps:

$$\mathbf{x}_l = \mathcal{H}_l([\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{l-1}]) \quad (5.8.2)$$

where $[\cdot]$ denotes channel-wise concatenation and $\mathcal{H}_l$ represents a composite transformation consisting of batch normalization, ReLU activation, and convolution. This dense connectivity pattern ensures that low-level features remain directly accessible throughout the network depth, promoting feature reuse and strengthening gradient flow. For the same input resolution, DenseNet-201 produces a final convolutional feature map $\mathbf{F}_{\text{DenseNet}} \in \mathbb{R}^{4 \times 4 \times 2048}$.

After extracting features from both backbones, the resulting tensors are ready for in-

tegration. The direct concatenation of the two 2048-channel feature maps produces a 4096-channel tensor, substantially increasing computational cost and potentially introducing representational imbalance. To address this, 1×1 convolutional projection layers reduce the channel dimension from $C_{\text{in}} = 2048$ to $C_{\text{proj}} = 256$. For a given feature tensor $\mathbf{F} \in \mathbb{R}^{4 \times 4 \times 2048}$, the projection operation computes:

$$\hat{\mathbf{F}}_{i,j,k} = \sigma_{\text{ReLU}} \left(\sum_{c=1}^{2048} \mathbf{W}_{k,c} \mathbf{F}_{i,j,c} + \mathbf{b}_k \right) \quad (5.8.3)$$

where $\mathbf{W} \in \mathbb{R}^{256 \times 2048}$ denotes learnable weights, $\mathbf{b} \in \mathbb{R}^{256}$ represents biases, and σ_{ReLU} is the ReLU activation function. This transformation compresses the channel dimension from 2048 to 256 while preserving spatial resolution, yielding projected tensors $\mathbf{P}_R, \mathbf{P}_D \in \mathbb{R}^{4 \times 4 \times 256}$. This projection preserves spatial resolution while reducing computational complexity.

Following projection, feature integration is performed using a Selective Feature Aggregation mechanism based on a trainable scalar parameter α_{fusion} . The fused representation is defined as:

$$\mathbf{F}_{\text{Fused}} = \alpha_{\text{fusion}} \mathbf{P}_R + (1 - \alpha_{\text{fusion}}) \mathbf{P}_D \quad (5.8.4)$$

where α is initialized to 0.5 and optimized jointly with all network parameters, dynamically adjusting the relative contribution of each backbone.

Attention-Based Feature Refinement and Aggregation

The fused feature representation is refined using the Convolutional Block Attention Module (CBAM), which sequentially applies channel and spatial attention to emphasize informative features. The overall process is illustrated in Figure 5.3.

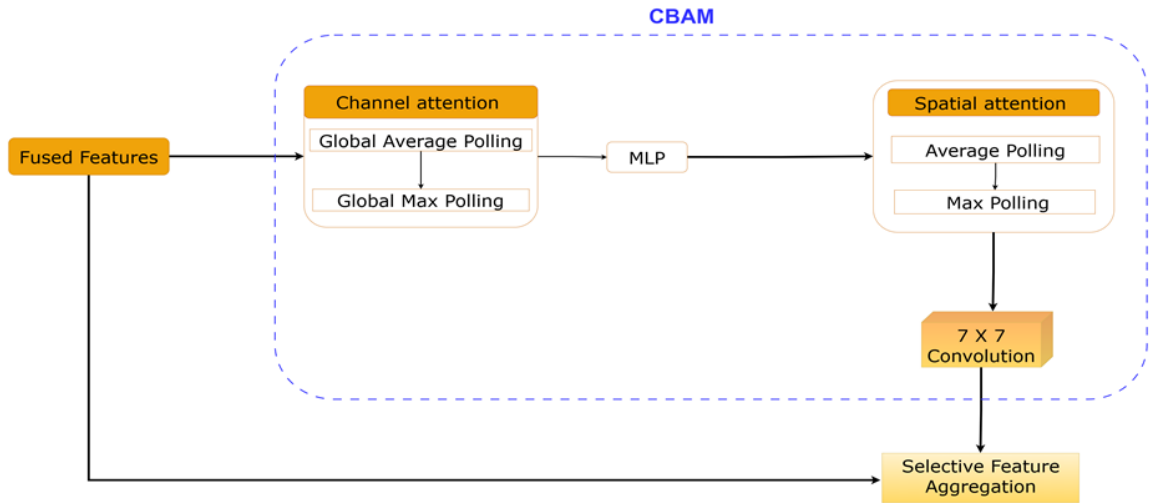


Figure 5.3. Attention-Based Feature Refinement and Aggregation process.

Channel Attention Module. The channel attention (CA) mechanism computes a weight for each feature channel, capturing its relative importance. The computation is performed using both global average pooling (GAP) and global max pooling (GMP) across spatial dimensions to encode complementary global descriptors:

$$\mathbf{z}_{\text{avg}} = \text{GAP}(\mathbf{F}_{\text{fused}}) \in \mathbb{R}^{256}, \quad \mathbf{z}_{\text{max}} = \text{GMP}(\mathbf{F}_{\text{fused}}) \in \mathbb{R}^{256} \quad (5.8.5)$$

These descriptors are passed through a shared multi-layer perceptron (MLP) with one hidden layer to compute channel attention weights:

$$\mathbf{M}_c = \sigma_{\text{sig}}(\text{MLP}(\mathbf{z}_{\text{avg}}) + \text{MLP}(\mathbf{z}_{\text{max}})) \in \mathbb{R}^{256} \quad (5.8.6)$$

where σ_{sig} denotes the sigmoid activation function. The channel-refined feature tensor is then computed as:

$$\mathbf{F}' = \mathbf{M}_c \odot \mathbf{F}_{\text{fused}} \quad (5.8.7)$$

where $\odot$ denotes element-wise multiplication. This operation reweights channels according to their relevance, enhancing features associated with vascular patterns for biometric identification or lesion textures for pathology classification.

Spatial Attention Module. Following channel refinement, the spatial attention (SA) mechanism highlights important locations within the feature map. Spatial attention uses channel-wise pooling (average and max) to condense the feature map into two 2D maps:

$$\mathbf{F}_{\text{avg}}^s = \frac{1}{256} \sum_{c=1}^{256} \mathbf{F}'_c, \quad \mathbf{F}_{\text{max}}^s = \max_{c=1, \dots, 256} \mathbf{F}'_c \quad (5.8.8)$$

These maps are concatenated and convolved with a 7×7 kernel to compute the spatial attention map:

$$\mathbf{M}_s = \sigma_{\text{sig}}\left(f^{7 \times 7}([\mathbf{F}_{\text{avg}}^s, \mathbf{F}_{\text{max}}^s])\right) \in \mathbb{R}^{H \times W} \quad (5.8.9)$$

where $[\cdot]$ denotes concatenation along the channel dimension. The final spatially-refined feature map is:

$$\mathbf{F}'' = \mathbf{M}_s \odot \mathbf{F}' \quad (5.8.10)$$

This process enables the network to focus on the diagnostically salient regions, ensuring

that fine-grained microvascular and lesion-level features are emphasized while irrelevant retinal background is suppressed.

Selective Feature Aggregation. After attention refinement, the feature maps are aggregated to produce a task-adaptive unified representation. This module applies trainable channel-wise and spatial weights to balance contributions from different regions and feature maps. Unlike static concatenation, selective aggregation allows the network to adaptively emphasize features depending on the current task gradient:

$$\mathbf{F}_{\text{agg}} = \beta_{\text{attn}} \mathbf{F}'' + (1 - \beta_{\text{attn}}) \mathbf{F}_{\text{fused}} \quad (5.8.11)$$

where β_{attn} is a trainable scalar parameter that dynamically regulates the contribution of the attention-refined representation relative to the original fused features. By learning this coefficient during optimization, the network acquires the capacity to adaptively balance refinement and preservation. This formulation is particularly important in a multitask setting, where excessive emphasis on attention-weighted activations could potentially suppress subtle but identity-discriminative vascular cues. Conversely, insufficient attention weighting might reduce sensitivity to localized pathological manifestations. The learnable aggregation mechanism therefore ensures architectural flexibility, enabling the model to retain residual structural information essential for biometric recognition while simultaneously benefiting from attention-guided enhancement of pathology-relevant features. An overview of the proposed process is presented in Figure 5.4.

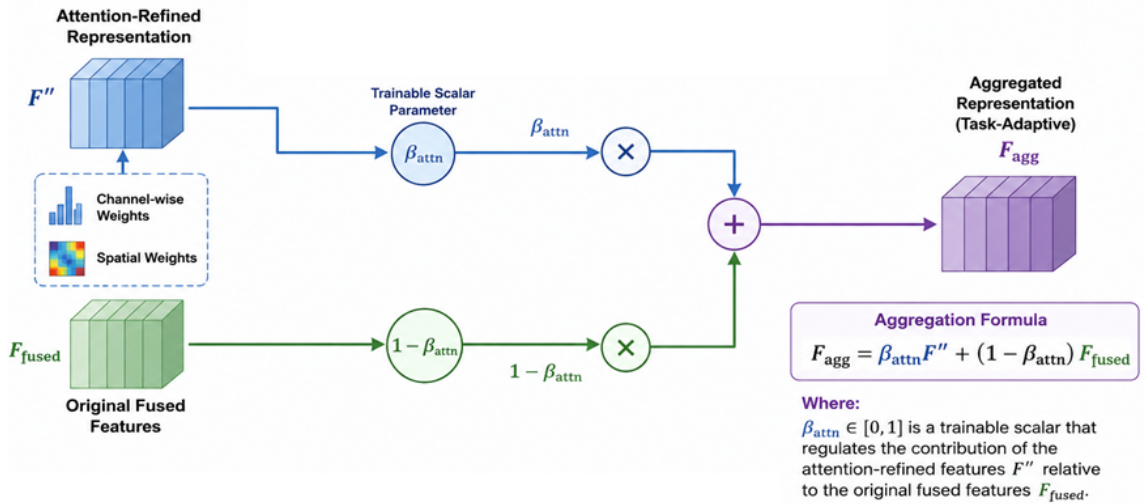


Figure 5.4. Selective Feature Aggregation process.

Shared Representation Learning

The objective of this stage is to transform the refined features into a compact latent embedding that jointly supports biometric identification and pathology classification. The aggregated feature map $\mathbf{F}_{\text{agg}} \in \mathbb{R}^{256 \times 4 \times 4}$ is subsequently reduced via Global Average Pooling (GAP) to a feature vector:

$$\mathbf{z} = \text{GAP}(\mathbf{F}_{\text{agg}}) \in \mathbb{R}^{256}. \quad (5.8.12)$$

This vector is transformed through a fully connected layer with $d_{\text{shared}} = 512$ neurons, followed by batch normalization, ReLU activation, and dropout ($p_{\text{dropout}} = 0.5$), yielding the shared latent representation $\mathbf{h} \in \mathbb{R}^{512}$.

The shared embedding is optimized under a weighted multitask loss:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{id}} \mathcal{L}_{\text{id}} + \lambda_{\text{path}} \mathcal{L}_{\text{path}} \quad (5.8.13)$$

where $\mathcal{L}_{\text{id}}$ is identity classification loss (categorical cross-entropy), $\mathcal{L}_{\text{path}}$ is pathology classification loss (binary cross-entropy), and $\lambda_{\text{id}} = 0.5$, $\lambda_{\text{path}} = 1.0$ are task weighting coefficients. A schematic representation of the overall framework is provided in Figure 5.5.

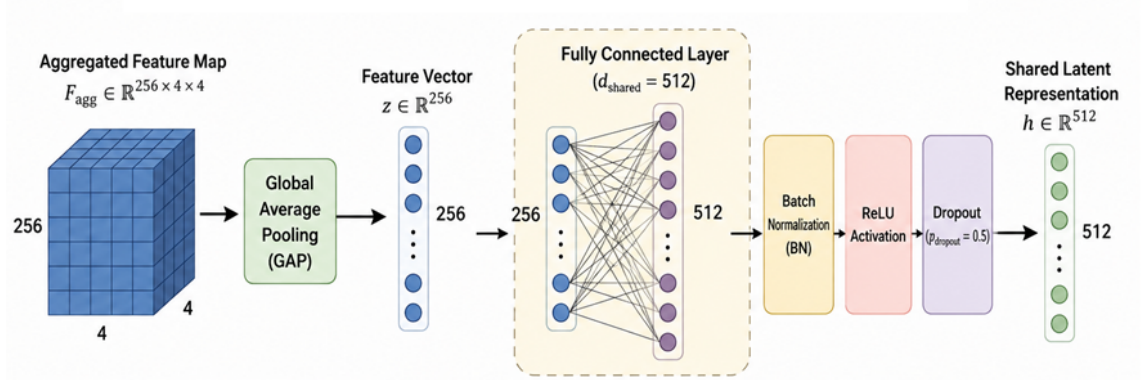


Figure 5.5. Shared Representation Learning Stage.

Task-Specific Output Heads

From the shared representation $\mathbf{h}$, two specialized output heads produce task-specific predictions.

For biometric identification, a fully connected layer with softmax activation outputs identity probabilities:

$$\hat{\mathbf{y}}_{\text{id}} = \text{softmax}(\mathbf{W}_{\text{id}} \mathbf{h} + \mathbf{b}_{\text{id}}) \quad (5.8.14)$$

where $\mathbf{W}_{\text{id}} \in \mathbb{R}^{N_{\text{id}} \times 512}$ and N_{id} is the number of enrolled identities.

For pathology classification, independent sigmoid-activated neurons handle 1 prediction:

$$\hat{\mathbf{y}}_{\text{path}} = \sigma_{\text{sig}}(\mathbf{W}_{\text{path}}\mathbf{h} + \mathbf{b}_{\text{path}}) \quad (5.8.15)$$

where $\mathbf{W}_{\text{path}} \in \mathbb{R}^{N_{\text{path}} \times 512}$ and N_{path} is the number of pathology classes.

Both branches are trained jointly through the combined loss objective $\mathcal{L}_{\text{total}}$, ensuring that gradients from identity discrimination and pathology recognition collectively shape the shared feature space. This dual supervision encourages the network to learn representations that are simultaneously discriminative for biometric purposes and sensitive to clinically relevant morphological variations. The workflow of this step is illustrated in Figure 5.6.

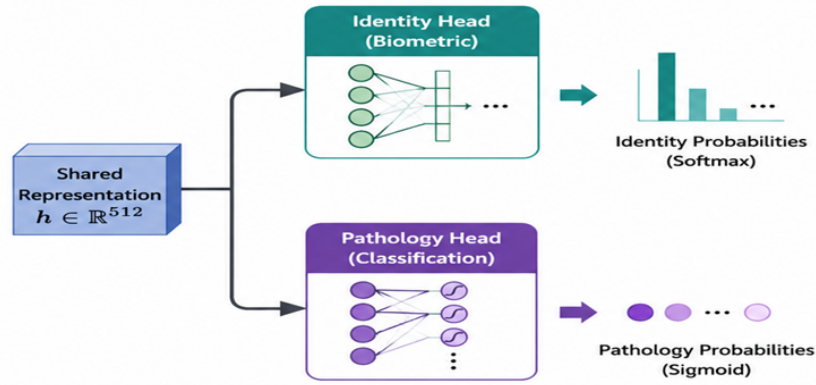


Figure 5.6. Task-Specific Output Heads Step.

5.8.3 Training Protocol

Prior to training, a stratified K-fold data partitioning strategy is employed to ensure a balanced and representative split of the dataset across all classes. Specifically, the data is divided using a stratified 80%–20% split, where class distributions are preserved in both subsets. The 80% portion is used for model training, while the remaining 20% is reserved for independent testing. This stratification ensures that both training and testing sets maintain consistent class proportions, thereby reducing sampling bias and improving the reliability of performance evaluation.

The model is trained using the Adamax optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with a two-stage fine-tuning strategy. In the first stage (epochs 1–500), backbone layers remain frozen while newly introduced layers are trained with learning rate $\eta = 10^{-3}$. In the second stage (epochs 501–1000), the final convolutional layers of each backbone are unfrozen and learning rate is reduced to $\eta = 10^{-4}$. Learning rate scheduling reduces η by factor $\gamma = 0.5$ if no improvement occurs for 8 epochs.

To mitigate class imbalance, particularly in pathology classification, sample weights are computed as:

$$w_k = \frac{N}{K \cdot N_k} \quad (5.8.16)$$

where N is the total number of samples, K is the number of classes, and N_k represents the number of samples in class k .

To prevent overfitting and stabilize training in the high-capacity dual-backbone architecture, several regularization techniques are employed, including L2 weight decay with coefficient $\lambda_{\text{wd}} = 10^{-4}$, dropout layers with rates $p_{\text{dropout}} = 0.5, 0.4$, and 0.3 at different stages, and batch normalization. Model checkpoints are saved based on validation performance for pathology classification to ensure reproducible and stable outcomes.

Training is conducted with a batch size of $B = 8$ images per iteration. The multitask loss is weighted with $\lambda_{\text{id}} = 0.5$ for biometric identification and $\lambda_{\text{path}} = 1.0$ for pathology classification, reflecting the relative importance of each task. Collectively, this protocol provides a stable and reproducible learning framework, preserving discriminative vascular features while maintaining sensitivity to disease-related morphological variations, enabling the unified network to perform effectively across both tasks.

5.9 Results and Discussion

The proposed multitask framework was evaluated on three public datasets: ODIR-2019 (8 classes), 39-CAS (39 classes), and FIVES (4 classes). This multi-dataset design ensures evaluation across diverse clinical scenarios.

5.9.1 Results on ODIR-2019 Dataset

The ODIR-2019 dataset comprises fundus images from both left and right eyes of patients, with annotations covering $N_{\text{path}} = 8$ diagnostic categories: Normal, Diabetic Retinopathy (DR), Glaucoma, Cataract, Age-related Macular Degeneration (AMD), Hypertension, Myopia, and Other Diseases. This dataset presents a challenging multi-label classification scenario where multiple conditions may co-exist within the same patient.

Pathology Classification Performance

The results presented in Table 5.1 demonstrate strong overall performance of the proposed model on the ODIR-2019 dataset. The system achieves high precision (0.95 micro, 0.97 macro), indicating that most predicted disease labels correspond to correct diagnoses with very few false positives. The recall value of 0.88 suggests that the majority of pathological cases are successfully detected, although a small proportion of positive instances may remain undetected.

The resulting F1-scores (0.91 micro and 0.92 macro) reflect a well-balanced trade-off between precision and recall across different disease categories. The overall accuracy of 0.853 confirms reliable multi-label classification performance at the image level. Furthermore, the low Hamming Loss of 0.024 indicates that only a small fraction of label assignments are incorrect, highlighting the stability of the multi-label predictions.

Finally, the very high macro ROC-AUC value of 0.993 demonstrates excellent discriminative capability, showing that the model can effectively separate pathological and non-pathological conditions. Overall, these results confirm the robustness and reliability of the proposed framework for retinal pathology classification on the ODIR-2019 dataset.

Table 5.1. Overall Evaluation Metrics on ODIR-2019

Metric	$\mathcal{A}$	P_{micro}	P_{macro}	R_{micro}	R_{macro}	$F1_{\text{micro}}$	$F1_{\text{macro}}$	$\mathcal{L}_H$	$\text{AUC}_{\text{macro}}$
Value	0.95	0.88	0.91	0.97	0.88	0.92	0.853	0.024	0.993

Per-class analysis (Table 5.2) indicates very strong classification performance across most retinal conditions. Cataract and Myopia achieve near-perfect results ($F1 = 0.99$ – 1.00), reflecting highly reliable detection. Glaucoma ($F1 = 0.94$) and the Normal class ($F1 = 0.93$) also demonstrate strong predictive accuracy. Diabetic Retinopathy ($F1 = 0.89$) and Age-related Macular Degeneration (AMD) ($F1 = 0.91$) exhibit high precision but comparatively lower recall (0.81 – 0.84), suggesting that some cases, particularly in early stages, may remain undetected. Hypertension presents the lowest performance ($F1 = 0.83$), which may be attributed to the limited representation of this class and the similarity of its vascular features to other retinal conditions.

Table 5.2. Per-Class Pathology Classification Results on ODIR-2019

Class	P	R	$F1$
Normal	0.91	0.94	0.93
Diabetic Retinopathy	0.97	0.81	0.89
Glaucoma	0.99	0.90	0.94
Cataract	0.99	0.99	0.99
AMD	1.00	0.84	0.91
Hypertension	0.96	0.73	0.83
Myopia	1.00	0.99	1.00
Other Diseases	0.95	0.87	0.90

Biometric Identification Performance

Table 5.3 presents the biometric identification results on the ODIR-2019 dataset. The model achieves an identification accuracy $\mathcal{A}_{\text{id}} = 0.9898$, with F1-score $F1 = 0.9897$. The Genuine Accept Rate $\text{GAR} = 0.9898$, while both the False Reject Rate FRR and False Accept Rate FAR are 0.0102. These results indicate that the model reliably extracts discriminative features for individual identification from retinal fundus images. The high precision and recall values further demonstrate that the model maintains a near-perfect balance between correctly accepting genuine identities and rejecting impostors, highlighting the feasibility of integrating retinal biometrics for secure authentication.

Table 5.3. Biometric Identification Results on ODIR-2019

Metric	$\mathcal{A}_{\text{id}}$	P	R	$F1$	GAR	FRR	FAR
Value	0.9898	0.9930	0.9898	0.9897	0.9898	0.0102	0.0102

5.9.2 Results on 39-CAS Dataset

The 39-CAS dataset represents a significantly more fine-grained and clinically diverse benchmark, comprising $N_{\text{path}} = 39$ retinal pathology categories, including vascular disorders, optic disc abnormalities, macular diseases, degenerative conditions, inflammatory pathologies, and rare retinal syndromes. In addition, the dataset includes $N_{\text{id}} = 502$ identity classes, making it a comprehensive evaluation platform for the proposed unified multitask framework. Compared to ODIR-2019, 39-CAS introduces higher granularity and greater inter-class similarity, particularly among disease subtypes and progressive stages.

Pathology Classification Performance

The results in Table 5.4 demonstrate exceptionally strong classification performance on the 39-CAS dataset. The model achieves very high overall accuracy (0.9953), indicating that nearly all labels are correctly predicted. The micro-averaged precision, recall, and F1-score are identical (0.9953), reflecting highly consistent predictions across the dataset. Slightly lower macro precision (0.9769) and recall (0.9701) suggest minor variation between classes, although performance remains uniformly high. The extremely low Hamming Loss (0.0002) indicates that incorrect label assignments are extremely rare, while the macro ROC-AUC of 1.0000 confirms excellent class separability. Overall, these results demonstrate that the proposed framework provides highly accurate and reliable pathology classification on the 39-CAS dataset.

Table 5.4. Overall Evaluation Metrics on 39-CAS

Metric	$\mathcal{A}$	P_{micro}	P_{macro}	R_{micro}	R_{macro}	$F1_{\text{micro}}$	$F1_{\text{macro}}$	$\mathcal{L}_H$	$\text{AUC}_{\text{macro}}$
Value	0.9953	0.9953	0.9769	0.9953	0.9701	0.9953	0.9953	0.0002	1.0000

The per-class analysis on the 39-CAS dataset provides detailed insight into the classification performance across different retinal pathology groups.

Normal and Baseline Categories. Table 5.5 presents the classification results for Normal and baseline retinal categories. Both the *Normal* and *Tessellated fundus* classes achieved perfect precision, recall, and F1-scores (1.0000), indicating flawless identification of normal retinal structures without false positives or false negatives. This demonstrates that the model effectively captures physiological retinal features and can reliably distinguish them from pathological patterns.

The *Large optic cup* class showed slightly lower, yet still excellent performance ($P = R = F1 = 0.9833$). The minor reduction in metrics is likely due to its subtle structural similarity with early glaucomatous changes, which can introduce slight ambiguity in feature boundaries. Despite this, the results confirm the robustness and clinical reliability of the proposed framework in establishing a strong baseline for retinal classification.

Table 5.5. Classification Performance for Normal / Baseline Categories on 39-CAS

Class	P	R	$F1$
Normal	1.0000	1.0000	1.0000
Tessellated fundus	1.0000	1.0000	1.0000
Large optic cup	0.9833	0.9833	0.9833

Diabetic Retinopathy Categories. Table 5.6 presents the classification results for the various Diabetic Retinopathy (DR) categories on the 39-CAS dataset. All DR classes achieved exceptionally high precision, recall, and F1-scores, ranging from 0.9832 to 1.0000, indicating near-perfect identification of DR severity levels without significant false positives or false negatives.

Specifically, **DR1** reached perfect performance ($P = R = F1 = 1.0000$), demonstrating that the model can accurately identify mild DR cases. **DR2** and **DR3** also exhibited excellent results ($F1 = 0.9832$ and 0.9958 , respectively), showing reliable detection of moderate to severe DR stages. The blurred fundus categories, both without PDR and with suspected PDR, maintained similarly high performance ($F1 = 0.9959$ and 0.9958), reflecting the model’s robustness to imaging artifacts and its ability to extract meaningful features even in challenging conditions.

Table 5.6. Classification Performance for Diabetic Retinopathy Categories on 39-CAS

Class	P	R	$F1$
DR1	1.0000	1.0000	1.0000
DR2	0.9832	0.9832	0.9832
DR3	0.9917	1.0000	0.9958
Blur fundus without PDR	0.9917	1.0000	0.9959
Blur fundus with suspected PDR	1.0000	0.9916	0.9958

Glaucoma and Optic Nerve–Related Pathologies. Table 5.7 presents the classification performance for glaucoma and optic nerve–related pathologies on the 39-CAS dataset. The results demonstrate excellent accuracy across all classes, with F1-scores ranging from 0.9872 to 1.0000.

The **Possible glaucoma** class achieved perfect precision (1.0000) and slightly lower recall (0.9748), indicating a very small number of false negatives. Conversely, **Optic atrophy** showed perfect recall (1.0000) and slightly lower precision (0.9756), suggesting minimal false positives. All other classes, including **Disc swelling and elevation**, **Dragged disc**, **Congenital disc abnormality**, and **Myelinated nerve fiber**, reached perfect or near-perfect F1-scores (≥ 0.9958), demonstrating the model’s ability to capture subtle anatomical variations and pathological features in the optic nerve.

Table 5.7. Classification Performance for Glaucoma and Optic Nerve–Related Pathologies on 39-CAS

Class	P	R	$F1$
Possible glaucoma	1.0000	0.9748	0.9872
Optic atrophy	0.9756	1.0000	0.9877
Disc swelling and elevation	1.0000	1.0000	1.0000
Dragged disc	1.0000	1.0000	1.0000
Congenital disc abnormality	0.9917	1.0000	0.9958
Myelinated nerve fiber	1.0000	1.0000	1.0000

Retinal Vascular Diseases. Table 5.8 presents the classification performance for retinal vascular diseases on the 39-CAS dataset. All vascular pathology classes achieved exceptionally high precision, recall, and F1-scores, ranging from 0.9916 to 1.0000, indicating very reliable detection of diverse vascular abnormalities.

The **BRVO** class achieved near-perfect performance ($F1 = 0.9917$), while **CRVO** and **Preretinal hemorrhage** reached perfect scores ($F1 = 1.0000$), demonstrating the model’s

ability to accurately detect both branch and central retinal vein occlusions as well as hemorrhagic events. Similarly, **RAO** showed slightly lower recall (0.9833), suggesting minimal false negatives, whereas vessel tortuosity and severe hypertensive retinopathy were detected with near-perfect F1-scores (0.9959).

Table 5.8. Classification Performance for Retinal Vascular Diseases on 39-CAS

Class	P	R	$F1$
BRVO	0.9917	0.9917	0.9917
CRVO	1.0000	1.0000	1.0000
RAO	1.0000	0.9833	0.9916
Vessel tortuosity	0.9917	1.0000	0.9959
Severe hypertensive retinopathy	0.9917	1.0000	0.9959
Preretinal hemorrhage	1.0000	1.0000	1.0000

Macular and Peripheral Retinal Disorders. Table 5.9 presents the classification performance for macular and retinal disorders on the 39-CAS dataset. Overall, the model achieved high precision, recall, and F1-scores, ranging from 0.9833 to 1.0000, demonstrating reliable detection of a wide spectrum of macular and retinal pathologies.

The **ERM** and **Chorioretinal atrophy / coloboma** classes reached perfect scores ($F1 = 1.0000$), indicating flawless identification. Other classes, including **Maculopathy** ($F1 = 0.9833$), **MH** ($F1 = 0.9916$), **CSCR** ($F1 = 0.9959$), and **Pathological myopia** ($F1 = 0.9958$), also showed near-perfect performance, with only minor false positives or negatives. The slightly lower F1 for **Rhegmatogenous RD** (0.9874) may be due to subtle structural variability in retinal detachment presentations.

Table 5.9. Classification Performance for Macular & Retinal Disorders on 39-CAS

Class	P	R	$F1$
Maculopathy	0.9833	0.9833	0.9833
ERM	1.0000	1.0000	1.0000
MH	1.0000	0.9833	0.9916
CSCR	0.9917	1.0000	0.9959
VKH disease	0.9917	1.0000	0.9959
Pathological myopia	1.0000	0.9917	0.9958
Rhegmatogenous RD	0.9916	0.9833	0.9874
Chorioretinal atrophy / coloboma	1.0000	1.0000	1.0000
Peripheral retinal degeneration and break	0.9917	1.0000	0.9959

Retinal Dystrophies and Deposits. Table 5.10 summarizes the classification performance for retinal dystrophies and deposits on the 39-CAS dataset. Overall, the model achieved very high precision, recall, and F1-scores, ranging from 0.9791 to 1.0000, indicating highly reliable detection of both inherited and acquired retinal abnormalities.

The classes **Retinitis pigmentosa**, **Bietti crystalline dystrophy**, **Massive hard exudates**, and **Cotton-wool spots** all reached perfect F1-scores (1.0000), demonstrating flawless identification. The **Yellow-white spots / flecks** class showed slightly lower F1 (0.9791), likely due to the subtle and scattered nature of these deposits, which can create minor ambiguity in feature boundaries.

Table 5.10. Classification Performance for Retinal Dystrophies & Deposits on 39-CAS

Class	P	R	$F1$
Retinitis pigmentosa	1.0000	1.0000	1.0000
Bietti crystalline dystrophy	1.0000	1.0000	1.0000
Massive hard exudates	1.0000	1.0000	1.0000
Yellow-white spots / flecks	0.9832	0.9750	0.9791
Cotton-wool spots	1.0000	1.0000	1.0000

Other Retinal Conditions. Table 5.11 presents the classification performance for other retinal conditions on the 39-CAS dataset. Overall, the model achieved exceptionally high precision, recall, and F1-scores, ranging from 0.9958 to 1.0000, demonstrating highly reliable detection across a diverse set of less common retinal conditions.

The classes **Vitreous Particles**, **Laser Spots**, and **Silicon Oil in Eye** reached perfect F1-scores (1.0000), indicating flawless identification. The **Fundus Neoplasm** and **Fibrosis** classes showed slightly lower F1-scores (0.9958 and 0.9959, respectively), likely due to minor variability in lesion appearance, yet still reflecting excellent detection performance.

Table 5.11. Classification Performance for Other Retinal Conditions on 39-CAS

Class	P	R	$F1$
Fundus Neoplasm	1.0000	0.9917	0.9958
Vitreous Particles	1.0000	1.0000	1.0000
Fibrosis	0.9917	1.0000	0.9959
Laser Spots	1.0000	1.0000	1.0000
Silicon Oil in Eye	1.0000	1.0000	1.0000

Biometric Identification Performance

Table 5.12 summarizes the biometric identification performance on the 39-CAS dataset. The model achieved an accuracy of 87.11% ($\mathcal{A}_{\text{id}} = 0.8711$), indicating a high success rate in correctly identifying individual subjects. Precision (0.8491) and recall (0.8711) are similarly high, yielding an F1-score of 0.8527, which reflects a balanced performance between correctly identifying true matches and avoiding false positives.

The Genuine Accept Rate (GAR) matches the accuracy at 0.8711, while both the False Reject Rate (FRR) and False Accept Rate (FAR) are 0.1289, indicating low error rates in rejecting genuine identities or accepting impostors.

Table 5.12. Biometric Identification Performance Metrics on 39-CAS

Metric	$\mathcal{A}_{\text{id}}$	P	R	$F1$	GAR	FRR	FAR
Value	0.8711	0.8491	0.8711	0.8527	0.8711	0.1289	0.1289

5.9.3 Results on FIVES Dataset

The FIVES (Fundus Image Vessel Segmentation) dataset was originally developed for retinal vessel segmentation tasks. In this study, it is repurposed to assess the generalization capability of the proposed unified multitask framework. The dataset covers a broad demographic range (ages 4–83) and includes a variety of retinal conditions, specifically Age-related Macular Degeneration (AMD), Diabetic Retinopathy (DR), Glaucoma (G), and Normal (N) fundus images without diagnosed ocular disease. Furthermore, the images encompass diverse quality degradations, including illumination distortions, blur, and contrast variations, making the dataset particularly suitable for evaluating the model’s robustness and stability under realistic clinical conditions.

Pathology Classification Performance

Table 5.13 presents the overall performance of the proposed unified multitask framework on the FIVES dataset for pathology classification. The model achieved a high overall accuracy of 0.9733 ($\mathcal{A}$), demonstrating its ability to correctly classify the majority of retinal images across diverse conditions.

Both micro- and macro-averaged precision ($P_{\text{micro}} = 0.9733$, $P_{\text{macro}} = 0.9759$) and recall ($R_{\text{micro}} = 0.9700$, $R_{\text{macro}} = 0.9683$) are consistently high, indicating that the model performs well both overall and across individual classes, effectively balancing performance for frequent and less-represented pathologies. The corresponding F1-scores ($F1_{\text{micro}} = 0.9700$, $F1_{\text{macro}} = 0.9699$) further confirm a strong trade-off between precision and recall.

Additionally, the low hierarchical loss ($\mathcal{L}_H = 0.0150$) suggests that the model respects the structured relationships between different pathology classes, while the macro-averaged AUC of 0.9982 demonstrates near-perfect discriminative ability across all categories.

Table 5.13. Overall Pathology Classification Performance on FIVES

Metric	$\mathcal{A}$	P_{micro}	P_{macro}	R_{micro}	R_{macro}	$F1_{\text{micro}}$	$F1_{\text{macro}}$	$\mathcal{L}_H$	$\text{AUC}_{\text{macro}}$
Value	0.9733	0.9733	0.9759	0.9700	0.9683	0.9700	0.9699	0.0150	0.9982

Table 5.14 presents the per-class classification performance of the proposed unified multitask framework on the FIVES dataset. The results indicate consistently high performance across all retinal categories.

The Age-related Macular Degeneration (AMD) class achieved precision of 0.9555, recall of 0.9667, and F1-score of 0.9508, reflecting highly accurate detection of AMD cases. Diabetic Retinopathy (DR) was classified with precision 0.9557, recall 0.9733, and F1-score 0.9434, indicating slightly higher sensitivity but a minor reduction in overall balance. Glaucoma (G) showed precision 0.9550, recall 0.9333, and F1-score 0.9441, suggesting that while most glaucoma cases were correctly identified, a small fraction of false negatives occurred. The Normal (N) class achieved precision 0.9579, recall 0.9600, and F1-score 0.9539, confirming reliable recognition of healthy fundus images.

Table 5.14. Per-Class Pathology Classification Report on FIVES

Class	P	R	$F1$
A (AMD)	0.9555	0.9667	0.9508
D (DR)	0.9557	0.9733	0.9434
G (Glaucoma)	0.9550	0.9333	0.9441
N (Normal)	0.9579	0.9600	0.9539

Biometric Identification Performance

Table 5.15 summarizes the biometric identification performance of the proposed framework on the FIVES dataset. The model achieved an accuracy of 92.33% ($\mathcal{A}_{\text{id}} = 0.9233$), demonstrating strong ability to correctly identify individual subjects across diverse demographic and pathological conditions.

Precision (0.9301) and recall (0.9233) are both high, resulting in an F1-score of 0.9171, which indicates a balanced trade-off between correctly recognizing true identities and minimizing false matches. The Genuine Accept Rate (GAR) of 0.9233 further confirms

reliable recognition of genuine subjects, while the False Reject Rate (FRR) and False Accept Rate (FAR) of 0.0767 reflect low misidentification rates.

Table 5.15. Biometric Identification Performance on FIVES

Metric	$\mathcal{A}_{\text{id}}$	P	R	$F1$	GAR	FRR	FAR
Value	0.9233	0.9301	0.9233	0.9171	0.9233	0.0767	0.0767

5.9.4 Comparative Discussion Across Datasets

The proposed unified multitask framework demonstrates robust performance across the ODIR-2019, 39-CAS, and FIVES datasets, despite substantial differences in dataset scale, annotation protocols, and clinical characteristics. Classification performance remains consistently high across all datasets. On ODIR-2019, the framework achieves strong micro- and macro-averaged F1 scores ($F1_{\text{micro}} \approx 0.91$, $F1_{\text{macro}} \approx 0.92$) and a high macro-AUC ($\text{AUC}_{\text{macro}} = 0.993$), confirming reliable multi-label discrimination in complex clinical scenarios with co-existing diseases. On 39-CAS, near-perfect fine-grained classification is observed ($\mathcal{A} = 0.9953$, $\text{AUC}_{\text{macro}} = 1.0000$), highlighting the model’s ability to distinguish closely related retinal disease subtypes. On FIVES, classification performance remains high ($\mathcal{A} = 0.9733$, $F1_{\text{macro}} \approx 0.9699$, $\text{AUC}_{\text{macro}} = 0.9982$), demonstrating strong generalization despite heterogeneous image quality and the dataset’s original focus on vessel segmentation. Minimal differences between micro- and macro-averaged metrics indicate stable per-class performance and effective mitigation of class imbalance.

Identification performance is consistently strong, with expected variability influenced by the number of identity classes and inter-subject similarity. Accuracy is near-perfect on ODIR-2019 ($\mathcal{A}_{\text{id}} = 0.9898$), lower on 39-CAS ($\mathcal{A}_{\text{id}} = 0.8711$) due to an increased number of identities and finer inter-subject similarity, and improved on FIVES ($\mathcal{A}_{\text{id}} = 0.9233$), reflecting dataset characteristics. Observed trends in the False Accept Rate (FAR) and False Reject Rate (FRR) are consistent with these patterns. These results align with typical biometric system behavior, indicating that identity complexity modulates recognition performance while maintaining overall robustness.

The analysis confirms that the shared retinal feature representation supports both pathology classification and identity recognition without performance compromise. The framework effectively captures disease-specific morphological variations alongside individual-specific vascular micro-patterns, enabling dual clinical and biometric functionality.

The experimental findings demonstrate that the proposed framework is robust, generalizable, and clinically applicable, achieving high pathology detection ($\mathcal{A} \geq 0.973$) and strong biometric identification ($\mathcal{A}_{\text{id}} \geq 0.871$) across diverse datasets. These results validate

retinal fundus imaging as a highly discriminative modality capable of supporting integrated diagnostic and biometric applications within a single deep learning architecture.

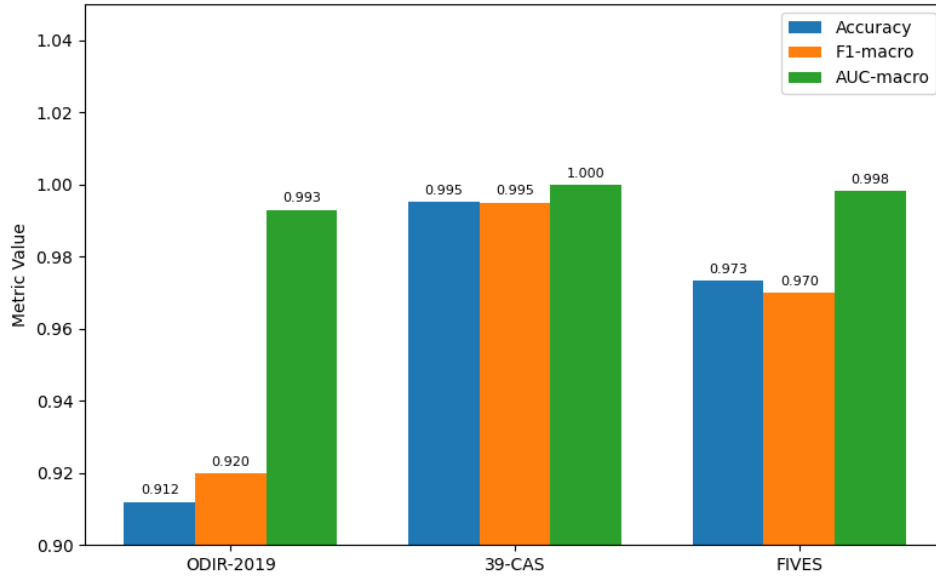


Figure 5.7. Comparative pathology classification performance across ODIR-2019, 39-CAS, and FIVES datasets.

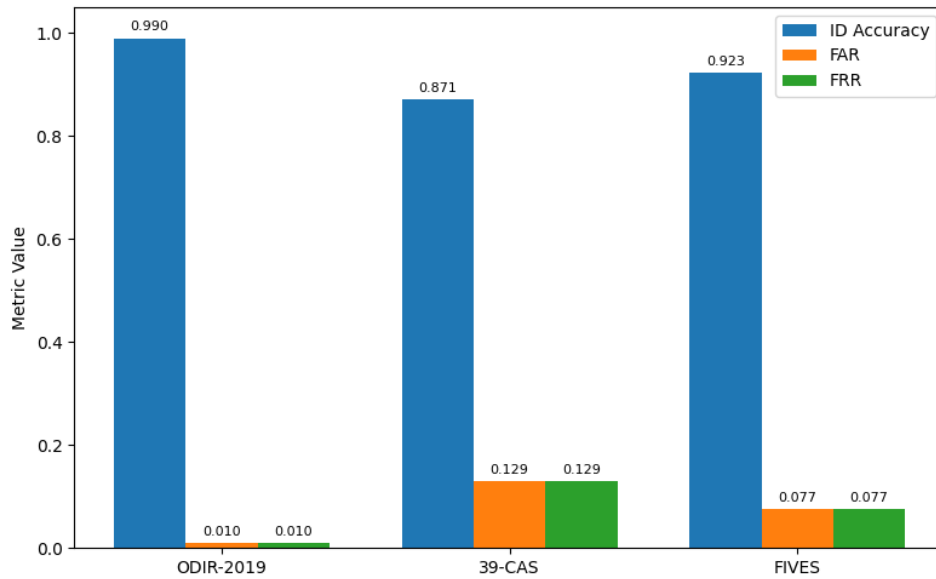


Figure 5.8. Comparative biometric identification performance across ODIR-2019, 39-CAS, and FIVES datasets.

The bar charts illustrate the unified multitask framework’s performance across the three datasets. For pathology classification, ODIR-2019 exhibits slightly lower accuracy and macro F1 due to its multi-label nature, whereas 39-CAS achieves near-perfect fine-grained discrimination. FIVES demonstrates strong generalization, maintaining high metrics despite heterogeneous image quality.

For biometric identification, ODIR-2019 achieves near-perfect accuracy with minimal FAR and FRR. Performance decreases on 39-CAS due to the larger number of identities and finer inter-subject similarity, while FIVES shows intermediate performance, reflecting a balance between dataset consistency and subject diversity.

5.9.5 Comparative Analysis with State-of-the-Art Methods

To evaluate the effectiveness of the proposed unified multitask framework, its performance is compared with several state-of-the-art approaches in retinal pathology classification and retinal biometric identification. These comparisons are based on published results obtained from fundus photography datasets. The objective is to assess how the proposed framework performs relative to existing solutions while highlighting its unique capability of simultaneously addressing diagnostic and biometric tasks within a single architecture.

Pathology Classification Performance

Table 5.16 presents a comparative evaluation of the proposed method against recent state-of-the-art approaches for retinal pathology classification using fundus photography datasets with varying complexity and number of classes. The results highlight clear performance differences across datasets and classification settings.

Cen et al. Cen et al. (2021) achieved 92.62% accuracy on the 39-CAS dataset, which contains 39 classes, demonstrating the difficulty of large-scale multi-class retinal disease classification. In comparison, Alqassab et al. Alqassab et al. (2026) reported a higher accuracy of 94.00% on the OIA-ODIR dataset with 8 classes, indicating improved performance in multi-label retinal disease recognition when leveraging enhanced feature extraction and training strategies. Similarly, Ray et al. Ray et al. (2026) obtained 91.60% accuracy on the FIVES dataset, reflecting competitive but still limited performance in general retinal pathology classification.

In contrast, the proposed method demonstrates strong and consistent performance across all evaluated datasets. It achieves 85.30% accuracy on the OIA-ODIR dataset, which is comparable to existing results considering the multi-label complexity of the task. More importantly, it significantly outperforms prior methods on more challenging settings, reaching 99.53% accuracy on the 39-CAS dataset with 39 classes, and 97.33% accuracy on the FIVES dataset.

These results indicate that the proposed framework is particularly effective in handling large-scale and fine-grained retinal classification problems. The substantial improvement on the 39-class scenario suggests that the model benefits from robust feature representation learning and effective generalization across highly diverse disease categories.

Overall, the comparative analysis demonstrates that while existing methods perform well on moderate-scale datasets, their performance tends to decrease with increasing classification complexity. In contrast, the proposed approach maintains strong accuracy across datasets with different numbers of classes, highlighting its scalability, robustness, and suitability for real-world retinal disease screening applications.

Table 5.16. Comparison of the proposed method with state-of-the-art pathology classification approaches on fundus photography datasets

Method	Dataset	Classes	Accuracy (%)
Cen et al. (2021)	39-cas	39	92.62
Alqassab et al. (2026)	OIA-ODIR	8	94.00
Ray et al. (2026)	FIVES	4	91.6
Our Method	OIA-ODIR	8	85.30
	39-CAS	39	99.53
	FIVES	4	97.33

Biometric Identification Performance

The results presented in Table 5.17 compare the proposed retinal biometric identification framework with recent state-of-the-art methods. Overall, the reported studies demonstrate high identification performance, particularly on relatively small-scale datasets.

For instance, Roy et al. (2020) and Sadeghpour et al. (2023) achieved strong results on the VARIA dataset, with accuracies of 97.90% and 98.60%, respectively. Similarly, Apap et al. (2024) reported perfect identification performance (100%) on both the VARIA and RIDB datasets, while Ochoa et al. (2024) achieved 96.70% accuracy on a private dataset. These results indicate that retinal biometric systems can achieve very high performance when evaluated under limited subject populations.

However, a key limitation of these approaches is the relatively small number of subjects used for evaluation, typically ranging from a few dozen to approximately 140 individuals. In such scenarios, the identification task is less challenging due to reduced inter-subject variability and a smaller search space. Moreover, certain recent works, such as Sadeghpour et al. (2023), focus on template protection and privacy-preserving biometric representations rather than direct identification, which limits direct comparability in terms of recognition performance.

In contrast, the proposed method is evaluated on significantly larger and more diverse datasets, providing a more realistic assessment of scalability. On the OIA-ODIR dataset, the model achieves 98.98% accuracy across approximately 5000 subjects, demonstrating strong identification capability at large scale. On the 39-CAS dataset, which contains 502 subjects,

the method achieves 87.11% accuracy, reflecting the increased difficulty associated with fine-grained discrimination in a larger identity space. Furthermore, on the FIVES dataset with 600 subjects, the proposed approach achieves 92.33% accuracy, confirming its robustness across datasets with different acquisition conditions and population distributions.

Overall, while several existing methods report near-perfect accuracy, these results are primarily obtained under small-scale experimental settings. The proposed framework distinguishes itself by maintaining strong performance under significantly larger subject populations, thereby providing a more realistic evaluation of retinal biometric identification in practical deployment scenarios. These findings highlight the scalability, robustness, and generalization capability of the proposed method for real-world biometric systems.

Table 5.17. Comparison of the proposed method with state-of-the-art retinal biometric identification approaches

Method	Dataset	Subjects	Application	Accuracy (%)
(Aicha and Eddine, 2024)	RIDB	20	Identification	99.998
(Apap et al., 2024)	VARIA	139	Identification	100.00
	RIDB	20		100.00
(Ochoa-Astorga et al., 2024)	Local database	40	Identification	96.70
(Sadeghpour et al., 2023)	VARIA	139	Template Protection	98.60
Our Method	OIA-ODIR	5000	Identification	98.98
	39-CAS	502	Identification	87.11
	FIVES	600	Identification	92.33

5.10 Conclusion

This chapter presented a unified multitask deep learning framework for the joint analysis of retinal fundus images, addressing the conventional separation between pathology classification and biometric identification. By leveraging a shared representation, the proposed approach enables the simultaneous learning of disease-related features and identity-specific vascular patterns within a single architecture.

The framework combines complementary backbone networks, attention-based feature refinement, and a selective aggregation mechanism to effectively capture both global morphological characteristics and fine-grained retinal structures. The joint optimization of task-specific objectives further allows knowledge transfer between pathology classification and biometric identification, improving overall learning efficiency.

Experimental results on the ODIR-2019, 39-CAS, and FIVES datasets demonstrate that

the proposed model achieves high performance across both tasks, while maintaining robustness to variations in dataset characteristics and imaging conditions. The findings confirm that multitask learning can effectively exploit the intrinsic relationship between retinal structure and pathology without compromising task-specific accuracy.

Despite these promising results, certain limitations remain, including the absence of temporal modeling and limited interpretability of the learned representations. These aspects provide opportunities for further refinement of the proposed approach.

Overall, this chapter establishes the feasibility and effectiveness of a unified multitask framework for retinal image analysis, providing a solid foundation for the subsequent developments presented in this thesis.

Chapter 6

General Conclusion and Future Perspectives

6.1 Introduction

This thesis investigated the integration of biometric recognition and pathology analysis within a unified framework for retinal image processing. By exploiting the dual nature of retinal fundus images, which simultaneously encode identity-specific vascular structures and clinically relevant biomarkers, this work proposes a comprehensive methodology that bridges the long-standing gap between biometric authentication and medical diagnosis.

The research contributions are organized into three complementary components addressing key limitations in existing literature: (i) the lack of integrated biometric pipelines, (ii) insufficient hyperparameter optimization strategies, and (iii) the separation between biometric and diagnostic modeling. This chapter synthesizes the main findings, highlights their scientific and practical implications, discusses limitations, and outlines future research directions.

6.2 Synthesis of Core Contributions

6.2.1 Robust Retinal Authentication Framework

The first contribution introduced a complete and robust pipeline for retinal biometric verification, integrating preprocessing, vessel segmentation using Residual Attention U-Net, Gaussian-weighted reconstruction, and multi-modal feature extraction.

The experimental results demonstrated that:

- The proposed segmentation model achieved consistently high Dice scores (above 0.90)

across multiple benchmark datasets, confirming its robustness and generalization capability.

- The combination of morphological, topological, and texture descriptors yielded highly discriminative representations.
- Ensemble classifiers, particularly XGBoost, achieved near-perfect verification performance, with Equal Error Rates approaching zero.
- The framework maintained strong cross-dataset generalization without retraining.

This contribution establishes a reliable and reproducible foundation for retinal biometric systems and demonstrates the importance of tightly coupled pipeline design.

6.2.2 Systematic Hyperparameter Optimization

The second contribution addressed the critical yet often overlooked role of hyperparameter optimization. A comprehensive evaluation of Grid Search, Random Search, Bayesian Optimization, and CMA-ES was conducted.

The findings revealed that:

- CMA-ES consistently provided superior performance and robustness across datasets.
- Optimal configurations were highly dataset-dependent, highlighting the limitations of manual tuning.
- Bayesian optimization showed strong performance but reduced stability in complex scenarios.
- Grid Search was computationally inefficient and less effective in high-dimensional spaces.

This contribution provides a principled optimization framework, emphasizing the necessity of adaptive strategies for achieving reliable performance in medical biometric systems.

6.2.3 Unified Multitask Learning Framework

The third contribution proposed a novel multitask architecture that jointly performs biometric identification and multi-label pathology classification within a unified model.

Key results include:

- The dual-backbone design with attention mechanisms effectively captured complementary features.
- High pathology classification performance was achieved across multiple datasets.

- Biometric identification accuracy remained strong despite joint training.
- Shared representation learning improved generalization through implicit regularization.

This contribution demonstrates that biometric recognition and clinical diagnosis are complementary objectives that can be simultaneously optimized within a single architecture.

6.3 Overall Contribution to the Field

Collectively, the proposed contributions advance the field of retinal image analysis along three key dimensions:

Methodological: The work introduces an integrated pipeline design, a systematic optimization strategy based on evolutionary algorithms, and a unified multitask learning paradigm.

Empirical: Extensive validation across multiple public datasets demonstrates strong generalization and establishes new performance benchmarks.

Practical: The proposed systems show clear potential for clinical deployment, enabling simultaneous patient identification and disease screening while maintaining computational efficiency and addressing privacy considerations.

6.4 Limitations of the Current Work

Despite the promising results, several limitations remain:

- The lack of longitudinal evaluation limits conclusions regarding long-term biometric stability.
- The authentication pipeline is not fully optimized end-to-end, allowing potential error propagation.
- Hyperparameter optimization remains computationally expensive.
- Limited interpretability of deep models may hinder clinical trust and regulatory approval.
- Dataset diversity, particularly in rare pathologies and demographics, remains insufficient.

Addressing these challenges is essential for transitioning from research prototypes to real-world clinical systems.

6.5 Future Research Directions

Future work should focus on the following directions:

- Longitudinal studies to evaluate temporal stability and disease progression effects.
- Fully end-to-end learning frameworks using metric learning and Siamese architectures.
- Development of explainable AI techniques to improve transparency and clinical acceptance.
- Privacy-preserving learning approaches such as federated learning and differential privacy.
- Multimodal integration with OCT and clinical metadata.
- Model optimization for real-time and edge deployment.
- Expansion toward broader clinical applications, including systemic disease prediction.

6.6 Final Remarks

This thesis demonstrates that retinal fundus images can serve as a unified modality for both biometric authentication and medical diagnosis. By integrating segmentation, feature extraction, optimization, and multitask learning into a cohesive framework, this work establishes a foundation for next-generation medical biometric systems.

The results confirm that retinal vascular patterns provide a powerful and reliable representation capable of supporting both identity recognition and clinical decision-making. The proposed methodologies achieve high accuracy, strong generalization, and practical feasibility for deployment.

Ultimately, this research contributes to the broader vision of intelligent healthcare systems that unify security and clinical functionality. Continued interdisciplinary collaboration will be essential to translate these advances into real-world impact, paving the way toward safer, more efficient, and patient-centered medical technologies.

Bibliography

- Manish Agnihotri, Aditya Rathod, Daksh Thapar, Gaurav Jaswal, Kamlesh Tiwari, and Aditya Nigam. Learning domain specific features using convolutional autoencoder: A vein authentication case study using siamese triplet loss network. In *ICPRAM*, pages 778–785, 2019.
- Mokhtari Aicha and Hadj Slimane Zine Eddine. Secure personal identification through fundus image-based transfer learning. In *2024 2nd International Conference on Electrical Engineering and Automatic Control (ICEEAC)*, pages 1–5. IEEE, 2024.
- M Usman Akram, Anam Akber Salam, Sajid Gul Khawaja, Syed GH Naqvi, and Shoab A Khan. Ridb: a dataset of fundus images for retina based person identification. *Data in Brief*, 33:106433, 2020.
- S. Aleem, B. Sheng, P. Li, P. Yang, and D. D. Feng. Fast and accurate retinal identification system: Using retinal blood vasculature landmarks. *IEEE Transactions on Industrial Informatics*, 15(7):4099–4110, 2018. doi: 10.1109/TII.2018.2881343.
- Ahmed I. M. Alqassab, Miguel Á. Luque-Nieto, and Mazin A. Mohammed. Identification of multiple ocular diseases using a hybrid quantum convolutional neural network with fundus images. *Scientific Reports*, 16:6798, 2026. doi: 10.1038/s41598-026-12345-6.
- George J. Annas. Hipaa regulations: A new era of medical-record privacy? *New England Journal of Medicine*, 348(15):1486–1490, 2003. doi: 10.1056/NEJMLim035027.
- Adrian Apap, Amey Bhole, Laura Fernández-Robles, Manuel Castejón-Limas, and George Azzopardi. Explainable multi-layer cosfire filters robust to corruptions and boundary attack with application to retina and palmprint biometrics. *Neural Computing and Applications*, 36(30):19231–19245, 2024.
- Atrin Arya, Sana Ayromlou, Armin Saadat, Purang Abolmaesumi, and Xiaoxiao Li. Federated impression for learning with distributed heterogeneous data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 215–225. Springer, 2024.

- Reza Azad, Amirhossein Kazerouni, Moein Heidari, Ehsan Khodapanah Aghdam, Amirali Molaei, Yiwei Jia, Abin Jose, Rijo Roy, and Dorit Merhof. Advances in medical image analysis with vision transformers: a comprehensive review. *Medical image analysis*, 91: 103000, 2024.
- S Bhanuse et al. Multi-stage generative adversarial network model for segmenting retinal vascular structures in eye disease prediction. *IEEE Transactions on Medical Imaging*, 44(2):345–358, 2025.
- A. Bilal, A. Imran, T. I. Baig, X. Liu, H. Long, A. Alzahrani, and M. Shafiq. Improved support vector machine based on cnn-svd for vision-threatening diabetic retinopathy detection and classification. *PLoS ONE*, 19(1):e0295951, 2024.
- Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. *Classification and regression trees*. Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA, 1984.
- Ling-Ping Cen, Jian Ji, Jing-Wei Lin, Si-Tong Ju, Hong-Jie Lin, Tai-Ping Li, Yun Wang, Jian-Feng Yang, Yu-Fan Liu, Song Tan, et al. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. *Nature Communications*, 12(1):4828, 2021.
- Chunhui Chen, Joon Huang Chuah, Raza Ali, and Yizhou Wang. Retinal vessel segmentation using deep learning: a review. *IEEE Access*, 9:111985–112004, 2021.
- Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016.
- Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*, pages 1597–1607. PMLR, 2020.
- Takwa Chihaoui, Rostom Kachouri, Hejer Jlassi, Mohamed Akil, and Kamel Hamrouni. Retinal identification system based on optical disc ring extraction and new local sift-ruk descriptor. *Advances in Systems, Signals and Devices*, 8:113–126, 2018.
- I. Glenn Cohen and Michelle M. Mello. Hipaa and protecting health information in the 21st century. *JAMA*, 320(3):231–232, 2018. doi: 10.1001/jama.2018.5630.
- Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3): 273–297, 1995.

- Thomas M. Cover and Peter E. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.
- R. Devi, P. Keerthika, P. Suresh, P. Sarangi, M. Sangeetha, C. Sagana, and K. Devendran. Retina biometrics for personal authentication. In *Machine Learning for Biometrics: Concepts, Algorithms and Applications*, pages 87–104. Academic Press, 2020. doi: 10.1016/B978-0-323-85209-8.00005-5.
- 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.
- Andre Esteva, Katherine Chou, Serena Yeung, Ali Madani, Abhijit Ravuri, Jayashree Kalpathy-Cramer, et al. Deep learning-enabled medical computer vision. *NPJ Digital Medicine*, 4(1):5, 2021.
- European Parliament and Council of the European Union. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. Official Journal of the European Union, L 119, 4.5.2016, p. 1–88, 2016. Available at: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.
- EYRINA Inc. High-resolution, automated retinal imaging for biometric and clinical analysis. White paper, EYRINA Inc., 2023.
- Damian JJ Farnell, Fraser N Hatfield, Paul Knox, Michael Reakes, Sarah Spencer, David Parry, and Simon P Harding. Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators. *Journal of the Franklin Institute*, 345(7):748–765, 2008.
- Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1): 119–139, 1997.
- Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001.
- Maria Frucci, Daniel Riccio, Gabriella Sanniti di Baja, and Luca Serino. Using direction and score information for retina based person verification. *Expert Systems with Applications*, 94:1–10, 2018.
- Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine Learning*, 63(1):3–42, 2006.

- Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. In *International Conference on Learning Representations (ICLR)*, 2018.
- N. Gim, M. Blazes, C. I. Sánchez, L. Zalunardo, G. Corradetti, T. Elze, N. Honda, N. K. Waheed, A. M. Cairns, M. V. Canto-Soler, A. Domalpally, M. Durbin, D. Ferrara, J. Hu, P. Nair, S. R. Sadda, T. D. L. Keenan, C. S. Lee, and the RIMR Consortium. Retinal imaging in an era of open science and privacy protection. *Experimental Eye Research*, 252:110341, 2025. doi: 10.1016/j.exer.2025.110341.
- Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, volume 2, pages 1735–1742. IEEE, 2006.
- Huaqing He, Li Lin, Zhiyuan Cai, and Xiaoying Tang. Joined: Prior guided multi-task learning for joint optic disc/cup segmentation and fovea detection. In *International Conference on Medical Imaging with Deep Learning*, pages 477–492. PMLR, 2022.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020.
- Á. S. Hervella, J. Rouco, J. Novo, et al. End-to-end multi-task learning for simultaneous optic disc and cup segmentation and glaucoma classification in eye fundus images. *Applied Soft Computing*, 116:108347, 2022.
- Robert B Hill. Apparatus and method for identifying individuals through their retinal vasculature patterns, 1986.
- Daniel T Hogarty, Joseph P Hogarty, and Alex W Hewitt. Smartphone use in ophthalmology: What is their place in clinical practice? *Survey of Ophthalmology*, 65(2):250–262, 2020.
- AD Hoover, Valentina Kouznetsova, and Michael Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical imaging*, 19(3):203–210, 2000.
- Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.

- Min Hu, Yanfeng Jiang, Zhiwei Luo, Weizhong Lan, and Weiwei Dai. Integrating biometric and multimodal imaging data for early prediction of myopia onset. *Scientific Reports*, 15(1):31416, 2025.
- Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.
- K. Huang, S. Huang, G. Chen, et al. An end-to-end multi-task system of automatic lesion detection and anatomical localization in whole-body bone scintigraphy by deep learning. *Bioinformatics*, 39(1):btac753, 2023.
- O. Iqbal, W. Jadoon, I. A. Khan, M. Afzal, and A. Aldhahri. Deep learning-based interpretable framework for simultaneous segmentation of retinal multi-lesions and multi-categorical disease classification. *International Journal of Computational Intelligence Systems*, 18(1):275, 2025.
- Ehsaneddin Jalilian, Georg Wimmer, Andreas Uhl, and Mahmut Karakaya. Deep learning based off-angle iris recognition. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4048–4052. IEEE, 2022.
- Kai Jin, Xingru Huang, Jie Zhou, Yun Li, Yaxin Yan, Yitian Sun, Xiaobo Zhang, Juan Ye, et al. Fives: A fundus image dataset for artificial intelligence based vessel segmentation. *Scientific Data*, 9(1):475, 2022.
- Sharif Amit Kamran, Khondker Fariha Hossain, Alireza Tavakkoli, Stewart Lee Zuckerbrod, Kenton M Sanders, and Salah A Baker. Rv-gan: Segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network. In *international conference on medical image computing and computer-assisted intervention*, pages 34–44. Springer, 2021.
- Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30, pages 3146–3154, 2017.
- Rajesh Kumar, Jay Kumar, Abdullah Aman Khan, Hub Ali, Cobbinah M Bernard, Riaz Ullah Khan, Shaoning Zeng, et al. Blockchain and homomorphic encryption based privacy-preserving model aggregation for medical images. *Computerized Medical Imaging and Graphics*, 102:102139, 2022.
- Vinod Kumar, Abhidnya Surve, Devesh Kumawat, Brijesh Takkar, Shorya Azad, Rohan Chawla, Daraius Shroff, Atul Arora, Ramandeep Singh, and Pradeep Venkatesh. Ultra-wide field retinal imaging: a wider clinical perspective. *Indian Journal of Ophthalmology*, 69(4):824–835, 2021.

- S Lazzouni and M Messadi. A novel approach of retina-based biometric authentication using multifractal analysis and wavelet fusion. *Biomedical Signal Processing and Control*, 99:106812, 2025.
- Ning Li, Tao Li, Chunyu Hu, Kang Wang, and Hong Kang. A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection. In *International Symposium on Benchmarking, Measuring and Optimization*, pages 177–193. Springer International Publishing, 2020. doi: 10.1007/978-3-030-66534-0_12.
- Xin Li, Wenhui Zhu, Xuanzhao Dong, Hao Wang, Yujian Xiong, Oana Dumitrascu, and Yalin Wang. nnmobilenet++: Towards efficient hybrid networks for retinal image analysis. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 282–292, 2026.
- Y. Ling, Y. Wang, W. Dai, et al. Mtanet: Multi-task attention network for automatic medical image segmentation and classification. *IEEE Transactions on Medical Imaging*, 43(2):674–685, 2023.
- Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.
- Jiazhen Liu, Xirong Li, Qijie Wei, Jie Xu, and Dayong Ding. Semi-supervised keypoint detector and descriptor for retinal image matching. In *European Conference on Computer Vision*, pages 593–609. Springer, 2022.
- Mingyuan Liu, Lu Xu, and Jicong Zhang. Learning large margin sparse embeddings for open set medical diagnosis. In *International conference on medical image computing and computer-assisted intervention*, pages 548–558. Springer, 2023.
- J Luo et al. Retinaguard: Obfuscating retinal age in fundus images for biometric privacy preserving. *IEEE Transactions on Information Forensics and Security*, 21:1123–1138, 2026.
- Z. Luo, C. Liu, D. Xiao, Z. Yu, Y. Wang, and T. Zhu. Retinaguard: Obfuscating retinal age in fundus images for biometric privacy preserving. *arXiv preprint arXiv:2509.06142*, 2025.
- Junyan Lyu, Yiqi Zhang, Yijin Huang, Li Lin, Pujin Cheng, and Xiaoying Tang. Aadg: Automatic augmentation for domain generalization on retinal image segmentation. *IEEE Transactions on Medical Imaging*, 41(12):3699–3711, 2022.

- Sarifuddin Madenda, Diana Tri Susetianingtias, Dea Adlina, Rini Arianty, et al. Retinal biometric identification using convolutional neural network. 45(6):865–872, 2021.
- M. S. A. Malik, Q. Zahra, I. U. Khan, M. Awais, and G. Qiao. An efficient retinal vessels biometric recognition system by using multi-scale local binary pattern descriptor. *Journal of Medical Imaging and Health Informatics*, 10(10):2481–2489, 2020. doi: 10.1166/jmihi.2020.3180.
- J. B. Mazumdar and S. R. Nirmala. Biometric authentication system based on texture features of retinal images. *International Journal of Biometrics*, 10(3):195–213, 2018. doi: 10.1504/IJBM.2018.093624.
- Tom M. Mitchell. *Machine Learning*. McGraw-Hill, New York, NY, USA, 1997.
- Nankai University Intelligent Computing System Lab. Oia-odir: Ophthalmic image analysis - odir dataset. <https://github.com/nkicsl/OIA>, 2019.
- V Nardone et al. A post-quantum cryptography enabled feature-level fusion framework for privacy-preserving multimodal biometric recognition. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 7(1):78–93, 2025.
- Jesús Eduardo Ochoa-Astorga, Linni Wang, Weiwei Du, and Yahui Peng. Enhanced vascular bifurcations mapping: Refining fundus image registration. *Electronics*, 13(9): 1736, 2024.
- Marcos Ortega, Manuel G Penedo, Jose Rouco, Noelia Barreira, and Maria J Carreira. Retinal verification using a feature points-based biometric pattern. *EURASIP Journal on Advances in Signal Processing*, 2009(1):235746, 2009.
- P. Patil and S. Jagtap. Multi-modal biometric system using finger knuckle image and retina image with template security using polyu and drive database. *International Journal of Information Technology*, 12:1043–1050, 2020. doi: 10.1007/s41870-020-00501-0.
- S Poornima and S Subramanian. Experimental analysis of biometric system using various multimodal fusion algorithms. In *Journal of Physics: Conference Series*, volume 2318, page 012037. IOP Publishing, 2022.
- J. Ross Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.
- B. M. S. Rani and A. Jhansi Rani. A hybrid biometric identification and authentication system with retinal verification using awn classifier for enhancing security. In *First International Conference on Artificial Intelligence and Cognitive Computing: AICC 2018*, pages 561–569. Springer Singapore, 2019. doi: 10.1007/978-981-13-1580-0_54.

- Rahul Ray, S Gopal Krishna Patro, Sudarson Jena, Atul Vikas Lakra, Mohammad Khishe, Mekhrbonu Rakhimova, and Bakhodir Rakhimov. Leveraging retinal vessel segmentation for improved disease classification. *Intelligence-Based Medicine*, page 100391, 2026.
- C. Rieck, C. Mai, L. Eisentraut, and R. Buettner. A novel transformer–cnn hybrid deep learning architecture for robust broad-coverage diagnosis of eye diseases on color fundus images. *IEEE Access*, 2025.
- Nicola Rieke, Jonny Hancox, Wenqi Li, Fausto Milletari, Holger R Roth, Shadi Albarqouni, Spyridon Bakas, Mathieu N Galtier, Bennett A Landman, Klaus Maier-Hein, et al. The future of digital health with federated learning. *NPJ Digital Medicine*, 3(1):119, 2020.
- R Rodiah et al. Retinal biometric identification using convolutional neural network. *Journal of Physics: Conference Series*, 1918(4):042142, 2021.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer, 2015.
- N. D. Roy and A. Biswas. Fast and robust retinal biometric key generation using deep neural nets. *Multimedia Tools and Applications*, 79(9):6823–6843, 2020.
- N. D. Roy, S. Goswami, S. Goswami, and A. Biswas. Biometric template generation framework using retinal vascular structure. In J. K. Mandal, S. Mukhopadhyay, P. Dutta, and K. Dasgupta, editors, *Computational Intelligence, Communications, and Business Analytics*, volume 1031 of *Communications in Computer and Information Science*, pages 245–256. Springer Singapore, 2019. doi: 10.1007/978-981-13-8581-0_20.
- Masoud Sabaghi, S Reza Hadianamrei, Mehdi Fattahi, Mohammad Reza Kouchaki, and Ali Zahedi. Retinal identification system based on the combination of fourier and wavelet transform. *Journal of Signal and Information Processing*, 3(1):35, 2012.
- M. Sadeghpour, A. Arakala, A. Stephen, and K. Horadam. Protection of sparse retinal templates using cohort-based dissimilarity vectors. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 5(2):233–243, 2023. doi: 10.1109/TBIOM.2023.3239866.
- Mathew Salvaris, Daniel Dean, and Wee Hyong Tok. *Deep learning with medical applications*. Apress, Berkeley, CA, 2018.
- TS Sasikala. Multimodal secure biometrics using attention efficient-net hash compression framework. *Digital Signal Processing*, 160:105018, 2025.

- Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823, 2015.
- Mahsa Shabani and Pascal Borry. The impact of the gdpr on health research. *European Journal of Human Genetics*, 28(3):285–286, 2020. doi: 10.1038/s41431-019-0563-6.
- C Simon and I Goldstein. A new scientific method of identification. *New York State Journal of Medicine*, 35(18):901–906, 1935.
- Laxman Singh, Ashish Kumar, and Richa Golash. A review on advancing biometric authentication through integrating multimodal fusion with synthetic data augmentation for adaptive systems. *The European Physical Journal Plus*, 141(1):36, 2026.
- J Son et al. Retinal vessel segmentation in fundoscopic images with generative adversarial networks. *arXiv preprint arXiv:1706.09318*, 2017.
- Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509, 2004.
- S. Sultan and M. F. Ghanim. Human retina based identification system using gabor filters and gda technique. *Journal of Communications Software and Systems*, 16(3):243–253, 2020.
- M. Sushith, N. Malligeswari, M. Anlin Sahaya Infant Tinu, and M. Jaiganesh. Hypergraph-based capsule temporal memory network for efficient and explainable diabetic retinopathy detection in retinal imaging. *Scientific Reports*, 2025.
- Daksh Thapar, Gaurav Jaswal, Aditya Nigam, and Vivek Kanhangad. Pvsnet: Palm vein authentication siamese network trained using triplet loss and adaptive hard mining by learning enforced domain specific features. In *2019 IEEE 5th international conference on identity, security, and behavior analysis (ISBA)*, pages 1–8. IEEE, 2019.
- George T Timberlake and Michael Kennedy. *The direct ophthalmoscope how it works and how to use it*. University of Kansas, 2nd edition, 2005.
- Andreas Uhl, Christoph Busch, Sébastien Marcel, and Raymond NJ Veldhuis. *Handbook of vascular biometrics*. Springer, 2020.
- United States Congress. Health insurance portability and accountability act of 1996. Public Law 104-191, 104th Congress, 1996. Available at: <https://www.congress.gov/104/plaws/publ191/PLAW-104publ191.pdf>.

- Paul Voigt and Axel Von dem Bussche. The eu general data protection regulation (gdpr): A practical guide. *Springer International Publishing*, 2017. doi: 10.1007/978-3-319-57959-7.
- Melvin Womack. The eyes have it. *Sensor Review*, 14(4):15–16, 1994.
- Y. Yan, L. Yang, and W. Huang. Fundus-danet: Dilated convolution and fusion attention mechanism for multilabel retinal fundus image classification. *Applied Sciences*, 14(18): 8446, 2024.
- Jing Yang, Yuxi Ma, Jin-Gang Yu, Feng Gao, Shuting Yang, Du Cai, Jiacheng Wang, and Liansheng Wang. Bridging knowledge discrepancy in retinal image analysis through federated multi-task learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 35–44. Springer, 2025.
- Zhen Yu, Ruiye Chen, Peng Gui, Wei Wang, Imran Razzak, Hamid Alinejad-Rokny, Xiaomin Zeng, Xianwen Shang, Lei Zhang, Xiaohong Yang, et al. A cross population study of retinal aging biomarkers with longitudinal pre-training and label distribution learning. *npj Digital Medicine*, 8(1):344, 2025.
- Luiz A Zanlorensi, Diego Rafael Lucio, Alceu de Souza Britto Junior, Hugo Proença, and David Menotti. Deep representations for cross-spectral ocular biometrics. *IET Biometrics*, 9(2):68–77, 2020.
- Jiacheng Zhang, Rong Jin, and Wenqiang Liu. Enhancing semi-supervised contrastive learning through saliency map for diabetic retinopathy grading. *IET Computer Vision*, 18(8):1127–1137, 2024.
- Y Zhang and Q Yang. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(12):5586–5609, 2021.
- Yan Zhao, Xiuying Wang, Tongtong Che, Guoqing Bao, and Shuyu Li. Multi-task deep learning for medical image computing and analysis: A review. *Computers in Biology and Medicine*, 153:106496, 2023.