https://doi.org/10.15388/vu.thesis.871
https://orcid.org/0000-0002-8758-8294

VILNIUS UNIVERSITY

Sandra Virbukaitė

Methodology for Applying Deep Learning Algorithms for Glaucoma Identification

DOCTORAL DISSERTATION

Technological Sciences,
Informatics Engineering (T 007)

VILNIUS 2025

The dissertation was prepared between 2020 and 2024 at Vilnius University.

Academic Supervisor – Dr. Jolita Bernatavičienė (Vilnius University, Technological Sciences, Informatics Engineering – T 007).

This Doctoral Dissertation will be Defended at a Public Meeting of the Dissertation Defence Panel:

Chairman – Prof. Dr. Julius Žilinskas (Vilnius University, Technological Sciences, Informatics Engineering – T 007).
Members:
Prof. Dr. Olga Kurasova (Vilnius University, Technological Sciences, Informatics Engineering – T 007),
Assoc. Prof. Dr. Viktor Medvedev (Vilnius University, Technological Sciences, Informatics Engineering – T 007),
Prof. Dr. Audris Mockus (University of Tennessee, USA, Technological Sciences, Informatics Engineering – T 007),
Prof. Dr. Pavel Stefanovič (Vilnius Gediminas Technical University, Technological Sciences, Informatics Engineering – T 007).

The dissertation will be defended at a public meeting of the Dissertation Defence Panel at 12:00 p.m. on 15th of December 2025, in room 203 at Vilnius University Faculty of Mathematics and Informatics Institute of Data Science and Digital Technologies.
Address: Akademijos g. 4, LT-04812, Vilnius, Lithuania
Tel. +370 5 210 9300; e-mail: info@mii.vu.lt

The text of this dissertation can be accessed at the Library of Vilnius University, as well on the website of Vilnius University: *https://www.vu.lt/lt/naujienos/ivykiu-kalendorius*.

https://doi.org/10.15388/vu.thesis.871
https://orcid.org/0000-0002-8758-8294

VILNIAUS UNIVERSITETAS

Sandra Virbukaitė

Giliojo mokymosi algoritmų taikymo metodika glaukomai identifikuoti

DAKTARO DISERTACIJA

Technologijos mokslai,
Informatikos inžinerija (T 007)

VILNIUS 2025

Disertacija rengta 2020–2024 metais Vilniaus universitete.

Mokslinė vadovė:
dr. Jolita Bernatavičienė (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija – T 007).

Gynimo taryba:
Pirmininkas – prof. dr. Julius Žilinskas (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija – T 007).
Nariai:
prof. dr. Olga Kurasova (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija – T 007),
doc. dr. Viktor Medvedev (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija – T 007),
prof. dr. Audris Mockus (Tenesio universitetas, JAV, technologijos mokslai, informatikos inžinerija – T 007),
prof. dr. Pavel Stefanovič (Vilniaus Gedimino technikos universitetas, technologijos mokslai, informatikos inžinerija – T 007).

Disertacija ginama viešame Gynimo tarybos posėdyje 2025 m. gruodžio 15 d. 12:00 val. Vilniaus universiteto Matematikos ir informatikos fakulteto Duomenų mokslo ir skaitmeninių technologijų instituto 203 auditorijoje. Adresas: Akademijos g. 4, LT-04812, Vilnius, Lietuva, tel. +370 5 210 9300; el. paštas: info@mii.vu.lt.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir Vilniaus universiteto interneto svetainėje adresu: *https://www.vu.lt/lt/naujienos/ivykiu-kalendorius*.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Dr. Jolita Bernatavičienė for her valuable guidance, support, and insightful advice throughout the course of this thesis. I am also grateful to the reviewers, Prof. Dr. Olga Kurasova and Assoc. Prof. Dr. Viktor Medvedev, for their thoughtful comments and constructive suggestions that helped improve the quality of this work.

I would like to extend my appreciation to Prof. Dr. Jurgita Markevičiūtė for her collaboration in the preparation of the research paper, as well as to all the professors at the Institute of Data Science and Digital Technologies for their high-quality lectures, valuable insights, and continuous encouragement, which have greatly contributed to the development of my knowledge and research skills.

I would also like to sincerely thank Ms. Danutė Rimeisienė, the Doctoral Studies Administrator, for her kind assistance, organizational support, and helpful guidance throughout my doctoral studies.

ABSTRACT

Glaucoma is a leading cause of irreversible blindness, and its diagnosis depends on accurate assessment of the cup-to-disc ratio (CDR) from retinal fundus images. This research develops a multi-level methodology for applying deep learning (DL) algorithms to automate glaucoma detection, while addressing key challenges related to expert annotation variability, dataset inconsistencies, and the lack of standardised diagnostic thresholds. A comprehensive analysis of convolutional neural networks (CNNs) for optic disc (OD) and optic cup (OC) segmentation was conducted using multi-expert annotated datasets. Experimental results demonstrate that image resolution impacts segmentation accuracy, with 512×512 bicubic interpolation providing the most precise results. An ensemble CNN approach enhanced segmentation performance, improving the Dice coefficient by up to 4% for OD and 5% for OC, with majority voting yielding the most consistent results. Statistical analysis confirmed that expert subjectivity introduces significant variability in glaucoma labelling. To address threshold inconsistencies, standardised CDR thresholds were proposed based on statistical analysis across datasets, introducing a three-stage classification: normal, glaucoma suspect, and glaucoma. Among CDR metrics, such as vertical (VCDR), horizontal (HCDR), and area-based (ACDR), VCDR emerged as the most consistent indicator for glaucoma stage classification. These findings highlight the importance of standardising expert labelling and selecting appropriate metrics, demonstrating that addressing annotation variability and establishing well-defined glaucoma diagnostic thresholds can lead to more accurate DL models for automated glaucoma diagnosis, thereby improving early detection and clinical decision-making.

AAM	Augmented attention module
ACDR	Ratio of the cup area to the disc area
ANOVA	Analysis of variance
CDR	Cup-to-disc ratio
CNN	Convolutional neural network
DSC	Depth-wise separable convolution
FCN	Fully convolutional network
FOV	Field of view
HCDR	Ratio of the horizontal cup diameter to the horizontal disc diameter
IoU	Intersection over union
MAE	Mean absolute error
MSA	Multiscale weight-shared attention module
OC	Optic cup
OCT	Optical coherence tomography
OD	Optic disc
ReLU	Rectified linear unit activation
ROI	Region of interest
VCDR	Ratio of the vertical cup diameter to the vertical disc diameter
VDD	Vertical disc diameter
VCD	Vertical cup diameter

α	attention coefficient
CE	cross-entropy loss
$dice$	dice coefficient
g	gating signal
W_x,W_g	linear transformation matrices
b_g, θ	bias terms
σ_1, σ_2	ReLU and sigmoid activation functions
L	dice loss
N	batch size
U	upsampling
$\hat{Y}_\mathrm{b}$	predicted probabilities
Y_b	ground truth

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Glaucoma is a leading cause of irreversible blindness worldwide, and its diagnosis often relies on structural analysis of retinal fundus images [59]. A key indicator is a cup-to-disc ratio (CDR), which quantifies the relationship between the optic disc (OD) and the optic cup (OC) [10, 71]. Several CDR metrics are commonly used, including the area (ACDR) [12], vertical (VCDR) [10], and horizontal (HCDR) [44] ratios. Accurate estimation of these metrics requires precise segmentation of OD and OC. This task is traditionally performed manually by ophthalmologists but is increasingly being automated through deep learning (DL)-based computer vision techniques such as convolutional neural networks (CNNs).

Although automated glaucoma identification systems offer advantages in speed and consistency, DL-based approaches face several challenges. First, training data annotations are influenced by the annotator's expertise and subjective interpretation, which can result in inconsistent OD and OC boundaries. This variability introduces noise into the ground truth, directly affecting the performance of DL-based models. Second, variations in the definition and application of CDR thresholds, that is the specific values of the CDR used to separate glaucomatous from healthy cases, introduce inconsistencies that limit the comparability of automated glaucoma detection systems.

Therefore, this dissertation investigates these challenges by evaluating the impact of multi-expert annotations with varying expertise on DL-based OD and OC segmentation. Furthermore, it analyses variability in the definition of CDR thresholds across studies and proposes methods for standardising CDR measurement.

Research Problem

Automating glaucoma diagnosis using DL-based models has the potential to enhance the efficiency and accuracy of eye health assessment. Such models can analyse fundus images and segment key retinal structures, such as OD and OC, enabling precise measurement of CDR. Among DL approaches, CNNs are particularly prominent due to their ability to learn hierarchical features from retinal images automatically. However, despite their potential, several challenges arise in the development of automated glaucoma detection systems.

The primary research problem relates to the diversity of datasets, which is a crucial factor in developing DL-based models. Many current DL-based automated glaucoma diagnosis tools rely on the same dataset for both training and evaluation. This approach results in models that are well-adapted to specific image types but may struggle to perform accurately on images from different populations or imaging devices.

A second research challenge concerns image preprocessing. Fundus images captured using different cameras vary in resolution, and their large size may not fit CNN input dimensions, requiring resizing before being processed. The region of interest (ROI) extraction and resizing processes can affect image quality, potentially impacting the accuracy of OD and OC segmentation, which is critical for accurate CDR measurement.

Another key issue is annotation variability in fundus image datasets. CNNs and other DL architectures rely on annotated datasets for training. Still, the segmentation of OD and OC can vary significantly among experts due to differences in experience and interpretation. These inconsistencies in labelling introduce noise into the ground truth data, affecting CNN performance.

Furthermore, the CDR metric and threshold variability create additional challenges. Different studies and clinical guidelines use varying CDR measurement approaches and threshold values to determine glaucoma presence, leading to inconsistencies in diagnosis. Without a standardised CDR threshold, automated glaucoma detection systems may produce inconsistent results, as outcomes can differ depending on the dataset or clinical guidelines used, making it challenging to define uniform diagnostic standards.

This thesis addresses these challenges by evaluating the impact of dataset diversity, image preprocessing techniques, expert annotation variability, and inconsistencies in CDR metrics and threshold definitions. By proposing a multi-level methodology, this research aims to improve the development and comparability of DL-based glaucoma detection systems, ultimately contributing to more effective glaucoma screening.

Actuality

Ophthalmology diseases such as glaucoma, diabetic retinopathy, macular degeneration, and retinal detachment are major contributors to vision

impairment and blindness globally. Early diagnosis and timely monitoring are critical to preserving vision and preventing irreversible damage [11]. Among these diseases, glaucoma poses a significant challenge due to its asymptomatic progression until advanced stages, making it the second leading cause of blindness worldwide. Glaucoma impacts approximately 3–5% of the global population aged 40 to 80. In 2013, it was estimated that 64.3 million people within this age group were living with glaucoma worldwide. This number was expected to increase to 76 million by 2020 and is projected to reach 111.8 million by 2040 [63]. The prevalence of glaucoma in Lithuania has increased over the past decade, rising from 37.7 cases per 1,000 individuals in 2015 to 40.7 cases per 1,000 in 2019 and to 41.5 cases per 1,000 in 2023 [20].

The CDR, a key metric derived from fundus images, is pivotal in glaucoma diagnosis. Ophthalmologists can carry out CDR measurement manually, relying on their clinical expertise, or with the assistance of automated computer-based methods such as CNNs. Automated systems hold the advantage of enhanced speed and consistency in OD and OC segmentation needed for CDR calculation. When deploying automated segmentation algorithms that use the CDR as a threshold to differentiate glaucoma cases from normal ones, notable variability emerges in how this threshold is defined and applied across different studies. The development of automated systems for glaucoma diagnosis is commonly based on annotated eye fundus image datasets. Publicly accessible eye fundus image datasets include images that are described and annotated by multiple experts with varying levels of expertise and diagnostic perspectives. However, there has been no evaluation of the impact of annotations by different experts on CNN training. Also, the lack of consensus on the standardised CDR thresholds and metrics is notable, which may impact the comparability of results across different studies.

Research Object

The focus of this study is on the following research objects:

- The key retinal structures, including OD and OC, to detect pathological changes associated with various eye conditions.
- DL methods for automatic detection of glaucoma.

- Quantitative metrics such as the CDR for glaucoma identification.

Research Aim and Objectives

This research aims to investigate and systematically evaluate the primary factors influencing the accuracy of DL models in glaucoma diagnosis and, based on the results obtained, to propose and validate a multi-level methodology that contributes to more accurate and efficient computer-aided glaucoma diagnosis from fundus images. The objectives include the following:

- To conduct an analytical review of DL methods applied to eye fundus image segmentation, identifying the primary factors influencing segmentation outcomes.
- To evaluate the effectiveness of image preprocessing methods, such as resizing and cropping, and to select appropriate methods for the proposed methodology.
- To develop DL methods for glaucoma identification, addressing the generalisation problem for eye fundus images of varying quality.
- To establish thresholds for glaucoma risk group (stage) identification through statistical analysis, incorporating insights from multiple experts.
- To develop a multi-level methodology that considers the factors influencing segmentation outcomes and contributes to more accurate and reliable computer-aided glaucoma diagnosis.

Research Methods

This thesis incorporates a comprehensive literature review covering DL models for glaucoma diagnosis, commonly used eye fundus image datasets, data preprocessing techniques, and the diagnostic metrics employed in the field. The following research methods are applied in this study:

- A comprehensive literature review of the different types of eye fundus cameras used for capturing eye fundus images, along with

an examination of various image preprocessing techniques and DL architectures for image segmentation.

- The development or modification of DL models for glaucoma identification.
- Image preprocessing techniques, including ROI extraction, resizing with various interpolation techniques, normalisation, and augmentation, to ensure the images are optimally prepared for DL model training.
- Experimental approaches and evaluation on publicly available datasets, such as Chákṣu IMAGE (further Chákṣu), DRISHTI-GS, REFUGE, and RIM-ONE r3 (further RIM-ONE), applying quantitative indicators such as the Dice coefficient (Dice) and the Intersection over Union (IoU), sometimes called the Jaccard coefficient, and statistical methods such as Analysis of Variance (ANOVA), Levene's test, and Student's t-tests.

Scientific Novelty

This thesis proposes a multi-level methodology for eye fundus image preprocessing, OD and OC segmentation, and the identification of glaucoma in fundus images. These enhance the accuracy and standardisation of automated glaucoma diagnostic systems by unifying the preprocessing of eye fundus images and their annotations, as well as the evaluation of metrics, providing more consistent solutions for early glaucoma detection and improved clinical decision-making. The scientific novelty can be outlined in a few key points:

- Demonstrated that image preprocessing techniques, specifically resizing and interpolation methods, have a significant impact on the accuracy of segmenting critical anatomical structures for glaucoma diagnosis, namely OD and OC.
- Introduced an ensemble of DL models for OD and OC segmentation, achieving a marked improvement in Dice compared to single-model approaches. Notably, OD segmentation by Dice improved by an average of 4%, and OC segmentation by Dice improved by an average of 5% across various datasets.

- Proposed thresholds for ACDR, VCDR, and HCDR to classify eye health into three categories: normal, glaucoma suspect, and glaucoma. The proposed "glaucoma suspect" stage enhances early glaucoma detection capabilities.
- Highlighted the variability in expert labeling for OD and OC segmentation and its impact on DL methods, particularly CNN performance. Statistical analysis demonstrated significant differences in CDR-based glaucoma detection across experts, underscoring the subjective nature of manual assessments.
- Identified VCDR as the most consistent CDR metric for glaucoma detection, compared to ACDR and HCDR.

Practical Value of the Research

This research introduces a novel multi-level DL-based methodology for glaucoma diagnostics, with a particular focus on early-stage detection. By employing DL algorithms, the proposed approach, including the CNN ensemble method, enables automated and objective assessment of glaucoma stages, reducing dependence on subjective expert evaluations.

The integration of specific CDR thresholds facilitates systematic classification of eye health into normal, glaucoma suspect, and glaucoma stages. This systematic categorization enhances early detection and enables timely intervention.

A comparative analysis of OD and OC segmentation results revealed no statistically significant differences between the DL-based segmentation and expert annotations, indicating that CNN models can be trusted as a tool for automated glaucoma diagnostics, making them viable for real-world clinical applications.

Statements to be Defended

1. Different image resolutions and interpolation methods affect the segmentation results. The best results were obtained when resizing images to 512×512 px using bicubic interpolation.
2. Among the three evaluated metrics, ACDR, HCDR, and VCDR, the VCDR showed the most consistent results in distinguishing glau-

coma stages, suggesting it is the most suitable metric for glaucoma assessment.

3. Variability in expert annotations of fundus images statistically significantly affects the segmentation performance of CNNs.

4. The proposed ensemble of CNNs demonstrates improvements in segmentation accuracy across several datasets. Specifically, Dice for OD segmentation increased by an average of 4%, while OC segmentation accuracy improved by an average of 5% across various datasets.

Approbation and Publications of the Research

The results of this thesis were published in 2 periodic scientific journals indexed by Clarivate Web of Science (Clarivate WoS), 1 peer-reviewed scientific journal, and 2 peer-reviewed scientific conference proceedings. The results were presented at 2 international and 5 national scientific conferences.

Articles published in international research journals with a citation index in the Clarivate WoS database:

1. Sandra Virbukaitė, Jolita Bernatavičienė, Daiva Imbrasienė. Glaucoma Identification Using Convolutional Neural Networks Ensemble for Optic Disc and Cup Segmentation, *IEEE Access*. IEEE. eISSN 2169-3536. 2024, Vol. 12, p. 82720–82729. DOI: 10.1109/ACCESS.2024.3412185.

2. Sandra Virbukaitė, Jolita Bernatavičienė. Impact of Eye Fundus Image Preprocessing on Key Objects Segmentation for Glaucoma Identification, *Nonlinear Analysis: Modelling and Control*. Vilnius University Press. ISSN 1392-5113. eISSN 2335-8963. 2024, Vol. 29, p. 96–110. DOI: 10.15388/namc.2024.29.33669.

Papers in other peer-reviewed scientific periodic journals:

1. Sandra Virbukaitė, Jolita Bernatavičienė. Deep Learning Methods for Glaucoma Identification Using Digital Fundus Images, *Baltic Journal of Modern Computing*. University of Latvia. ISSN 2255-8942. eISSN 2255-8950. 2020, Vol. 8, p. 520–530. DOI: 10.22364/bjmc.2020.8.4.03.

Papers in peer-reviewed scientific conference proceedings:

1. Sandra Virbukaitė, Jolita Bernatavičienė. Deep Neural Networks application for Cup-to-Disc Ratio Estimation in Eye Fundus Images, *2023 18th Conference on Computer Science and Intelligence Systems (FedCSIS), Warsaw, Poland*. IEEE, 2023. ISSN 2300-5963. Vol. 35, p. 1191–1195. DOI: 10.15439/2023F944.

2. Sandra Virbukaitė, Jolita Bernatavičienė. Image Resizing Impact on Optic Disc and Optic Cup Segmentation, *WSCG 2022: full papers proceedings: 30. International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*. University of West Bohemia, 2022. p. 306–309. DOI: 10.24132/CSRN.3201.39.

Presentations in international scientific conferences:

1. Sandra Virbukaitė. Deep Neural Networks application for Cup-to-Disc Ratio Estimation in Eye Fundus Images, *2023 18th Conference on Computer Science and Intelligence Systems (FedCSIS), Warsaw, Poland*. Warsaw, Poland, September 17–20, 2023.

2. Sandra Virbukaitė. Image Resizing Impact on Optic Disc and Optic Cup Segmentation, *30th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*. Pilzen, Czech Republic, May 17-20, 2022.

Presentations in national scientific conferences:

1. Sandra Virbukaitė. How the Expertise of Different Experts Influences the Learning of Convolutional Neural Networks, *Artificial intelligence technologies in medicine: research and diagnostics*. Vilnius, Lithuania, October 23, 2024.

2. Sandra Virbukaitė. Application of Convolutional Neural Networks for Glaucoma Identification, *Artificial intelligence technologies in medicine: research and diagnostics*. Vilnius, Lithuania, October 20, 2023.

3. Sandra Virbukaitė. Interpolation Methods Impact on Eye Fundus Optic Disc and Optic Cup Segmentation, *Data Analysis Methods for Software Systems. 13th International Workshop*. Druskininkai, Lithuania, December 1-3, 2022.

4. Sandra Virbukaitė. Impact of Images Quality Variety and Resizing Level on Eye Fundus Optic Disc Segmentation, *Data Analysis Methods for Software Systems. 12th International Workshop*. Druskininkai, Lithuania, December 2-4, 2021.

5. Sandra Virbukaitė. Image Quality Impact on Optic Disc Segmentation Accuracy, *Computer Days*. Klaipėda, Lithuania, September 23-24, 2021.

Outline of the Thesis

The thesis consists of an Introduction, three Chapters, a Conclusion, and a Summary in the Lithuanian language. The Introduction provides an overview of the research and the dissertation. Chapter 1 reviews the literature, Chapter 2 describes DL methods, evaluation metrics, and the proposed methodology for image segmentation in glaucoma identification, while Chapter 3 presents the DL-based experiments and the obtained results for glaucoma identification.

The bibliographic references are included at the end of the thesis. The dissertation consists of 132 pages, 12 figures, and 28 tables.

1. LITERATURE REVIEW

This chapter provides an overview of the key research areas relevant to automated glaucoma detection from eye fundus images. It covers eye fundus imaging and datasets, DL methods for image segmentation, CNN ensemble approaches for glaucoma identification, commonly used measures and thresholds, and related studies conducted in Lithuania.

1.1. Eye Fundus Imaging and Datasets

Eye fundus imaging refers to the process of capturing images of the interior surface of the eye using various types of fundus cameras. These images contain key anatomical structures, including the macula, fovea, neuroretinal rim, retinal blood vessels, and the OD, with the OC located at its center. These structures are visualised in Figure 1.1. OD, also called the optic nerve head, is a round disc-shaped yellow-orange or pink area at the back of the eye where the retina and optic nerve connect. The OD is responsible for the transmission of information, in the form of electrical impulses, from the retina to the brain via the optic nerve. The appearance and health of the OD are crucial in assessing eye diseases, including glaucoma. OC is a small indentation called the physiological cup located in the centre of the OD. The OC size varies among individuals and is assessed relative to the OD size. Monitoring the size of the OC is vitally important in evaluating eye condition, as enlargements or other changes in the OC's size can be an early indicator of glaucoma. The Macula is a distinct area in the central part of the retina, crucial for the central vision that enables us to perceive fine details with clarity. It is essential for tasks like reading, driving, and recognising faces. Additionally, the macula is responsible for colour vision, allowing us to experience the full range of colours in vivid detail. The fovea is a small central depression inside the macula that contains the highest concentration of cone photoreceptor cells and is responsible for sharp central vision. It plays a vital role in providing sharp central vision, enabling precise visual tasks like reading fine text or threading a needle. Blood vessels are the central retinal artery and its branches that supply the retina by providing it with nutrients and oxygen, and removing metabolic waste and carbon dioxide. Changes in the structure or function of these vessels can indicate or contribute to systemic and ocular diseases, such

as hypertension or diabetic retinopathy. The neuroretinal rim refers to the region of tissue located between the area of the OD and the OC. The thinning of this rim is a typical structural change commonly associated with glaucomatous damage to the OD.

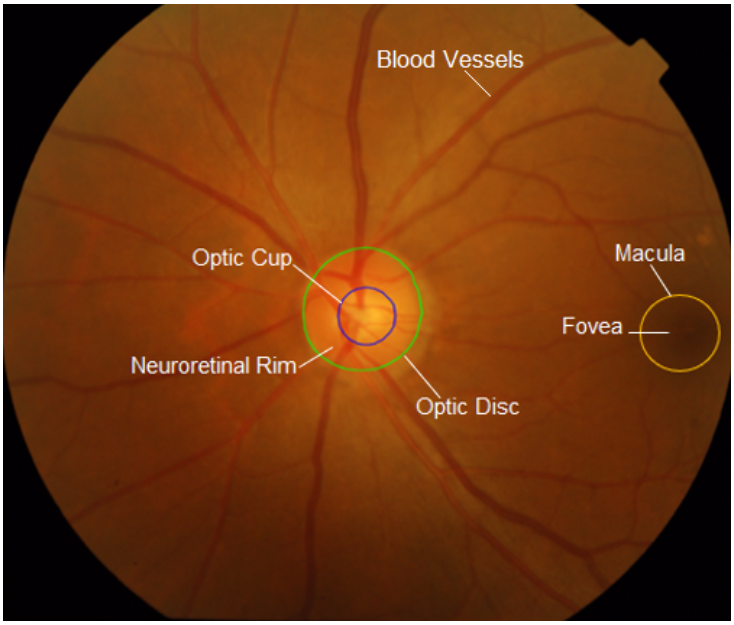


Figure 1.1: Key structures in a fundus image [57].

Fundus images that display these key structures may be stored in publicly available datasets for use in research and diagnostic applications. A summary of the features and contents of these datasets is presented in Table 1.1, followed by a more detailed overview below.

The REFUGE dataset [47] was developed as part of the Retinal Fundus Glaucoma Challenge and comprises 1,200 high-quality colour fundus images from Chinese female patients. Of these, 800 images were acquired using the Canon CR-2 AF [2] with a resolution of 1634×1634 pixels (px), while the remaining 400 were obtained with a Zeiss Visucam 500 fundus camera [8] at a resolution of 2124×2056 px. The images were collected across various hospitals and clinical studies, ensuring the anonymisation of personal data. The REFUGE dataset provides extensive manual annotations, including pixel-wise localisation of the fovea, and detailed markings of OD and OC. These annotations were created by seven independent glaucoma specialists, each with 5 to 10

years of experience. To ensure consistency, a majority voting system was used to generate a single segmentation per image. This segmentation underwent a final quality check by a senior glaucoma specialist with over 10 years of expertise. In this dataset, cases were classified as glaucomatous based on VCDR greater than 0.7 and the presence of visual field defects. The dataset is divided into three equal subsets: a training set featuring images captured with the Zeiss Visucam 500, and validation and testing subsets composed of lower-resolution images obtained with the Canon CR-2 AF. The Zeiss Visucam 500 is specifically designed to support detailed diagnoses of common eye diseases, including diabetic retinopathy and glaucoma, by delivering exceptionally high-quality fundus images. The CR-2 AF supports high-quality stereo photography, allowing clinicians to capture two sequential retinal images to create a stereo pair.

RIM-ONE [19] is a collection of glaucomatous and normal retinal images obtained from various Spanish hospitals [18] using a non-mydriatic Kowa WX 3D stereo fundus camera [23, 46] to capture its images. This release includes 85 normal images, 35 glaucoma suspect images, and 39 glaucoma images, stored in the JPG format, with a resolution of 2144×1424 px per image for stereo pairs. RIM-ONE provides manual segmentation of both OD and OC. These segmentations, saved in the PNG format, were independently performed by two experts, and the final segmentation masks were averaged for consistency.

The DRISHTI-GS dataset [57] consists of 70 glaucoma and 31 normal colour fundus images stored in PNG format, acquired with a 30-degree Field of View (FOV) and a resolution of 2896×1944 px. These images were captured at Aravind Eye Hospital in Madurai and focused on the OD region. The dataset includes images of male and female patients aged 40–80. Each image was manually segmented by four clinical experts with varying levels of experience (3, 5, 9, and 20 years). The dataset provides pixel locations [x,y] representing the averaged boundaries of OD and OC. These boundaries were determined by dividing the image region into 80 equal angular sectors centred on OD and averaging the manual markings in each sector. A segmentation soft map for OD and OC was generated by fusing the individual manual segmentations of all four experts. Additionally, the dataset includes CDR values for each image, as annotated by the four experts.

The Chákṣu dataset [30] consists of 1,345 retinal colour fundus images collected from Indian subjects aged 18 to 76 years, with all personal information anonymised. The images were captured using three different non-mydriatic fundus cameras: the Remidio Fundus-on-Phone (FoP) [49], the Forus 3Nethra Classic [40], and a Bosch handheld fundus camera [15]. The Remidio Fundus-on-Phone (FoP) is a retinal imaging device that leverages smartphone technology to capture high-quality fundus images. The Forus 3Nethra Classic fundus camera is a cutting-edge tool designed to acquire high-resolution images of the retina and the posterior segment of the eye. The Bosch handheld fundus camera is a portable imaging device developed to capture detailed images of the ocular fundus, including the retina, macula, and OD. The images captured by these cameras are divided into three datasets accordingly:

- REMIDIO: 1,074 images with a resolution of 2448×3264 px, including 150 classified as glaucoma and 924 as normal.
- FORUS: 126 images at a resolution of 2048×1536 px, with 18 identified as glaucoma and 108 as normal.
- BOSCH: 145 images captured at 1920×1440 px, where 20 are labelled as glaucoma and 125 as normal.

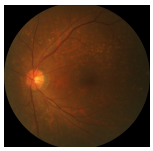
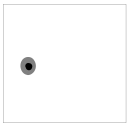
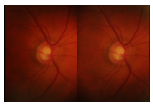
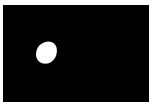
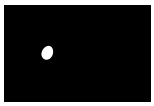
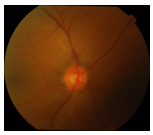
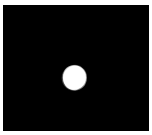
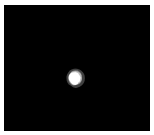
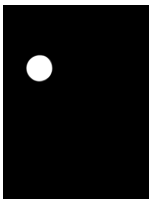
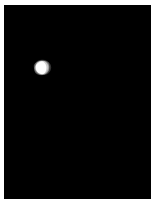
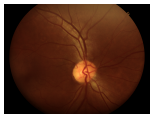
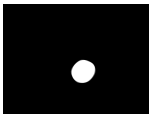
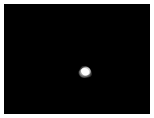
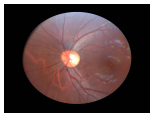
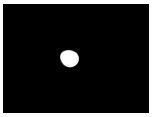
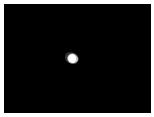
The images are stored in JPEG or PNG formats, and the dataset includes detailed ground truths for both OD and OC. A binary classification indicating glaucoma or non-glaucoma status is provided for each image, based on evaluations by five experts with 5 to 15 years of experience in ophthalmology. In addition to the individual annotations, the dataset provides single fused ground truths generated using various algorithms, including median, mean, and majority voting.

The quality of the images depends on the camera features, so their specifications are presented in Table 1.2. Visualizations of some digital fundus cameras are shown in Figure 1.2.

1.2. Deep Learning Methods for Image Segmentation in Glaucoma Identification

Traditional methods of evaluating eye diseases like glaucoma rely heavily on the expertise of clinicians to manually analyse fundus images,

Table 1.1: Summary of fundus image datasets.

Dataset and camera	Image	OD Label	OC Label	Size, px	Experts
REFUGE (Canon CR-2 AF)				1634×1634 and 2124×2056	7
RIM-ONE (Kowa WX 3D)				2144×1424	2
DRISHTI-GS (30-degree FOV)				2896×1944	4
REMIDIO (Remidio Fundus-on-Phone)				2448×3264	5
FORUS (Forus 3Nethra Classic)				2048×1536	5
BOSCH (handheld Bosch)				1920×1440	5

measure CDR, and assess subtle changes in the optic nerve. This process can be slow, labour-intensive, and prone to human error. In contrast, with the availability of diverse, high-quality retinal image datasets, automated systems powered by advanced technologies such as DL methods could quickly and consistently analyse vast amounts of retinal images, providing rapid assessments. Therefore, the use of DL methods in image segmentation has become increasingly important in the field of ophthalmology, particularly for the identification and diagnosis of glaucoma. DL techniques, specifically in the area of image segmentation, allow for the automatic identification and analysis of critical features in fun-

Table 1.2: Technical specifications of digital fundus cameras.

Fundus Camera	Kowa WX 3D	Zeiss Visucam 500	Canon CR-2	Remidio FOP NM-10	Forus 3Nethra Classic	Bosch Hand-held
Type of Camera	3D Stereo Fundus Camera	Non-Mydriatic Fundus Camera	Non-Mydriatic Retinal Camera	Mydriatic and non-mydriatic	Non-mydriatic	Mydriatic and non-mydriatic
Type of Image	Normal/SP/Stereo	Color, red-free, blue and red images	Color, Digital Red-Free, Digital Cobalt	Color, red free and infrared	Color	Color, red free and infrared
Picture angle (degrees)	45 and 34	45 and 30	45	40	45	Unknown
Minimum pupil diameter (mm)	4.0 and 3.5	4.0 and 3.3	4.0 and 3.3	3.0	Unknown	3.5
Dioptric compensation for patient eyes (D)	-12 to +13, -32 to -10, +10 to +35	-35 to +35	-10 to +15, -31 to -7, +11 to +33	-30 to +30	Unknown	-20 to +20
Light source	Unknown	Xenon flash lamp, 22 flash levels, Max. 60 Ws	Infrared LED, White LED y	Infrared and cool white LED	Unknown	Unknown
Capture image resolution (Megapixels)	Unknown	5.0	20.2	Unknown	3.0	5.0

Figure 1.2: Eye fundus cameras from top left: Forus 3Nethra Classic [40], Zeiss Visucam 500 [8], Canon CR-2 AF [2], Kowa WX 3D [23], Remidio FOP NM-10 [49], Handheld Bosch [15].

dus images, such as OD and OC. Leveraging detailed annotations and manual segmentations from the eye fundus image datasets, DL models can be trained to recognise subtle patterns in the retina, enabling early detection of glaucoma and potentially preventing irreversible vision loss. Furthermore, an overview of the most recent studies on the application of DL for glaucoma detection is presented, highlighting the advantages and disadvantages of the proposed approaches.

GDCSeg-Net, an encoder-decoder network designed in [82] for OD and OC segmentation. The model addresses image segmentation challenges across various datasets by employing a mixed training strategy, enabling it to perform effectively on diverse image datasets. GDCSeg-Net incorporates two key components. One of them is a multiscale weight-shared attention (MSA) module. This module captures multi-scale features using four parallel depth-wise separable convolutions with dilation rates of 1, 3, 5, and 7. The shared weights across these convolutions reduce the number of model parameters, improving computational efficiency and minimising overfitting. The second one is a densely connected depth-wise separable convolution (DSC) module. This module fuses the multi-scale features extracted by the MSA module layer-by-layer through dense connections, ensuring effective feature integration. The network was validated on multiple publicly available

datasets, including DRISHTI-GS, MESSIDOR, IDRiD, and REFUGE.

Two CNN models with 39 layers each for segmenting OD and OC to calculate CDR were developed by the authors in [69]. CDR was computed based on the segmentation outputs of these models. The authors identified a key issue in existing CNN models: resolution loss during the training process caused by down-sampling layers, which can result in a loss of critical image details. To address this, the authors introduced a novel approach by incorporating an up-sampling layer for every down-sampling layer in the network. By adding more layers to both CNN models, a greater number of image features are extracted, enhancing the segmentation process. However, the 39-layer deep CNN architecture also demands significant processing power and memory. The proposed method allows the lost resolution to be restored, ensuring that the output image resolution matches the input resolution. The model was trained and evaluated on the DRISHTI-GS dataset. This approach is promising for applications requiring precise segmentation, but its implementation needs to be balanced against computational efficiency and dataset limitations.

The authors of [21] proposed a Recurrent Fully Convolution Network (RFC-Net) for the automatic joint segmentation of OD and OC. This architecture is designed to capture high-level contextual information while preserving spatial details, ensuring accurate segmentation results. Images are processed at multiple scales to capture both global and local features, improving segmentation precision. The network generates intermediate outputs at different stages to guide learning and refine segmentation results. The images are transformed into a polar coordinate system to better represent the circular structure of OD and OC, facilitating segmentation. Four recurrent units are included to capture dependencies across spatial regions, enabling the network to learn detailed features over multiple iterations. Connections between earlier and later layers are added to preserve fine-grained spatial details and improve gradient flow during training.

The study [27] introduced G-Net, a system designed to differentiate glaucomatous eyes from non-glaucomatous ones using two modified versions of the U-Net architecture [50]. The modifications aimed to improve the model's ability to extract relevant features and enhance segmentation accuracy. The first modification involved increasing the size

of the convolutional filters in the original U-Net model. This adjustment allows the network to capture more features from the images, which is crucial for precise segmentation tasks like OD and OC delineation. The second modification focused on the removal of the dropout layer from the U-Net architecture. Since the model was trained on a limited dataset, the authors believed that dropout could hinder performance, so they opted to remove it to improve learning. The system is divided into two models. The first model is dedicated to OD segmentation, providing an accurate boundary for the region of interest. The second model works with cropped OD regions (512×512 px) extracted using the segmentation output from the first model. By focusing exclusively on the OD, the model can more effectively analyse the enclosed OC, leading to enhanced segmentation performance.

In [36], a dual DL-based approach was developed that aimed to automate both glaucoma detection and OD and OC segmentation. This method combined a modified U-Net architecture, named X-Unet, for segmentation and a modified version of DeepLab+3 for glaucoma classification. The X-Unet architecture improves upon the original U-Net by incorporating several techniques, such as multiple inputs, squeeze and excitation blocks, and feature suppression mechanisms. Because of the multiple inputs, the network accepts more raw pixel information during training, enabling it to better capture relevant features. Squeeze and excitation blocks are used to adaptively weigh features from different convolutional layers, allowing the network to focus on more informative features. The Feature Suppression mechanism suppresses irrelevant features derived from global information, helping to improve the focus on the region of interest. For glaucoma classification, the authors modified the DeepLab+3 architecture by replacing the last layer with a global average pooling layer, followed by a fully connected layer, to predict the likelihood of glaucoma. Additionally, the encoder employed Atrous Spatial Pyramid Pooling and cascaded convolutions to capture multi-scale, pixel-wise contextual information. The decoder refined the low-level features using the multi-scale context extracted by the encoder. The final results were obtained by averaging the outputs from multiple models.

The authors in [54] proposed the Attention U-Net, which enhances the limitation of the traditional U-Net architecture, that lacks sufficient

feature information in its initial layers, by introducing attention gates at each skip connection. These gates enable the network to emphasise crucial regions of the input image while suppressing less relevant background information, leading to improved segmentation accuracy. The model architecture is composed of three primary components: an encoder, a decoder, and attention gates embedded in the skip connections to refine feature selection. The standard U-Net encoder is substituted to enhance segmentation performance with pre-trained deep convolutional networks such as VGG19 [53], Inception-v3 [60], and ResNet50 [24], which serve as backbones in the contracting path. Among these alternatives, ResNet50 proved to be particularly effective due to its residual learning framework, which addressed the vanishing gradient problem.

An attention-guided network named Residual Attention U-Net (RAUNet) was presented in [45] for the segmentation of the cataract surgical instrument. RAUNet improves segmentation accuracy by utilising an encoder-decoder framework that generates high-resolution masks while effectively preserving both global context and fine-grained details. RAUNet employs a pre-trained ResNet34 [24] as its encoder to extract rich semantic features while optimising model efficiency and accelerating inference. Leveraging ImageNet training, this backbone improves feature representation and enhances adaptability for medical imaging applications. In the decoder, a newly designed Augmented Attention Module (AAM) integrates multi-scale features to enhance segmentation accuracy. Furthermore, transposed convolution is applied for up-sampling, facilitating the restoration of fine structural details. The AAM improves feature selection by capturing semantic relationships across multiple channels. In medical imaging, various anatomical structures correspond to distinct feature activations. AAM strengthens this differentiation by incorporating global context from high-level feature maps and combining it with spatial details from lower-level feature maps. While high-level semantics help highlight essential regions, the global context of lower-level features aids in suppressing unnecessary background information, leading to more precise segmentation.

A modification of the U-Net named UNet++, developed in [81], builds upon the traditional U-Net by introducing a more refined encoder-decoder structure with redesigned skip connections and deep super-

vision, leading to improved feature fusion and segmentation accuracy. Unlike U-Net, where encoder feature maps are directly transferred to the decoder, UNet++ enhances this process through dense convolutional blocks in the skip pathways. These blocks refine the feature representations by progressively aligning the semantic information of the encoder and decoder feature maps. In UNet++, the depth of the pyramid determines the number of convolutional layers within the skip pathways, facilitating a smoother transition between encoder and decoder feature representations. The skip connection between nodes incorporates three convolutional layers, each following a concatenation step that merges the prior convolution's output with the upsampled features from the lower level. This structured refinement process eases optimisation by delivering semantically enhanced feature maps to the decoder.

1.3. Glaucoma Identification Using CNNs Ensemble

In recent years, ensemble learning-based CNN architectures have been increasingly applied in the development of automated glaucoma assessment systems through the analysis of fundus images.

Two ensemble models for segmenting OD and OC, which rely on a modified U-Net architecture integrated with three FCNs, were presented by the authors in [29]. In each ensemble model, the final segmentation results were derived by averaging the predictions from all three individual FCNs. The study found that FCNs trained for two-class segmentation (OD and OC separately) outperformed those trained for three classes simultaneously. The proposed ensemble models improved segmentation accuracy by averaging multiple FCN outputs, mitigating labelling errors. However, the authors acknowledged challenges in obtaining consistent fundus image annotations due to inter-reader variability and the relatively small dataset size. Additionally, the authors emphasised the importance of accurate ROI detection in improving OD and OC segmentation. Their experiments compared Mask R-CNN with a modified U-Net-based FCN for ROI detection, finding that while both performed well, Mask R-CNN produced fewer over-labelling errors, making it more dependable for subsequent segmentation steps. In the proposed approach, for OD segmentation, the raw ROIs were provided as inputs to the model. For OC segmentation, the model was fed with masked ROIs (ROIs where the OD was previously segmented) to focus specifi-

cally on the OC. The method was trained and evaluated on the REFUGE and RIGA datasets.

The authors in [10] developed an ensemble approach for the segmentation of OD and OC aimed at calculating CDR. The method combined a U-Net architecture for segmentation with Random Sample Consensus (RANSAC) to ensure that the predicted shapes for OD and OC resemble ellipses. For glaucoma classification, the authors employed transfer learning using MobileNet-v2, a pre-trained network on the ImageNet 1K challenge dataset. The method was trained on a combined set of images sourced from multiple datasets, such as RIM-ONE and DRISHTI-GS, and the results from OD and OC segmentation were merged to compute a Glaucoma likelihood score.

A method, introduced in [83], utilised an ensemble approach based on CNNs for glaucoma detection and OD and OC segmentation. The researchers focused on improving the segmentation accuracy while minimising the computational complexity. They employed a deep CNN model that was trained using numerous patches extracted from the DRISHTI-GS dataset, a widely used dataset for glaucoma research. The segmentation process was enhanced by using an entropy sampling technique, which effectively reduced the computational burden during training by selectively choosing the most informative patches from the data. Entropy sampling, as applied in this method, helped the model focus on the most relevant parts of the images, thus reducing the time and resources needed for training while still achieving high-quality segmentation. This approach ensured that CNN model did not waste resources processing redundant or irrelevant information, making the training process more efficient and scalable. The model's performance was evaluated on the DRISHTI-GS dataset.

The authors in [3] developed an OD segmentation system based on an ensemble of ten DL-based semantic segmentation models. These models included Gated Skip Connections (GSCs), U-Net, DoubleU-Net, CGNet, DeepLabV3+, SegNet, ERFNet, SQNet, LinkNet, and ESNet. The system aggregated the outputs of these individual models to produce a final segmentation result. The aggregation was performed using the Ordered Weighted Average operator, applied at each pixel of the input image. In the post-processing step, a threshold of 0.5 was applied to the aggregated results, and the class with the maximum activation

was chosen as the label for that pixel. The method was evaluated on fundus images captured from Hospital Sant Joan de Reus, where the ensemble consisting of Double U-Net, GSCs, and DeepLabV3+ models provided the best results.

In [68], a two-phase glaucoma classification framework utilising four pre-trained deep CNNs, such as AlexNet, NasNet-Large, Inception ResNet-v2, and Inception-v3 [60] was presented. By applying transfer learning, where the training process begins with pre-trained network weights, they effectively reduced the overall training time. The approach was evaluated using extracted ONH regions from three publicly available datasets (ACRIMA, ORIGA-Light, and RIM-ONE) along with private datasets obtained from local hospitals. Classification performance was enhanced by integrating the outputs of the four CNNs into a single ensemble model, where final decisions were determined using five voting strategies: Proportional Voting, Majority Voting, Averaging, Accuracy/Score-Based Weighted Averaging (ASWA), and Accuracy-Based Weighted Voting (AWV).

The authors of [31] employed an ensemble approach combining the DL models VGG-16 [55], MobileNet, and ResNet50 [24] to classify OD in fundus images from the REFUGE dataset. They implemented both hard voting and average voting techniques to aggregate the model predictions. Performance evaluation was conducted using several metrics, including binary accuracy, precision, recall, area under the curve (AUC), as well as counts of true positives, true negatives, false positives, and false negatives.

The study [14] introduced a deep ensemble model that used a stacking ensemble strategy. This approach integrated multiple deep CNN architectures, including Xception, Inception-v3 [60], DenseNet-201, MobileNet-V2, EfficientNet, VGG-16 [55], VGG-19 [53], GoogLeNet, AlexNet, ResNet-18, ResNet-50 [24], ResNet-101, and SqueezeNet. A support vector machine was employed as the final classifier to distinguish between glaucoma and normal fundus images.

1.4. Measures and Thresholds for Risk of Glaucoma Identification

When employing automated segmentation algorithms that rely on CDR as a threshold for distinguishing between glaucomatous and normal

cases, substantial variability is observed in the threshold selection across studies utilising CNN-based methods, and the appropriate threshold remains unclear. Table 1.3 presents the variations in CDR metric and its thresholds across studies, followed by a detailed review of the studies.

Table 1.3: Variation in CDR metric and its thresholds across studies

References	**Metric**	**Threshold**
[75]	VCDR	> 0.7
[80], [33], [6]	VCDR	> 0.5
[71]	VCDR	$\geqslant 0.6$
[39]	VCDR	> 0.6
[56]	VCDR	> 0.4
[44]	ACDR VCDR and HCDR	$\geqslant 0.3$ $\geqslant 0.5$
[10]	VCDR	$= 0.6$
[13]	ACDR	> 0.3

MRSNet, proposed in [75], is a segmentation network built with an encoder-decoder architecture. The encoder incorporated a convolutional block enhanced by large kernel convolutional attention, enabling efficient multi-scale feature extraction and improved spatial attention while minimizing computational cost. To enhance segmentation performance, the model employed a multi-resolution image combination strategy and incorporated a compression and excitation module for adaptive input feature extraction. The network was trained using a five-fold cross-validation strategy and validated experimentally on three widely-used datasets: REFUGE, DRISHTI-GS, and RIM-ONE. Additionally, performance was evaluated by calculating the absolute error, defined as the absolute difference between the predicted VCDR and the gold standard VCDR. Higher VCDR (> 0.7) indicated an increased risk of glaucoma, while lower values suggested otherwise. Smaller absolute errors reflected more accurate segmentation results.

A network called EfficientNet and Attention-based Residual Depth-wise Separable Convolution (EARDS) [80] was designed to enhance the segmentation of OD and OC for glaucoma diagnosis. The model leveraged EfficientNet-b0 as an encoder, which enabled effective boundary representation by efficiently extracting multi-scale features. This joint segmentation approach improved precision in delineating OD and OC boundaries, which are critical for calculating CDR. The EARDS network was evaluated on two widely used datasets, DRISHTI-GS and REFUGE.

VCDR value greater than 0.5 was used as a threshold to identify potential glaucoma cases. The EARDS network's use of depth-wise separable convolutions, combined with an attention mechanism, helped to reduce computational complexity while retaining high segmentation accuracy. This made it particularly suitable for resource-constrained environments without compromising on performance.

A two-task network named CDRNet [71] was designed for precise measurement of CDR. The framework consisted of two interconnected networks: one focused on weakly supervised image segmentation, leveraging an extended multiple instance learning (MIL) formulation and smooth maximum approximation, and the other performed bounding-box regression, generating class-specific bounding box predictions at a single scale while preserving the original image resolution. Additionally, a class-specific bounding-box normaliser and enhanced IoU (eIoU) were proposed to improve prediction accuracy. By combining these tasks, the model ensured accurate segmentation and localisation of OD and OC. VCDR served as the diagnostic threshold for glaucoma detection, with cases flagged as glaucoma when VCDR $\geqslant 0.6$, while VCDR < 0.6 was classified as normal.

The authors of [39] proposed an encoder-decoder architecture called Responsive Fusion Attention U-ConvNext, tailored for the semantic segmentation of OD and OC in fundus images. A novel Dual-Path Response Fusion Attention (DPRFA) module was introduced to enhance the integration of encoder and upsampled decoder feature maps, improving segmentation accuracy. To further optimise the training process, the authors designed a composite loss function that combines cross-entropy, Dice, and IoU losses, ensuring accurate model learning across datasets. The network was developed in four variants, each with different sizes, and trained on two widely-used datasets, DRISHTI-GS and REFUGE. The model also calculated VCDR, where a threshold of VCDR > 0.6 is used to classify eyes as glaucomatous.

A region-based deep CNN (R-DCNN) [33] was designed for the joint segmentation of OD and OC, aiming to achieve precise measurement of CDR for glaucoma detection. The architecture leveraged ResNet34 as the backbone for feature extraction and incorporated a disc proposal network for OD segmentation, a cup proposal network for OC segmentation, and an attention module to enhance communication

between the two networks. The attention mechanism improved feature alignment, resulting in more accurate segmentation. The R-DCNN was evaluated on an in-house dataset annotated by four experienced ophthalmologists (5-8 years of expertise) and the public datasets DRISHTI-GS and RIM-ONE. For glaucoma screening, the model used CDR threshold of VCDR > 0.5, indicating suspicion of glaucoma.

The authors of [44] developed a U-Net-based architecture that incorporated transfer learning models such as Inception V3 and Inception ResNet V2 to segment OD and OC for glaucoma prediction. The model was trained and evaluated on three major datasets: DRISHTI-GS, RIM-ONE, and REFUGE. The segmentation results were assessed using various metrics for CDR, including VCDR, HCDR, and ACDR. To classify eyes as abnormal, the model applied a threshold value of 0.5 for VCDR and HCDR, and 0.3 for ACDR. Eyes exceeding these thresholds were considered indicative of glaucoma, while those below the thresholds were classified as normal.

A multimodal classification approach for glaucoma diagnosis, presented in [56], integrated both multimodal data and features extracted from retinal fundus images. The methodology incorporated a hybrid fusion strategy that combined the strengths of six machine learning (ML) models, an ensemble model, and two deep CNNs for classification. Additionally, the authors developed a hybrid deep network named VGG-CapsNet, which merged two powerful DL frameworks: VGG-16 and CapsNet. This approach was validated through experiments on three public datasets REFUGE, ORIGA, and ACRIMA, as well as on four different combinations of these datasets. For classification, the model used CDR threshold of 0.4 to distinguish between normal (CDR < 0.4) and abnormal (CDR $\geqslant 0.4$) eyes.

A system, designed in [6] as a binary classifier, combined an ensemble of five neural networks: Inception-ResNet-V2, NasNet, Xception, Inception, and a tree-learning algorithm. The model classified retinal fundus images into two categories: normal (with VCDR ≤ 0.5) and glaucoma suspect (with VCDR > 0.5). The neural network ensemble was trained on images from several datasets, including the Age-Related Eye Disease Study (AREDS), the Singapore Malay Eye Study (SiMES), and the RIM-ONE dataset. The system was validated on the Online Retinal Fundus Image Dataset for Glaucoma Analysis (ORIGA).

1.5. Related Research in Lithuania

Research on automated glaucoma identification in Lithuania has evolved over nearly two decades, with early efforts primarily focused on developing image processing and classical ML methods for the structural analysis of the OD and OC.

One of the earliest studies in 2007 by Treigys and Šaltenis [66] explored the potential of neural networks for disease classification, demonstrating that, when combined with dimensionality reduction techniques, such models could effectively distinguish glaucomatous eyes based on OD structural parameters. Further work by Treigys et al. [65] proposed an automated algorithm for OD localisation and parameterisation in colour retinal images, combining morphological image processing, adaptive edge detection, and geometric modelling to achieve accurate OD detection. For OD centre localisation, the study utilised the iterative circular Hough transform. The study prepared by Treigys and Dzemyda [64] addressed the challenge of linear monitoring, introducing an automated image registration method to align retinal images taken at different times, thus enabling more precise disease progression analysis. Parallel to these developments, Bernatavičienė et al. [4] investigated automated rule induction methods for classifying ophthalmological data, emphasising the diagnostic value of OD and OC diameter ratio. Subsequent research by Buteikienė et al. [7] advanced these ideas by presenting a fully automated approach for OD detection using an elliptic curve model and Bayesian optimization, achieving a strong correlation (0.8) between automated and expert evaluations. Later, Stabingis et al. [58] extended the focus toward broader fundus analysis, developing a system for automated measurement of the artery-to-vein ratio, an important indicator in ocular disease detection, and addressing challenges associated with vessel tree extraction.

Together, these studies reflect a consistent effort to automate and standardise glaucoma-related diagnostics through classical image processing, geometric modelling, and early neural network methods. However, while these approaches established a strong foundation, they were limited by feature extraction.

In recent years, the emergence of DL has transformed ophthalmic image analysis, offering end-to-end feature learning and performance

on unseen data across diverse datasets. The 2023 study by Raudonis et al. [48] introduced a DL-based method for automatic microaneurysm detection in colour fundus images – an essential step for the early diagnosis of diabetic retinopathy. The method employed an ensemble of three networks, such as U-Net, ResNet34-UNet, and UNet++, to improve segmentation accuracy. Building on these DL advances, Çelik et al. [9] proposed a decision support system for the automated detection of optic nerve hypoplasia, a condition closely related to OD morphology and relevant for glaucoma screening. Their method integrated a U-Net architecture with a pretrained ResNet encoder to segment the OD and fovea.

These advancements highlight the ongoing evolution from classical, feature-based image processing methods toward DL–driven, data-centric diagnostic systems. Within this broader transformation, the application of DL techniques to glaucoma identification remains both timely and essential for achieving higher diagnostic accuracy and clinical applicability.

1.6. Conclusions of the Chapter

In conclusion, fundus imaging is an essential tool in ophthalmology for diagnosing and monitoring eye diseases, particularly glaucoma. The quality of these images is largely determined by the camera used, with technological advancements ranging from high-resolution systems such as the Forus 3Nethra Classic and Zeiss Visucam 500 to portable devices like the Bosch30 handheld fundus camera. Despite these advances, manual glaucoma diagnosis remains time-consuming and resource-intensive, requiring ophthalmologists to assess multiple parameters. Given that glaucoma is a progressive disease that can cause irreversible vision loss, there is a clear need for more efficient diagnostic methods.

Research on automated glaucoma identification in Lithuania has evolved over nearly two decades, initially focusing on image processing and classical ML methods for analysing OD and OC. In recent years, the emergence of DL has transformed ophthalmic image analysis globally. The works with DL models demonstrate significant progress, especially in OD segmentation, while OC segmentation continues to be more challenging. Moreover, most results have been obtained using CNNs trained on single datasets, leaving open the question of whether these

models can achieve consistent performance across multiple data sources. This highlights the need for methods with enhanced feature extraction in image segmentation. Architectures such as U-Net, Attention U-Net, RAUNet, and UNet++ show particular promise in this regard and have therefore been selected for the experimental part of this dissertation.

The goal of this research is to contribute toward developing a diagnostic system suitable for primary care, where affordable handheld fundus cameras allow family doctors to screen patients for glaucoma risk and refer them when necessary. Publicly available datasets such as REFUGE, RIM-ONE, REMIDIO, FORUS, BOSCH, and DRISHTI-GS provide a foundation for training DL models capable of supporting such systems.

Finally, while most existing studies rely on VCDR as the primary metric, alternatives such as ACDR and HCDR have also been investigated. For example, some studies suggest using ACDR with a threshold of greater than 0.3 for glaucoma detection [12], [44]. In other studies [44], HCDR greater than 0.5 is treated as an abnormality. Thus, despite intensive research in this field, the lack of standardised thresholds across studies arising from differences in datasets, expert labelling, and evaluation metrics makes it difficult to compare results and underscores the need for greater consistency in future research.

2. METHODS

This chapter presents the methodology employed in this study for glaucoma identification based on eye fundus images. It is divided into three main sections: image preprocessing techniques, metrics for evaluating segmentation quality, and research methods for glaucoma identification using fundus analysis.

2.1. Image Preprocessing for Fundus Images

Image preprocessing can significantly address some of the challenges faced in the development of automated systems for glaucoma detection, especially those related to computational efficiency, dataset limitations, model adaptability, and OC segmentation.

Preprocessing techniques such as noise reduction and image resizing help simplify input data while preserving critical features. By refining images and eliminating unnecessary details, these methods enhance computational efficiency. In case of data limitation, data augmentation techniques, such as rotations, scaling, and flipping, can artificially expand the size and variety of a training dataset, improving their ability to perform on unseen data. Techniques like contrast enhancement [83], adaptive histogram equalisation [29], or normalisation [83] can help standardise images by reducing variability across datasets, ensuring that models trained on one dataset are better equipped to perform effectively on others. Image preprocessing can also aid in OC segmentation by focusing on relevant regions of interest through ROI cropping, which highlights areas such as the OC, thereby further improving segmentation accuracy. Each of these techniques is worth discussing in more detail below.

Noise reduction involves eliminating unwanted random variations or distortions, commonly referred to as noise, from an image to enhance image clarity while retaining its essential features. Smoothing filters, including the mean filter, the Gaussian filter, and the median filter, are one of the noise reduction techniques. The mean filter reduces sharp intensity changes by averaging the values of neighboring pixels, while the Gaussian filter applies a weighted average based on a Gaussian distribution, achieving smoother results with less blurring. The median filter replaces each pixel value with the median of its neighborhood, making

it particularly effective for removing salt-and-pepper noise. Another type of filter is edge-preserving filters that reduce noise without compromising important edges in the image. The bilateral filter achieves this by considering both spatial proximity and intensity differences when smoothing. Similarly, the non-local means method averages similar patterns throughout the image, avoiding over-smoothing and maintaining structural details. One more noise reduction technique is deep learning-based approaches, such as denoising autoencoders and CNNs. These approaches utilize large datasets to learn complex patterns and details within images, enabling them to differentiate between relevant features and noise. Denoising autoencoders function by encoding a noisy image into a compact latent space, then reconstructing a cleaner version of the image, thus learning how to map noisy data to its noise-free counterpart.

While noise reduction techniques enhance image quality by removing irrelevant variations, their application must be carefully considered to avoid blurring important details. This becomes particularly relevant in the context of fundus images, which are typically large and contain peripheral details outside ROI, which occupies only a small portion of the image, such as the OD, macula, or blood vessels. The resolution of these images varies significantly depending on the capturing device, for instance, 2144×1424 px [16] or 1920×1440, 2048×1536 and 2448×3264 px [30], leading to inconsistencies in data representation. To address this, various cropping techniques are commonly applied to effectively isolate ROI. These techniques may involve manual annotation, automated segmentation methods, or heuristic-based cropping to focus on relevant areas. Cropping not only reduces computational overhead by eliminating extraneous regions but also enhances the accuracy of downstream tasks, such as segmentation. Furthermore, standardising images through consistent cropping ensures uniformity in data processing pipelines, enabling models to perform more effectively across datasets [27]. Different approaches to ROI extraction have been documented in Table 2.1, ranging from traditional ROI extraction techniques to advanced ML-based methods.

Even after ROI extraction, the resulting images can still require considerable computational resources due to their large size, which can hinder processing efficiency and model performance. To address this challenge, image resizing is a common practice. By reducing the dimen-

Table 2.1: ROI extraction methods.

References	Method	ROI size, px	ROI resized to, px
[36]	Disc-aware Ensemble Network [17] to identify ROI centre	600×600	128×128
[27]	512×512 region centred on OD	512×512	128×128
[35]	Circular Hough transform to locate the OD and determine its centre and radius	480×480 and 560×560	240×240
[73]	512×512 px field of view	512×512	256×256
[68]	Bounding box of 1.5 times OD radius and resizing according to the default input size of each Deep CNNs	-	256×256, 299×299, 331×331
[74]	Image opening and Gaussian blur to highlight the brightest pixels, followed by Canny edge detection to define ROI	-	640×640
[69]	Sobel edge detection method and the Watershed algorithm to locate the region of OD	512×512	-
[21]	YOLO-based models taking images of different sizes on the centre point of OD	400×400, 500×500, 550×550, 600×600, 650×650, 700×700, 750×750, 800×800, 850×850, 900×900	512×512

sions of images, the computational load can be significantly decreased while maintaining essential features for analysis. In many studies, images were resized to standard dimensions such as 128×128 [27, 36], 256×256 [28, 76], or 512×512 px [21].

The resizing process typically involves interpolation techniques, which estimate and generate new pixel values based on the existing ones. Among the most widely used interpolation methods are bilinear, bicubic, and nearest neighbor interpolation [22].

Bilinear interpolation [22] computes the new pixel value by averaging the nearest four pixels, providing smoother transitions. Bicubic interpolation [22] considers the nearest 16 pixels, produces even smoother results, but is computationally more expensive. Nearest neighbour interpolation [22], on the other hand, is the simplest method, assigning the value of the nearest pixel to the resized image, though it may result in a less smooth appearance.

In addition to noise reduction, ROI extraction, and resizing, image normalisation by scaling pixel values to a consistent range, such as $[0, 1]$ or $[-1, 1]$, is a common practice applied to ensure uniform input for the model. Similarly, contrast enhancement methods like Contrast Limited Adaptive Histogram Equalization (CLAHE) are often employed to improve the visibility of low-contrast images [69]. CLAHE is particularly helpful in enhancing fine details. However, some studies intentionally avoid procedures like contrast or illumination enhancement to preserve the natural variability of input data, making the learning of deep learning models more dynamic [68].

Beyond these preprocessing steps, augmentation techniques are also crucial in enhancing dataset diversity. These techniques vary significantly across studies, with each employing different strategies to introduce more variability. For instance, Juneja et al. [27] applied image rotation in 20° increments, ranging from 0° to 180°, along with horizontal and vertical flipping. Likewise, Liu et al. [36] utilised specific rotations at -90°, 180°, and 270°, along with flipping to increase variability. In contrast, Yan et al. [75] took a broader approach, incorporating scaling (both upscaling and downscaling), flipping, and elastic deformation to further expand the dataset.

Although the discussed image preprocessing techniques can enhance the performance of DL models, making them more reliable and efficient for applications in glaucoma detection, further advancements in other areas, such as model architectures and training strategies, are also necessary.

2.2. Metrics for Segmentation Quality Estimation

After image preprocessing is completed, the next crucial step in the development of automated glaucoma identification systems is to train deep learning models to perform the image segmentation task. To assess the

performance of segmentation algorithms, various metrics are employed to quantify the accuracy of the segmented areas compared to ground truth labels. Among the most frequently used metrics for evaluating segmentation quality are Dice and IoU, which both measure the degree of overlap between the predicted segmentation and the ground truth. These metrics are essential in determining how effectively the model identifies critical regions, such as OD and OC in retinal images, which are crucial for glaucoma diagnosis.

Dice is a statistical measure that evaluates the similarity between two sets, commonly applied in medical image segmentation tasks [26, 38, 67, 82]. Dice ranges from 0 to 1, with a higher value signifying better similarity between the predicted and true segmented regions. It is defined by the formula

$$\text{Dice} = \frac{2\left|S \cap L\right|}{|S| + |L|}, \tag{2.1}$$

S is the result based on segmentation, L – the ground truth label.

IoU evaluates the overlap between predicted and ground truth regions. IoU is calculated by the ratio of the intersection of the predicted and ground truth areas to their union

$$\text{IoU} = \frac{\left|S \bigcap L\right|}{\left|S \bigcup L\right|}, \tag{2.2}$$

Like Dice, IoU ranges from 0 to 1, with a higher score reflecting more accurate segmentation. IoU is often used alongside Dice to provide a well-rounded evaluation of segmentation performance [3, 29, 62].

As both Dice and IoU provide insights into the accuracy of segmenting OD and OC in retinal images, they are crucial for calculating CDR in glaucoma detection. CDR is an essential parameter in diagnosing glaucoma, as it reflects the relative size of the OC compared to the OD. CDR can be calculated as three different metrics:

- The ratio of the vertical cup diameter (VCD) to the vertical disc diameter (VDD) is named vertical CDR (VCDR) [80]:

$$\text{VCDR} = \frac{vertical\,cup\,diameter}{vertical\,disc\,diameter} \tag{2.3}$$

- The ratio of the horizontal cup diameter (HCD) to the horizontal disc diameter (HDD), known as horizontal CDR (HCDR) [44]:

$$\text{HCDR} = \frac{horizontal\ cup\ diameter}{horizontal\ disc\ diameter} \tag{2.4}$$

- The ratio of the cup area to the disc area, termed area CDR (ACDR) [12]:

$$\text{ACDR} = \frac{area\ cup}{area\ disc} \tag{2.5}$$

Considering the importance of accurately segmenting OD and OC for glaucoma assessment, the application of statistical methods is often required when analyzing the resulting measurements. The ANOVA test [41] is commonly used to assess the differences between the means of more than two groups. In the context of OD and OC segmentation, ANOVA can be applied to compare the segmentation results of different models or algorithms across multiple datasets or segmentation methods. This is particularly useful when testing various DL approaches or training configurations to determine which method yields the most accurate and consistent results in segmenting OD and OC regions.

The statistical Student's t-test [41] is commonly employed to test whether the difference between two distinct groups is statistically significant. In the context of OD and OC segmentation, the t-test can be employed to compare the performance of segmentation models when applied to retinal images from glaucoma patients versus those from normal individuals. This comparison can focus on segmentation metrics such as Dice or IoU to assess whether the model performs differently for glaucoma-affected images compared to healthy retina images.

2.3. Research Methodology for Glaucoma Identification Based on Eye Fundus Estimation

The high-level view of the proposed methodology for the DL-based approach for glaucoma identification is illustrated in Figure 2.1, which is grounded in the experimental results presented in Chapter 3.

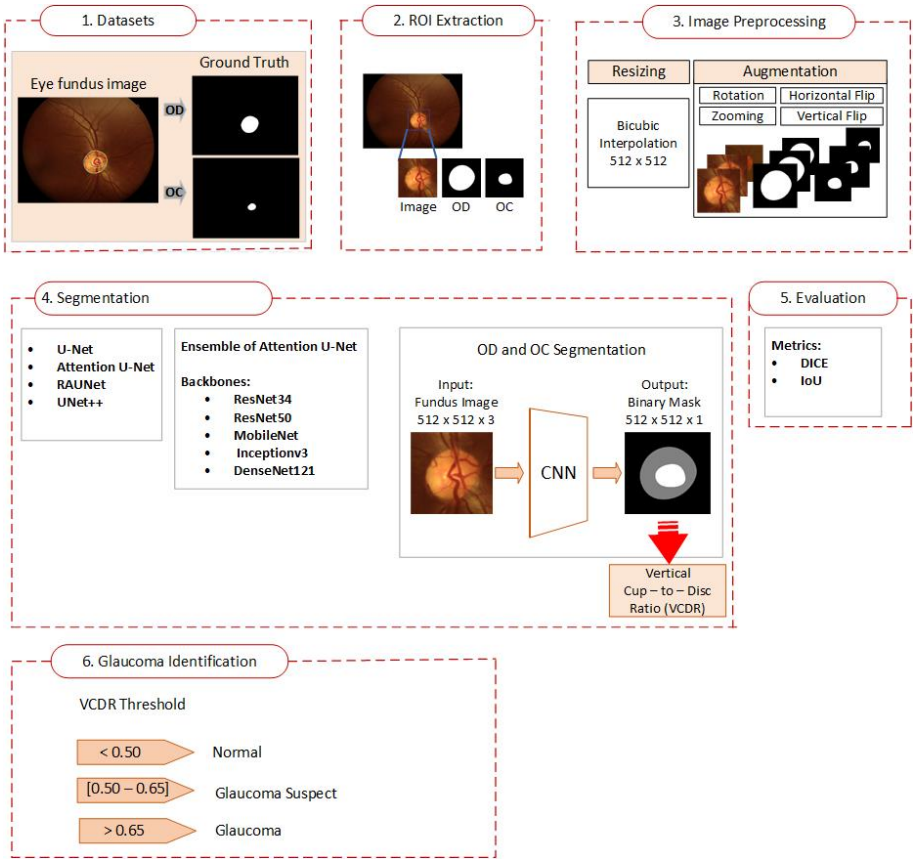


Figure 2.1: Proposed Methodology Workflow

The methodology follows a systematic and structured pipeline that integrates data preprocessing, model development, and performance evaluation. The process begins with the collection and integration of multiple publicly available eye fundus image datasets. To enhance the quality of the input data, an image preprocessing is employed. This includes ROI extraction, where the eye fundus image is cropped based on the OD boundary, as labelled. To further standardise the data, image resizing is applied, ensuring consistent input dimensions across different datasets. Data augmentation techniques such as flipping, rotation, and intensity variation are applied to the training set to increase data diversity, while the validation and test sets remain unchanged to ensure an unbiased performance evaluation. More details about the applied image preprocessing techniques can be found in Section 3.1. Finally, a mixed dataset is constructed by combining images from multiple sources before

splitting into training, validation, and test subsets, thereby ensuring that images from the same source can appear in different subsets, but individual samples are unique across them. The framework utilises multiple CNN architectures, such as Attention U-Net, RAUNet, and UNet++, to investigate the comparative effectiveness of various DL models for glaucoma detection. These architectures are trained under consistent experimental settings, with hyperparameter fine-tuning applied to optimise learning rates and batch sizes, where the modified Attention U-Net, with DenseNet121 as the pre-trained backbone, is shown as the most suitable. To ensure performance assessment, the trained CNN is evaluated using Dice and IoU, which measure segmentation accuracy of OD and OC regions. The experimental results indicate that VCDR is the most effective metric for classifying glaucoma stages (normal, glaucoma suspect, and glaucoma) when evaluated with the thresholds proposed in this study.

2.4. Conclusions of the Chapter

Traditional glaucoma diagnosis depends on expert evaluation of fundus images, making the process time-consuming. The availability of high-quality eye fundus image datasets has enabled DL-based automated systems to achieve notable success in OD segmentation. However, OC segmentation remains a challenge. Overcoming these limitations requires further research to improve segmentation performance across diverse datasets.

Image preprocessing techniques, such as noise reduction, contrast enhancement, data augmentation, and ROI extraction, play a critical role in enhancing model performance. These methods standardise image inputs, improving the adaptability of DL models across different datasets and ultimately leading to more accurate segmentation results.

To assess segmentation model performance, metrics like Dice and IoU are commonly used. Additionally, statistical methods such as ANOVA and the Student's t-test provide valuable insights by comparing model performance across different datasets and approaches. Advanced DL architectures, including Attention U-Net, RAUNet, and UNet++, offer improvements in feature extraction and segmentation accuracy.

Beyond segmentation accuracy, clinical glaucoma diagnosis relies on assessing structural changes in the optic nerve head. CDR remains

the most widely used metric for evaluating glaucoma, with different variations, such as area-based and diameter-based CDRs, serving as key indicators.

The results of this research and the principles of the methodology have been presented in multiple peer-reviewed articles [A.1], [A.2], [B.1], [C.1].

3.1. Impact of Eye Fundus Image Preprocessing on OD and OC Segmentation

Image preprocessing is a fundamental step in computer vision and medical image analysis, serving to refine raw fundus images before DL models process them. The primary objectives are to enhance image quality, suppress noise, normalise variations in lighting and resolution, and highlight clinically relevant structures such as OD and OC. By applying these techniques, preprocessing ensures that fundus images are in an optimal state for automated analysis, ultimately improving the accuracy of DL-based glaucoma detection systems.

A key aspect of preprocessing in fundus image analysis is addressing the variability in image resolution across datasets, which arises from differences in fundus camera specifications. Given that preprocessing also entails standardising input sizes for DL models, it is important to note that fundus image dimensions differ according to the acquisition device. This variability is evident in datasets such as DRISHTI-GS provides images at 2896×1944 px [57], RIM-ONE at 2144×1424 px [19], and REFUGE at 1634×1634 or 2124×2056 px [47]. Although these high-resolution images capture substantial retinal detail, they are generally too large for direct processing by CNNs and encompass extraneous background regions that are irrelevant for glaucoma assessment.

To address these challenges, ROI encompassing OD is extracted via cropping. ROI is defined as a square region centred on the OD, with sides twice the size of the OD. This ensures that the central area of the OD, along with surrounding diagnostically relevant retinal tissue, is captured while extraneous background is excluded. Following cropping, ROI dimensions vary across datasets: in DRISHTI-GS, they range from 674×674 to 1060×1060 px, in RIM-ONE from 456×456 to 890×890 px, and in REFUGE from 408×408 to 616×616 px. This preprocessing strategy ensures that CNN models receive inputs of consistent size while retaining critical anatomical features necessary for accurate glaucoma detection.

3.1.1. Image Resizing Level Impact

A key focus of the research is to analyse the influence of image size alignment techniques, specifically image resizing and interpolation methods, on the segmentation results. The dimensions of the cropped ROIs, discussed in Section 3.1, varied based on the original image sizes across different datasets. In this experiment, the DRISHTI-GS dataset (with ROI sizes ranging from 674×674 px to 1060×1060 px) and the RIM-ONE dataset (with ROIs ranging from 456×456 px to 890×890 px) were used. To standardise the varying sizes of cropped images, image resizing was applied, adjusting them to the common sizes of 128×128 [27, 36], 256×256 [28, 76], and 512×512 px [21, 73] using bilinear interpolation, a widely used method. The images were normalised by scaling the pixel values to the range [0,1]. Due to the limited number of images in the DRISHTI-GS (101 images) and RIM-ONE (159 images) datasets, data augmentation was applied to enhance the training set. Random image zooming by 20%, rotation by an angle of rotation from 0° to 45°, and horizontal and vertical flipping were applied to each image, which significantly increased the dataset size to 3,000 images for training. For evaluation, the test sets consisted of 30% of the eye fundus images from each dataset [43, 61], randomly selected to ensure representative coverage while avoiding overlap with the training and validation sets. A learning rate and batch size were selected using the KerasTuner framework, resulting in a learning rate of 0.1 and a batch size of 3 for training the U-Net in each experimental setup.

To assess the impact of image resizing on OD and OC segmentation, the original U-Net architecture [50] was implemented under multiple experimental setups:

- U-Net was trained on the DRISHTI-GS training dataset and evaluated separately on the DRISHTI-GS and RIM-ONE test sets.
- U-Net was trained using the RIM-ONE training dataset and tested independently on the RIM-ONE and DRISHTI-GS test sets.
- A mixed training dataset was created by merging all available datasets, and the trained model was evaluated separately on the DRISHTI-GS and RIM-ONE test sets.

The decision to use a mixed training dataset was motivated by the observation that DL models are often trained and validated on the same dataset (see Table 3.1). This practice can lead to overfitting, where the model becomes highly specialised in recognising a specific type of image but struggles to perform well on new data. By incorporating a mixed dataset, the model is exposed to a diverse range of images, enhancing its ability to apply its learning across different datasets and improving its applicability to real-world scenarios.

Table 3.1: OD and OC segmentation results by Dice across various methods.

Other methods	**Initial ROI size**	**ROI size after resizing**	**Dataset**	**OD**	**OC**
LU-Net [37]	448×448	-	REFUGE	0.9822	-
			DRISHTI-GS	0.9974	-
DDSCNet [35]	480×480	240×240	REFUGE	0.9600	0.8900
DDSCNet [35]	560×560	240×240	DRISHTI-GS	0.9780	0.9120
UNet [52]	256×256	128×128	RIM-ONE	0.9500	0.8200
GDCSeg-Net [82]	512×512	-	REFUGE	0.9640	0.8940
			DRISHTI-GS	0.9740	0.9000
			RIM-ONE	0.9560	0.8240
CNN [69]	128×128	-	DRISHTI-GS	0.9880	0.9710
RFC-Net [21]	400×400 – 900×900	512×512	DRISHTI-GS	0.9790	0.9060

The results of OD and OC segmentation, evaluated by Dice, are summarised in Table 3.1 for existing methods, and in Table 3.2 for our experiments with U-Net trained on images resized to 512×512, 256×256, and 128×128 px from DRISHTI-GS, RIM-ONE, and the mixed dataset.

Table 3.2: Dice for OD and OC segmentation on images resized to 512×512, 256×256, and 128×128 px from DRISHTI-GS, RIM-ONE, and the mixed dataset.

	512×512 px		**256×256 px**		**128×128 px**	
	OD	**OC**	**OD**	**OC**	**OD**	**OC**
Train and Test Datasets						
DRISHTI-GS vs DRISHTI-GS	0.976	0.905	0.966	0.877	0.935	0.859
DRISHTI-GS vs RIM-ONE	0.825	0.607	0.774	0.571	0.713	0.537
RIM-ONE vs RIM-ONE	0.966	0.907	0.956	0.877	0.921	0.826
RIM-ONE vs DRISHTI-GS	0.824	0.593	0.772	0.566	0.711	0.522
Mixed vs DRISHTI-GS	**0.953**	**0.899**	**0.939**	**0.845**	**0.912**	**0.801**
Mixed vs RIM-ONE	0.934	0.877	0.909	0.830	0.888	0.794

These experimental results, presented in [C.1] and [C.2] as well, indicate that higher image resolution leads to improved OD and OC segmentation. When the U-Net was tested on images from the same dataset used for training, the segmentation results were quite high. However, performance decreased when the U-Net was tested on images from a different dataset. In contrast, the mixed dataset approach yielded promising results. The best segmentation performance was achieved when the network was trained on a mixed dataset with images resized to 512×512 px. The applied bilinear interpolation for resizing resulted in the loss of OC boundary details. This addresses the need to explore alternative interpolation methods, such as nearest-neighbor and bicubic interpolation, to assess their impact on segmentation accuracy.

3.1.2. Interpolation Method Impact

The results of the previous experiment (see Subsection 3.1.1), where the impact of image resizing level was discussed, underscore the limitations of bilinear interpolation in maintaining OC boundaries. Although a higher resolution (512×512 px) enhanced segmentation results, the observed loss of OC boundary details suggests that bilinear interpolation may not effectively preserve spatial information.

To address this issue, this experiment investigates alternative interpolation methods, including nearest-neighbor and bicubic interpolation. Although the previous experiment showed that resizing the ROI to 512×512 px achieved the best OD and OC segmentation results, this experiment aims to fully explore other interpolation methods. Therefore, the same resizing dimensions of 128×128, 256×256, and 512×512 px are maintained. Since the mixed dataset approach yielded promising results, it is also applied in this experiment, further extended with images from the REFUGE dataset. As a result, the separate training datasets consisted of images from the DRISHTI-GS, RIM-ONE, and REFUGE datasets, with the ROIs resized to the 128×128, 256×256, and 512×512 px sizes and interpolation methods such as bilinear interpolation, nearest neighbor interpolation, and bicubic interpolation.

To enhance diversity and reduce the risk of overfitting, the number of images in each dataset was increased to 1.000 through the application of various image augmentation techniques, including rotation (from 0° to 45°), zooming by 20%, and horizontal and vertical flipping. To ensure consistency across the datasets, an equal number of images from each dataset was included into the mixed training dataset.

To broaden the application of CNNs, this experiment incorporated three modifications of U-Net Attention U-Net [54], Residual Attention U-Net (RAUNet) [45], and UNet++ [81], which have demonstrated advancements in object segmentation. These models were selected to be trained on the mixed datasets for the segmentation of OD and OC.

The training of the CNNs was conducted on a single-GPU machine [1] with 1 TB of RAM, using Keras with the TensorFlow framework. An early stopping technique was employed to minimise unnecessary training time. The Adam optimiser and binary cross-entropy loss function were used for training the three CNNs: Attention U-Net, UNet++, and RAUNet. The key hyperparameters, such as learning rate, batch size, and dropout rate, were individually optimized for each network using the KerasTuner framework. For UNet++, the training time per step (in seconds) based on image sizes of 512×512, 256×256, and 128×128 px was 250, 130, and 43, respectively. For RAUNet, the training times per step were 347, 173, and 63 seconds for the same image sizes. In the case of Attention U-Net, the training times per step were 392, 189, and 67 seconds for the image sizes 512×512, 256×256, and 128×128 px,

Table 3.3: Dice of OD and OC segmentation by different interpolation methods and image size.

CNN and Dataset	**512×512 px**		**256×256 px**		**128×128 px**	
	OD	**OC**	**OD**	**OC**	**OD**	**OC**
	BILINEAR					
UNet++						
REFUGE	0.963	0.855	0.954	0.844	0.935	0.829
RIM-ONE	0.962	0.811	0.938	0.728	0.927	0.715
DRISHTI-GS	0.964	0.855	0.899	0.849	0.944	0.824
RAUNet						
REFUGE	0.950	0.827	0.939	0.804	0.910	0.781
RIM-ONE	0.948	0.808	0.929	0.791	0.908	0.777
DRISHTI-GS	0.950	0.826	0.937	0.809	0.912	0.787
Attention U-Net						
REFUGE	0.966	0.864	0.965	0.855	0.938	0.842
RIM-ONE	0.963	0.820	0.961	0.809	0.938	0.782
DRISHTI-GS	0.967	0.860	0.961	0.859	0.947	0.845
	NEAREST NEIGHBOR					
UNet++						
REFUGE	0.961	0.858	0.949	0.848	0.934	0.822
RIM-ONE	0.961	0.836	0.937	0.769	0.913	0.659
DRISHTI-GS	0.965	0.867	0.958	0.839	0.952	0.800
RAUNet						
REFUGE	0.954	0.838	0.935	0.823	0.912	0.808
RIM-ONE	0.946	0.811	0.927	0.805	0.903	0.810
DRISHTI-GS	0.953	0.834	0.936	0.825	0.916	0.796
Attention U-Net						
REFUGE	0.968	0.865	0.966	0.852	0.942	0.848
RIM-ONE	0.963	0.839	0.962	0.825	0.939	0.798
DRISHTI-GS	0.968	0.872	0.966	0.869	0.953	0.854
	BICUBIC					
UNet++						
REFUGE	**0.964**	**0.862**	**0.957**	**0.857**	**0.944**	**0.849**
RIM-ONE	**0.965**	**0.846**	**0.951**	**0.827**	**0.936**	**0.793**
DRISHTI-GS	**0.970**	**0.873**	**0.966**	**0.859**	**0.956**	**0.836**
RAUNet						
REFUGE	**0.951**	**0.846**	**0.949**	**0.831**	**0.931**	**0.815**
RIM-ONE	**0.949**	**0.830**	**0.937**	**0.829**	**0.928**	**0.819**
DRISHTI-GS	**0.956**	**0.853**	**0.944**	**0.834**	**0.922**	**0.818**
Attention U-Net						
REFUGE	**0.973**	**0.874**	**0.969**	**0.873**	**0.950**	**0.859**
RIM-ONE	**0.977**	**0.855**	**0.967**	**0.828**	**0.952**	**0.813**
DRISHTI-GS	**0.979**	**0.877**	**0.968**	**0.870**	**0.959**	**0.862**

respectively.

The trained Attention U-Net, RAUNet, and UNet++ were tested separately on 50 test images of each dataset, meaning REFUGE, RIM-ONE, and DRISHTI-GS. The performance of OD and OC segmentation

was evaluated by Dice, which is used in most cases [26, 67, 82] to describe the similarity between the two images.

Dice for OD and OC segmentation, achieved using Attention U-Net, RAUNet, and UNet++ on mixed datasets with differently preprocessed images, is presented in Table 3.3. The results demonstrate that the most accurate segmentation was obtained when the eye fundus images were resized to 512×512 px using bicubic interpolation.

Among the CNNs, the best segmentation results presented in Table 3.4 were achieved with Attention U-Net, where the highest Dice of 0.979 for OD and 0.856 for OC were observed on the DRISHTI-GS test dataset. To evaluate the significance of these OD and OC segmentation results, statistical measures such as the mean and variance of Dice were calculated for Attention U-Net using 512×512 px resized images across three interpolation methods, such as bilinear, nearest neighbour, and bicubic. ANOVA test p-values indicate that differences in OD segmentation across interpolation methods are not statistically significant, whereas differences in OC segmentation are significant, emphasising the importance of preprocessing when segmenting the OC.

Table 3.4: Dice statistics for OD and OC segmentation on 512×512 px images using Attention U-Net.

	p-values*		**Bilinear**		**Nearest Neighbor**		**Bicubic**	
Statistics	**OD**	**OC**	**OD**	**OC**	**OD**	**OC**	**OD**	**OC**
DRISHTI-GS								
Mean	0.124	0.001	0.974	0.830	0.975	0.817	0.979	0.856
Variance	0.144	0.002	0.002	0.015	0.002	0.018	0.001	0.013
REFUGE								
Mean	0.185	0.003	0.966	0.838	0.967	0.847	0.973	0.856
Variance	0.177	0.001	0.000	0.006	0.000	0.005	0.000	0.002
RIM-ONE								
Mean	0.271	0.000	0.972	0.815	0.973	0.759	0.977	0.837
Variance	0.158	0.001	0.000	0.013	0.000	0.033	0.000	0.010

* p-values at significance level $\alpha = 0.05$ by ANOVA test for means comparison and by Levene's test for variances comparison.

Further analysis was conducted to determine whether there were significant differences in segmentation performance across different datasets and interpolation methods. Based on the p-value derived from the ANOVA and Levene's tests [85] with a significance level of $\alpha = 0.05$, there was no significant difference in Dice for OD segmentation across

test data from various datasets and image resizing methods. However, for OC segmentation, the interpolation method used for image resizing had a noticeable impact.

Since Attention U-Net, utilizing bicubic interpolation for image resizing, yielded the best results for both OD and OC segmentation, this method was selected for evaluating the statistical significance of image sizes 128×128, 256×256, and 512×512 px used in the resizing process. The analysis through ANOVA and Levene's tests revealed that the selected image size significantly impacts the segmentation outcomes for both OD and OC, as detailed in Table 3.5.

Table 3.5: Performance statistics of Attention U-Net for OD and OC segmentation using Dice on images resized to various dimensions via bicubic interpolation.

	p-value*		**512×512 px**		**256×256 px**		**128×128 px**	
Statistics	**OD**	**OC**	**OD**	**OC**	**OD**	**OC**	**OD**	**OC**
	DRISHTI-GS							
Mean	0.002	0.002	0.979	0.866	0.964	0.865	0.955	0.845
Variance	0.002	0.003	0.001	0.013	0.001	0.016	0.001	0.044
	REFUGE							
Mean	0.002	0.001	0.973	0.836	0.968	0.860	0.950	0.865
Variance	0.002	0.006	0.000	0.005	0.003	0.004	0.001	0.007
	RIM-ONE							
Mean	0.005	0.001	0.969	0.847	0.956	0.820	0.949	0.807
Variance	0.003	0.001	0.002	0.013	0.004	0.032	0.005	0.052

* p-value at significance level $\alpha = 0.05$ by ANOVA test for means comparison and by Levene's test for variances comparison.

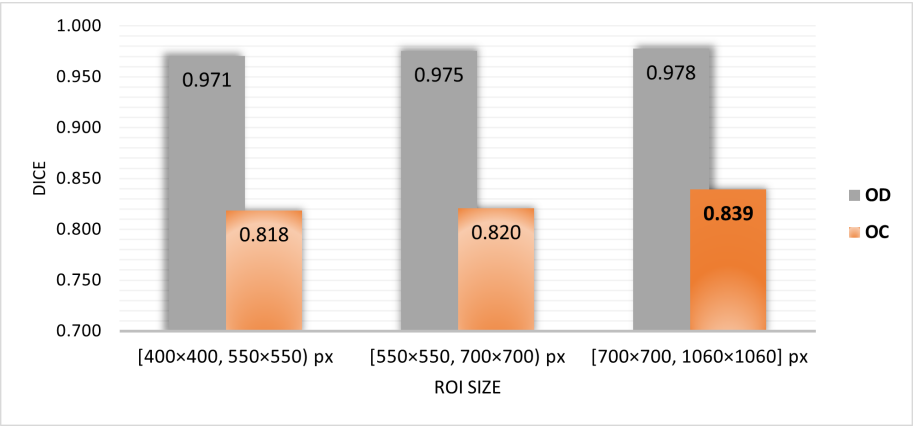


Figure 3.1: Distribution of Dice for OD and OC segmentation on images resized via bicubic interpolation from the initial ROI to 512×512 px.

The test data from the DRISHTI-GS, REFUGE, and RIM-ONE datasets were categorised into three groups based on their extracted ROI sizes: [400×400, 550×550), [550×550, 700×700), and [700×700, 1060×1060] px. These groups contained 57, 43, and 50 images, respectively. The histogram in Figure 3.1 illustrates the average Dice for OD (gray column) and OC (orange column) segmentation, using bicubic interpolation to resize the images to 512×512 px. The results indicate that, for OD segmentation, the initial ROI size has little effect on the segmentation performance. However, for OC segmentation, the resizing process is more sensitive, with better segmentation quality achieved from higher resolution images.

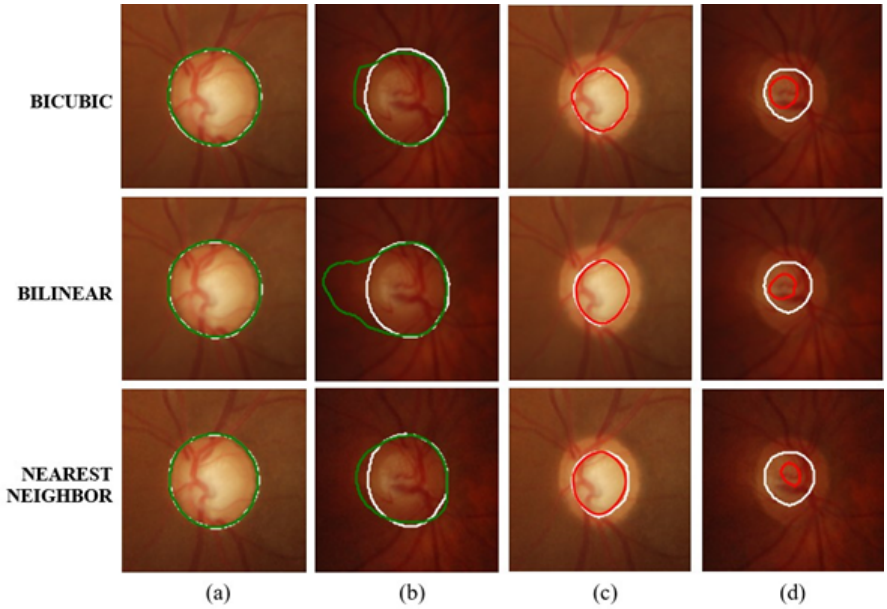


Figure 3.2: OD (green) and OC (red) segmentation results using different interpolation methods with Attention U-Net. (a) Best-case OD segmentation, (b) Worst-case OD segmentation, (c) Best-case OC segmentation, (d) Worst-case OC segmentation. The white circle indicates the ground truth.

Figure 3.2 presents visual test results from applying Attention U-Net to DRISHTI-GS dataset images resized using different interpolation methods, such as bilinear, nearest neighbor, and bicubic. In the figure, the green circle indicates the segmented OD, the red circle represents the segmented OC, and the white circle shows the ground truth. The first column (a) illustrates the best case of OD segmentation, the second column (b) displays the worst case for OD segmentation, the third column (c) shows the best case for OC segmentation, and the fourth column (d) presents the worst case for OC segmentation. While the best segmentation results for both OD and OC show no significant difference between the interpolation methods, the visual discrepancy becomes evident in the worst cases. Notably, the network struggles more with segmentation when bilinear interpolation is applied during the resizing of ROI images.

These experimental results, also presented in [A.2], demonstrate that image preprocessing significantly influences the segmentation out-

comes for OD and OC segmentation. A statistical analysis of the results revealed a notable difference in Dice for OC segmentation when images were resized using different interpolation methods, although OD segmentation remained unaffected by the interpolation technique applied.

Additionally, the size used in resizing, namely 128×128, 256×256, and 512×512 px, was found to have a statistically significant impact on both OD and OC segmentation. Among the various sizes tested, the best segmentation performance was achieved by resizing images to 512×512 px and applying bicubic interpolation. However, the training on images of this size requires more computational power compared to training on images of sizes 256×256 px and 128×128 px.

In terms of CNN performance, the Attention U-Net model provided the best segmentation results when trained on images resized to 512×512 px. The highest Dice were achieved on the following datasets: 0.979 for OD and 0.877 for OC on the DRISHTI-GS test dataset, 0.973 for OD and 0.874 for OC on the REFUGE test dataset, and 0.977 for OD and 0.855 for OC on the RIM-ONE test dataset (see Table 3.1).

3.2. Deep Learning Ensemble Approach for OD and OC Segmentation in Fundus Images

While image preprocessing techniques, such as interpolation methods and resizing strategies, have demonstrated an impact on the performance of a single CNN model, further improvements in segmentation accuracy can be explored through model combination strategies. Instead of relying on a single network, an ensemble approach leveraging multiple CNN architectures can improve performance across diverse datasets, mitigating the limitations of individual models. The following section explores the implementation of DL models ensemble for OD and OC segmentation, analysing its potential to improve segmentation performance beyond what is achieved by a single model. The fundamental idea behind ensemble learning is that different models may perform well in different areas, and by combining their predictions, the overall system achieves greater accuracy. The final prediction is obtained by aggregating the outputs through various voting techniques:

- Majority voting (MV) [51],
- Average voting (AV) [42],

- Weighted average voting (WAV) [32],
- Unanimous voting (UV) [70],
- Max voting.

The experimental research aims to assess the effectiveness of a DL ensemble method for segmenting OD and OC in eye fundus images. By integrating multiple DL models with different architectures, this approach leverages their combined strengths to improve segmentation accuracy, which is crucial for glaucoma diagnosis. The schema of the proposed ensemble method is illustrated in Figure 3.3.

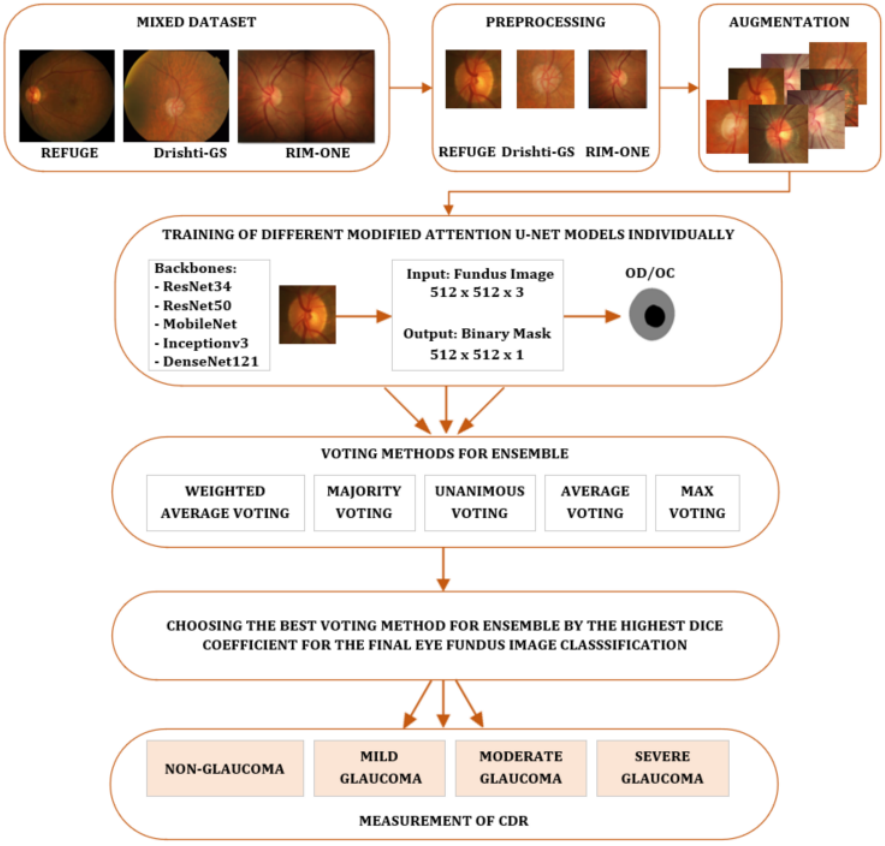


Figure 3.3: Flowchart of the proposed DL ensemble method.

The eye fundus images from the REFUGE, RIM-ONE, and DRISHTI-GS datasets were preprocessed following the same steps outlined in

Section 3.1. This included extracting the ROI and normalizing pixel values between 0 and 1. The extracted ROIs were then resized to 512×512 pixels using bicubic interpolation, as this method had previously demonstrated the best segmentation performance in the interpolation impact evaluation discussed in Subsection 3.1.2. The same image augmentation techniques were applied, including rotation (0° to 45°), zooming (20%), and horizontal and vertical flipping.

To train and validate the models, preprocessed images from the three datasets were combined into a mixed dataset. The original Attention U-Net, along with its modified versions using different pretrained networks ResNet34, ResNet50, MobileNet, Inceptionv3, and DenseNet121 were employed as backbone architectures. These models were designated as Model-1, Model-2, Model-3, Model-4, and Model-5, respectively. The training was conducted on a single GPU machine [1] with 1 TB of RAM using the Keras and TensorFlow frameworks. The Adam optimizer and Dice loss function were utilized, along with an early stopping technique to minimize unnecessary training time. Hyperparameters such as batch size, dropout rate, and learning rate were optimized separately for each model using the KerasTuner framework.

All trained CNNs were evaluated on separate test datasets from REFUGE, DRISHTI-GS, and RIM-ONE, each consisting of 50 eye fundus images representing different glaucoma stages and non-glaucoma cases. Performance metrics, including Dice and IoU, were used for evaluation. For the final OD and OC segmentation predictions, five different voting strategies, such as majority voting, weighted average voting, unanimous voting, max voting, and average voting, were applied.

Additionally, the experiment was extended to classify eye fundus images into non-glaucoma, mild glaucoma, moderate glaucoma, and severe glaucoma cases based on the calculated VCDR. CDR was computed using OD and OC segmentation outputs obtained from the ensemble method and compared against the ground truth VCDR values for each glaucoma stage and non-glaucoma cases.

OD and OC segmentation results, evaluated using Dice and IoU metrics, are summarised in Table 3.6. Among all models, Model-1, Model-4, and Model-5 demonstrated the highest segmentation performance based on these metrics and were selected to form the ensemble model. The final OD and OC segmentation predictions generated by the

Table 3.6: Segmentation results of OD and OC using different models.

Model	Training time / epoch	Test Dataset	OD		OC	
			Dice	IoU	Dice	IoU
Attention U-Net (original)	1 s	REFUGE	0.944	0.895	0.864	0.761
		DRISHTI-GS	0.961	0.925	0.897	0.814
		RIM-ONE	0.912	0.838	0.815	0.688
Model-1 (ResNet34)	370 ms	REFUGE	**0.958**	**0.919**	**0.887**	**0.797**
		DRISHTI-GS	**0.972**	**0.945**	**0.918**	**0.848**
		RIM-ONE	**0.972**	**0.946**	**0.893**	**0.807**
Model-2 (ResNet50)	396 ms	REFUGE	0.953	0.911	0.886	0.795
		DRISHTI-GS	0.972	0.946	0.915	0.843
		RIM-ONE	0.957	0.917	0.891	0.804
Model-3 (MobileNet)	370 ms	REFUGE	0.947	0.899	0.883	0.791
		DRISHTI-GS	0.964	0.931	0.905	0.826
		RIM-ONE	0.969	0.940	0.882	0.789
Model-4 (Inceptionv3)	470 ms	REFUGE	**0.952**	**0.909**	**0.883**	**0.791**
		DRISHTI-GS	**0.970**	**0.943**	**0.912**	**0.839**
		RIM-ONE	**0.974**	**0.950**	**0.892**	**0.805**
Model-5 (DenseNet121)	404 ms	REFUGE	**0.958**	**0.920**	**0.889**	**0.800**
		DRISHTI-GS	**0.973**	**0.948**	**0.912**	**0.838**
		RIM-ONE	**0.974**	**0.949**	**0.898**	**0.815**

ensemble model using five different voting methods majority voting, weighted average voting, unanimous voting, max voting, and average voting were evaluated separately on the REFUGE, DRISHTI-GS, and RIM-ONE test datasets.

The results, summarised in Table 3.7, reveal that the majority voting method achieved the highest Dice and IoU, making it the most effective approach for determining the final segmentation outcome.

Table 3.8 summarises OD and OC segmentation performance in terms of Dice and IoU for three models: the original Attention U-Net, the modified Attention U-Net with a pre-trained DenseNet121 backbone, and the proposed ensemble model utilising the majority voting method. Compared to the single original Attention U-Net, the ensemble approach improved OD segmentation by 2%, 2%, and 7% on REFUGE, DRISHTI-GS, and RIM-ONE, respectively. Similarly, OC segmentation had an increase of 3%, 2%, and 9% on the same datasets. When compared to

Table 3.7: Performance of five voting methods on test instance.

Voting technique	Test Dataset	OD		OC	
		Dice	**IoU**	**Dice**	**IoU**
Weighted average	REFUGE	0.956	0.916	0.886	0.795
	DRISHTI-GS	0.972	0.945	0.915	0.843
	RIM-ONE	0.975	0.950	0.895	0.810
Majority	REFUGE	**0.961**	**0.925**	**0.894**	**0.808**
	DRISHTI-GS	**0.974**	**0.950**	**0.916**	**0.845**
	RIM-ONE	**0.978**	**0.957**	**0.902**	**0.822**
Unanimous	REFUGE	0.952	0.908	0.877	0.781
	DRISHTI-GS	0.968	0.938	0.901	0.819
	RIM-ONE	0.973	0.948	0.886	0.796
Average	REFUGE	0.956	0.916	0.887	0.796
	DRISHTI-GS	0.972	0.945	0.914	0.842
	RIM-ONE	0.974	0.949	0.894	0.809
Max	REFUGE	0.957	0.917	0.890	0.801
	DRISHTI-GS	0.973	0.948	0.922	0.856
	RIM-ONE	0.971	0.943	0.894	0.809

Table 3.8: OD and OC segmentation performance (Dice and IoU) for three models

Model	Test Dataset	OD		OC	
		Dice	**IoU**	**Dice**	**IoU**
Attention U-Net (original)	REFUGE	0.944	0.895	0.864	0.761
	DRISHTI-GS	0.961	0.925	0.897	0.814
	RIM-ONE	0.912	0.838	0.815	0.688
Model-5	REFUGE	0.958	0.920	0.889	0.800
	DRISHTI-GS	0.973	0.948	0.912	0.838
	RIM-ONE	0.974	0.949	0.898	0.815
Proposed Ensemble (Majority voting method) [A.1]	REFUGE	**0.961**	**0.925**	**0.894**	**0.808**
	DRISHTI-GS	**0.974**	**0.950**	**0.916**	**0.845**
	RIM-ONE	**0.978**	**0.957**	**0.902**	**0.822**

the best-performing Attention U-Net variant with DenseNet121 as a backbone, the ensemble method further improved OD segmentation by 0.3%, 0.1%, and 0.4% on REFUGE, DRISHTI-GS, and RIM-ONE, respectively. OC segmentation also benefited, with Dice score increases of 0.5%, 0.4%, and 0.7% across these datasets.

OD and OC segmentation results achieved using the proposed ensemble approach were compared with those reported by other researchers (Table 3.9). The proposed ensemble method, which utilised a mixed dataset for training multiple DL models, outperformed some of the previous approaches where models were trained and tested on a

Table 3.9: Comparison of the proposed method with existing methods in OD and OC segmentation.

Method	**Train dataset**	**Test dataset**	**OD**		**OC**	
			Dice	**IoU**	**Dice**	**IoU**
CAE-BMAL [78]	REFUGE	REFUGE	0.9628	-	0.8786	-
		DRISHTI-GS	0.9620	-	0.8570	-
		RIM-ONE	0.8980	-	0.7910	-
Paired-Box RPN [79]	REFUGE ORIGA	REFUGE	0.9582	-	0.9027	-
		ORIGA	0.9634	-	0.8927	-
TUNet [34]	REFUGE DRISHTI-GS RIM-ONE	REFUGE	0.9613	-	0.9006	-
		DRISHTI-GS	0.9727	-	0.9025	-
		RIM-ONE	0.9690	-	0.8618	-
PY-Net [5]	REFUGE DRISHTI-GS RIM-ONE	REFUGE	0.9648	0.9323	0.9385	0.8847
		DRISHTI-GS	0.9710	0.9437	0.9338	0.8764
		RIM-ONE	0.9612	0.9256	0.9292	0.8737
BEAC-Net [25]	DRISHTI-GS RIM-ONE	DRISHTI-GS	0.8614	0.8385	0.8087	0.7633
		RIM-ONE	0.8582	0.8385	0.7333	0.6633
		66 Vision Tech	0.8267	0.8138	0.8057	0.7858
C2FTFNet [77]	DRISHTI-GS	DRISHTI-GS	0.9768	0.9531	0.9195	0.8538
EARDS [80]	REFUGE DRISHTI-GS	REFUGE	0.9549	0.9147	0.8872	0.8017
		DRISHTI-GS	0.9741	0.9497	0.9157	0.8493
EE-Unet [72]	REFUGE GAMMA	DRISHTI-GS	0.9624	0.8849	0.9228	0.8152
		RIM-ONE	0.9558	0.8803	0.8642	0.7619
Ensemble **Proposed** [A.1]	Mix of RIM-ONE REFUGE DRISHTI-GS	RIM-ONE	0.9610	0.9250	0.8940	0.8080
		REFUGE	0.9740	0.9500	0.9160	0.8450
		DRISHTI-GS	0.9780	0.9570	0.9020	0.8220

single dataset.

Table 3.10 displays the percentage of correctly classified outputs for segmented OD and OC using the ensemble method, categorizing test images into four distinct glaucoma stages and non-glaucoma cases based on the calculated VCDR, and compares these results with the ground truth VCDR.

Figure 3.4 presents visual samples generated using the proposed ensemble method. From left to right, the figure displays images of normal, mild glaucoma, moderate glaucoma, and severe glaucoma cases. The green and red circles highlight the segmented OD and OC, while the white circles represent the ground truth for each.

Table 3.10: Percentage of true VCDR vs. VCDR calculated from ensemble segmented OD and OC for each test dataset separately.

VCDR by glaucoma stages	**Test dataset**	**Amount of images**	**%VCDR (True vs. Calc.)**
(0–0.3] Normal	REFUGE DRISHTI-GS RIM-ONE	4 – 7	50 – 86
(0.3–0.4] Mild glaucoma	REFUGE DRISHTI-GS RIM-ONE	7 – 7	57 – 86
(0.4–0.7] Moderate glaucoma	REFUGE DRISHTI-GS RIM-ONE	34 10 23	92 90 87
Above 0.7 Severe glaucoma	REFUGE DRISHTI-GS RIM-ONE	5 40 13	100 95 92

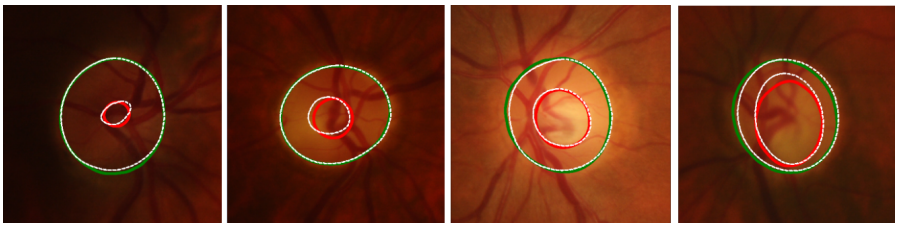


Figure 3.4: Results of the best ensemble method. Dashed circles represent ground truth OD and OC, while the green and red circles indicate segmented OD and OC, respectively. From left to right: normal, mild glaucoma, moderate glaucoma, and severe glaucoma.

The experimental results, presented in [A.1], can be summarized as follows:

- OD and OC segmentation accuracy improves using an ensemble of CNNs compared to a single model. The proposed CNN ensemble method increased OD segmentation by Dice by 2%, 2%, and 7% on REFUGE, DRISHTI-GS, and RIM-ONE datasets, respectively. Similarly, OC segmentation accuracy by Dice was increased by 3%, 2%, and 9% on these datasets.
- Among the five voting methods applied, the majority voting method provided the most accurate OD and OC segmentation results when combining predictions from three different models.

- Severe-stage glaucoma cases were classified with accuracy rates of 92%, 95%, and 100% on the RIM-ONE, DRISHTI-GS, and REFUGE datasets, respectively.
- Moderate-stage glaucoma cases were classified with accuracy rates of 87%, 90%, and 92% on the RIM-ONE, DRISHTI-GS, and REFUGE datasets.
- Mild-stage glaucoma cases were classified with accuracy rates of 86% and 57% on the RIM-ONE and REFUGE datasets. In instances of misclassification, mild-stage glaucoma cases were often mistakenly categorised as moderate-stage glaucoma.
- Non-glaucoma cases were classified with accuracy rates of 86% and 50% on the RIM-ONE and REFUGE datasets. When misclassified, non-glaucoma cases were frequently identified as mild-stage glaucoma.

3.3. Impact of Different Experts' Eye Fundus Image Annotations on CNN Learning

While VCDR was utilised in the previous experiment (Section 3.2) discussing DL ensembles for glaucoma identification, alternative measurements such as the HCDR [44] and the ACDR [12] can also be considered for further analysis. The choice of measurement is particularly important because variability in CDR reflects the inherent difficulty in accurately delineating OC boundaries, especially during the early and intermediate stages of glaucoma. This variability is further amplified during the annotation process, where differences in expert assessments can influence the consistency of labelled datasets.

Publicly available fundus image datasets, such as Chákṣu, contain images annotated by multiple experts with varying levels of expertise and diagnostic perspectives. As a result, differences in OD and OC boundary delineation can lead to inconsistencies in glaucoma identification, which is critical for training CNNs. This variability may negatively affect the performance of automated segmentation algorithms, limiting their performance across diverse populations. Figure 3.5 presents examples of glaucoma and normal cases annotated by five different experts. While the OD boundaries are generally well-defined, significant

discrepancies can be observed in the OC annotations.

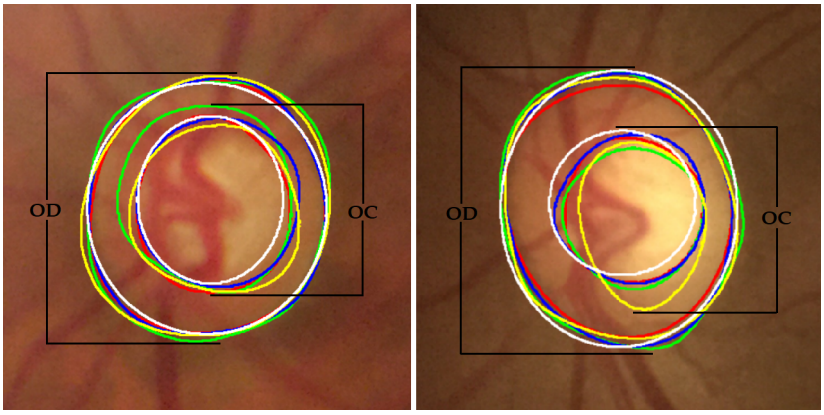


Figure 3.5: Left to right: Glaucoma and normal cases annotated by five experts. OD and OC annotations are color-coded as follows: red (Expert 1), green (Expert 2), blue (Expert 3), yellow (Expert 4), and white (Expert 5).

Therefore, this experiment aims to address several critical questions that emerged from the literature review:

- Are the assessments of different experts statistically significantly different?
- Can distinct CDR thresholds be defined for different metrics, which could then be used to develop an automated classifier for distinguishing between glaucoma and healthy eye cases?
- How to compare CNNs trained on images labelled by different experts?
- Which metric is the most appropriate for identifying glaucoma?
- What impact do labelling fundus images by different experts have on CNN training?

To address each of the questions above, the experiment was first conducted using a single CNN approach, focusing on expert variability and its impact on CNN training. Among the tested models, the modified Attention U-Net with a pre-trained DenseNet121 backbone achieved

the best segmentation performance in the experiment presented in Section 3.2. Therefore, it was selected for this study. The proposed workflow for implementing this experiment is presented in Figure 3.6.

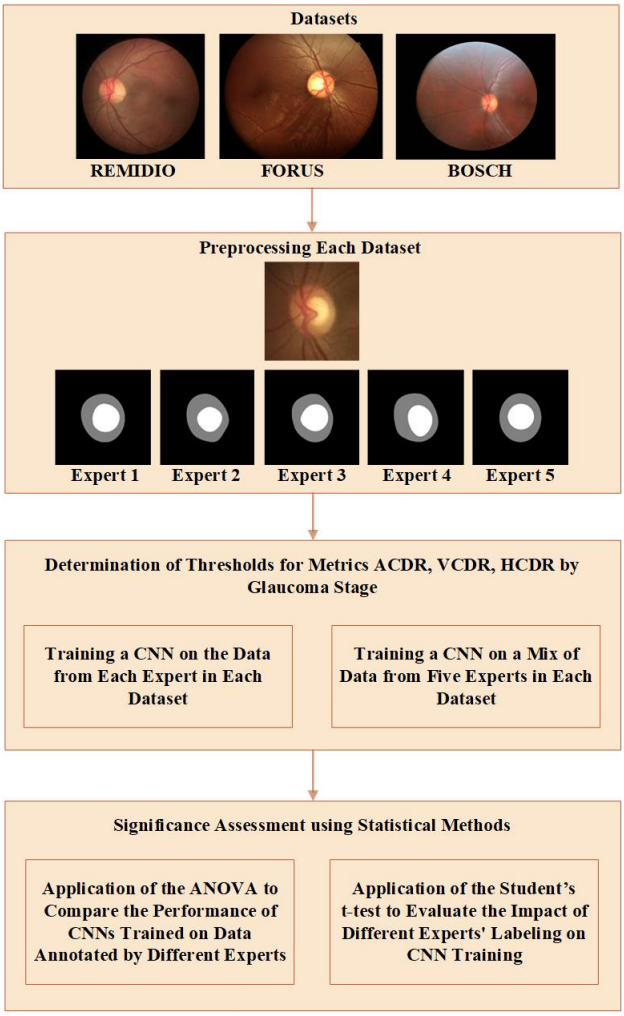


Figure 3.6: Experimental workflow illustrating the impact of different expert annotations of fundus images on CNN training

The three eye fundus image datasets REMIDIO, FORUS, and BOSCH from the Chákṣu dataset were preprocessed and used independently to train and test CNN. Ground truth masks for each dataset were generated based on annotations from five different experts. Due to the high resolution of eye fundus images in the Chákṣu dataset, ranging from 2448×3264, 2048×1536, to 1920×1440 px, the preprocessing phase

included ROI extraction, image resizing, and normalisation (scaling pixel values between 0 and 1). The ROI was determined by cropping the central OD region and resizing it to 512×512 px using bicubic interpolation, as this method had previously demonstrated optimal segmentation performance. To prevent overfitting, online image augmentation was applied, including random rotations between 0° and 45°, 20% zoom, and horizontal and vertical flips. Additionally, 20% of the images were set aside for validation. The preprocessed data were then used in two separate experiments, conducted on a single GPU machine [1] with 1 TB of RAM, utilising the Keras and TensorFlow frameworks. The experiments implemented the Adam optimiser and Dice loss, incorporating early stopping to optimise training efficiency.

The statistical significance of differences in eye fundus image labelling by various experts, as well as their impact on CNN performance for OD and OC segmentation, is assessed using the ANOVA test [85] and the Student's t-test [84]. The ANOVA test is used to examine if there are significant differences in labelling across multiple experts, while the Student's t-test is applied to compare differences between two distinct groups (glaucoma vs. normal cases). These statistical methods allow us to evaluate the effect of expert variability on CNN performance.

3.3.1. Evaluation of Statistically Significant Differences in Glaucoma Assessments Provided by Different Experts

The assessments of glaucoma and non-glaucoma cases by experts with varying levels of expertise were systematically examined to determine whether significant differences exist in their evaluations. For this analysis, we utilised eye fundus images and their corresponding ground truths from three datasets, such as REMDIDIO, FORUS, and BOSCH. Each dataset was annotated independently by multiple experts, where individual assessments classified the cases into normal and glaucoma based on their subjective evaluation of CDR.

To evaluate the variability in assessments, an independent sample Student's t-test was performed for each of the three calculated CDR metrics: ACDR, VCDR, and HCDR. The test was conducted separately for each dataset and each expert's annotations. Statistical significance was determined using a p-value at a significance level of $\alpha = 0.05$, with the results summarized in Table 3.11.

Table 3.11: p-value results of Student's t-test for each expert and metric across all datasets.

Expert	Glaucoma cases	Normal cases	ACDR p-value	VCDR p-value	HCDR p-value
			REMIDIO		
Expert 1	200	874	<0.001	<0.001	<0.001
Expert 2	347	727	<0.001	<0.001	<0.001
Expert 3	166	908	<0.001	<0.001	<0.001
Expert 4	202	872	<0.001	<0.001	<0.001
Expert 5	101	973	<0.001	<0.001	<0.001
			FORUS		
Expert 1	27	99	<0.001	<0.001	<0.001
Expert 2	35	91	<0.001	<0.001	<0.001
Expert 3	25	101	<0.001	<0.001	<0.001
Expert 4	27	99	<0.001	<0.001	<0.001
Expert 5	9	117	0.003	<0.001	<0.001
			BOSCH		
Expert 1	44	246	<0.001	<0.001	<0.001
Expert 2	114	176	<0.001	<0.001	<0.001
Expert 3	52	238	<0.001	<0.001	<0.001
Expert 4	44	246	<0.001	<0.001	<0.001
Expert 5	10	280	<0.001	<0.001	<0.001

The results indicate significant differences in the assessment of glaucoma and normal cases when evaluated at the individual expert level across all datasets. This highlights the impact of expert variability in glaucoma diagnosis and underscores the inherent subjectivity of CDR-based evaluations. For example, in the REMIDIO dataset, notable discrepancies were observed, with Expert 1 classifying 200 cases as glaucoma and 874 as normal, while Expert 5 identified only 101 glaucoma cases and 973 normal cases.

This discrepancy indicates that each expert applies their own threshold or interpretation when distinguishing between glaucoma and normal cases, leading to variability in the overall assessment. Similar trends were observed in the FORUS and BOSCH datasets, where differences in CDR thresholds and individual expertise contributed to variations in the classification outcomes.

3.3.2. Determination of Thresholds for Different Metrics: ACDR, VCDR, and HCDR

To explore the applicability of establishing threshold values for ACDR, VCDR, and HCDR that could support the development of an automated classifier for glaucoma detection, an in-depth analysis was performed on the REMIDIO, FORUS, and BOSCH datasets.

For this analysis, the eye fundus images in each dataset were assigned glaucoma or normal labels based on a majority voting approach, thereby reducing the impact of individual subjectivity in expert assessments. This method combined the annotations from multiple experts to establish a consensus ground truth classification. Once these labels were finalized, the corresponding ACDR, VCDR, and HCDR values, as determined by each expert, were examined. To visualize and analyze the variability in these metrics, boxplot charts were generated (Figure 3.7), illustrating the distribution of ACDR, VCDR, and HCDR values for glaucoma and normal cases across the REMIDIO, FORUS, and BOSCH datasets.

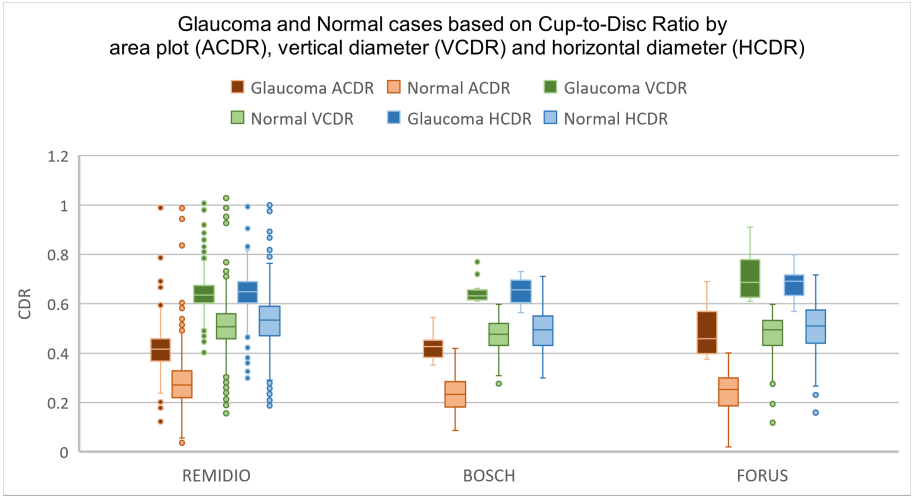


Figure 3.7: Glaucoma and normal cases classified by CDR using ACDR, VCDR, and HCDR.

Table 3.12: Mean and CI of eye health stages based on ACDR, VCDR, and HCDR for each dataset.

Stage by metric	**Statistics**	**REMIDIO**	**FORUS**	**BOSCH**
Glaucoma by ACDR	Mean $\pm$ 95% CI	0.468 ± 0.010	0.489 ± 0.037	0.421 ± 0.018
Normal by ACDR		0.247 ± 0.002	0.244 ± 0.008	0.236 ± 0.005
Glaucoma by VCDR		0.690 ± 0.010	0.716 ± 0.036	0.646 ± 0.015
Normal by VCDR		0.479 ± 0.002	0.475 ± 0.008	0.473 ± 0.005
Glaucoma by HCDR		0.677 ± 0.008	0.683 ± 0.024	0.653 ± 0.019
Normal by HCDR		0.505 ± 0.003	0.503 ± 0.010	0.490 ± 0.006

Next, the mean values and corresponding 95% confidence intervals (CIs) for ACDR, VCDR, and HCDR were computed, considering only the subset of images where all experts unanimously classified the cases as either glaucoma or normal. This approach reduced the influence of expert variability, allowing for a clearer distinction between the two groups across the datasets. The results, outlined in Table 3.12, provide the mean and CI values for ACDR, VCDR, and HCDR metrics within glaucoma and normal groups across all datasets. Key observations from this analysis are the following:

- Across all datasets, the mean values of ACDR, VCDR, and HCDR were consistently higher in glaucoma cases than in normal cases, indicating their potential as quantitative markers for the detection

of glaucoma.

- The CIs for the mean values of the glaucoma and normal groups varied across metrics within each dataset. A greater separation between these intervals indicates a stronger discriminative ability between the two groups.

According to these statistical results, threshold values for each of the three metrics ACDR, VCDR, and HCDR are suggested in Table 3.13 to be used as criteria for glaucoma evaluation. Additionally, a framework for assessing eye health that classifies cases into three distinct stages is proposed:

- **Normal**: Cases where CDR values are within the established thresholds corresponding to healthy eye conditions, indicating that OD and OC appear normal without signs of glaucomatous damage.
- **Glaucoma Suspect**: Cases where CDR values are higher than normal but have not yet reached the definitive thresholds that would confirm a diagnosis of glaucoma, individuals are categorised as being at risk. This early-stage intervention can significantly reduce the risk of disease progression and preserve vision by addressing any changes before they develop into more serious, irreversible damage.
- **Glaucoma**: Cases where CDR values exceed the established thresholds indicate a high likelihood of glaucoma.

This classification system improves diagnostic accuracy by capturing the full spectrum of glaucoma progression. It highlights the importance of identifying suspect cases, which are often overlooked but play a crucial role in early detection and intervention, ultimately helping to prevent vision loss.

Table 3.13: Suggested thresholds for evaluating eye health stages using ACDR, VCDR, and HCDR.

Stage	**Threshold by ACDR**	**Threshold by VCDR**	**Threshold by HCDR**
Normal	< 0.30	< 0.50	$\leqslant 0.50$
Glaucoma Suspect	[0.30, 0.50]	[0.50, 0.65]	(0.50, 0.65]
Glaucoma	> 0.50	> 0.65	> 0.65

3.3.3. The Method for Comparing CNNs Trained on OD and OC Labels Provided by Different Experts

Building upon the framework for glaucoma staging and assessment, the next step involves exploring the impact of expert variability on the performance of CNNs. Specifically, the focus is on the methodology for comparing CNNs trained on OD and OC labels provided by different experts. This comparison is crucial for understanding how differences in expert annotations may influence the training process and the overall accuracy of automated segmentation models.

In this experiment, the modified Attention U-Net, using DenseNet121 as its pre-trained backbone, was individually trained and tested on eye fundus images along with their respective OD and OC masks provided by each expert across the REMIDIO, FORUS, and BOSCH datasets. The model architecture integrates attention mechanisms to refine feature extraction for OD and OC segmentation while utilizing DenseNet121 to leverage deep feature representations. This approach allows the model to concentrate on clinically significant areas within the eye fundus images, leading to enhanced segmentation accuracy.

The dataset for testing was constructed by taking 25 eye fundus images of each case, glaucoma and normal, from each of the datasets. For each image, ground truth labels were independently provided by each expert to ensure a diverse range of expert opinions. This approach not only helped maintain consistency across the test sets but also introduced a layer of inter-expert variability, which is crucial for evaluating the performance of the CNN model under different labelling perspectives. CNN model's performance in segmenting OD and OC was then assessed using Dice and IoU. The detailed results of these evaluations, including the comparison between the different expert labels, are summarised in Table 3.14. This table provides insight into the model's ability to handle

Table 3.14: OD and OC segmentation results from CNNs trained on expert-specific datasets and a mixed-data strategy.

Train Dataset	Test Dataset	OD		OC	
		Dice	IoU	Dice	IoU
REMIDIO					
Expert 1	Expert 1	0.964	0.931	0.867	0.773
Expert 2	Expert 2	0.962	0.927	0.889	0.805
Expert 3	Expert 3	0.960	0.924	0.869	0.777
Expert 4	Expert 4	0.955	0.914	0.862	0.761
Expert 5	Expert 5	0.946	0.900	0.839	0.731
Mix of five experts	Expert 1	0.934	0.876	0.821	0.707
	Expert 2	0.953	0.911	0.865	0.770
	Expert 3	0.945	0.897	0.831	0.722
	Expert 4	0.952	0.909	0.853	0.749
	Expert 5	0.948	0.902	0.794	0.671
FORUS					
Expert 1	Expert 1	0.976	0.953	0.921	0.862
Expert 2	Expert 2	0.980	0.961	0.928	0.871
Expert 3	Expert 3	0.986	0.973	0.928	0.872
Expert 4	Expert 4	0.980	0.962	0.922	0.861
Expert 5	Expert 5	0.975	0.953	0.920	0.860
Mix of five experts	Expert 1	0.975	0.951	0.896	0.814
	Expert 2	0.970	0.941	0.857	0.754
	Expert 3	0.977	0.954	0.867	0.769
	Expert 4	0.969	0.939	0.855	0.753
	Expert 5	0.964	0.930	0.861	0.761
BOSCH					
Expert 1	Expert 1	0.974	0.949	0.927	0.869
Expert 2	Expert 2	0.972	0.945	0.922	0.861
Expert 3	Expert 3	0.973	0.947	0.875	0.792
Expert 4	Expert 4	0.953	0.916	0.884	0.803
Expert 5	Expert 5	0.960	0.924	0.891	0.810
Mix of five experts	Expert 1	0.950	0.905	0.868	0.771
	Expert 2	0.964	0.931	0.869	0.772
	Expert 3	0.964	0.930	0.808	0.697
	Expert 4	0.949	0.908	0.863	0.767
	Expert 5	0.952	0.909	0.812	0.693

variations in expert labelling and its overall segmentation accuracy across all test images.

The segmentation performance highlights the model's proficiency in precisely identifying OD and OC regions across different datasets and expert annotations. This step is crucial as it lays the foundation for calculating key metrics, including ACDR, VCDR, and HCDR. These metrics were calculated from OD and OC regions identified by CNNs, and the analysis focuses on how variations in expert-provided ground

truth labels affect the model's ability to detect glaucoma. This approach enables a better understanding of how expert differences impact the performance and accuracy of the model in diagnosing glaucoma.

To assess the impact of varying expert annotations on CNN performance, an ANOVA test was conducted to compare the segmentation results and the derived CDR-based metrics from models trained with expert-specific data. This statistical analysis was performed in three distinct tests:

- Test 1:Evaluating whether there is a statistically significant difference between glaucoma and normal cases.
- Test 2: Investigating whether notable differences exist between the assessments made by different experts.
- Test 3: Analyzing the interaction between the case stage, specifically glaucoma and normal, and the expert annotations.

The results, outlined in Table 3.15, were analyzed at a significance level of α = 0.05, leading to the following conclusions:

- Significant difference was observed in the classification of glaucoma and normal cases when CNNs were trained on datasets annotated by different experts. This variation underscores the need for consistency in ground truth labeling during model training and evaluation.
- The performance of CNNs trained on data annotated by different experts demonstrated a statistically significant difference. This

Table 3.15: ANOVA test p-value* results.

Test	**Dataset**	**ACDR p-value**	**VCDR p-value**	**HCDR p-value**
Test 1	REMIDIO	<0.001	<0.001	<0.001
	FORUS	<0.001	<0.001	<0.001
	BOSCH	<0.001	<0.001	<0.001
Test 2	REMIDIO	<0.001	<0.001	<0.001
	FORUS	<0.001	<0.001	<0.001
	BOSCH	<0.001	<0.001	<0.001
Test 3	REMIDIO	0.045	0.226	0.040
	FORUS	0.203	0.149	0.298
	BOSCH	<0.001	<0.001	0.001

* p-value at significance level α = 0.05 by ANOVA test.

highlights the direct influence of expert variability on segmentation results and the subsequent CDR calculations.

- The analysis found no significant interaction between the classification of glaucoma and normal stages and the annotations provided by different experts. This suggests that, although expert-specific annotations affect CNN performance, the classification of glaucoma stages is not dependent on individual labelling biases.

3.3.4. Evaluation of the Impact of Variation in Expert Labelling on CNN Training and the Selection of the Most Suitable Metric for Glaucoma Assessment

To further investigate the influence of expert labelling variability on CNNs' training process, the next step focuses on evaluating how these discrepancies affect the model's performance. This assessment is crucial in determining the most accurate metrics for accurate glaucoma detection. By analysing the effects of different expert annotations on the training outcomes, we can identify which metric, among the various CDRs, such as ACDR, VCDR, and HCDR, offers the most effective and consistent measure for glaucoma assessment.

In this experiment, the impact of variability in expert annotations on CNN training performance was evaluated, alongside the effectiveness of different CDR metrics. To mitigate inconsistencies introduced by expert variability, a mixed training dataset was constructed by randomly selecting eye fundus images with their corresponding OD and OC masks, each annotated by different experts from the REMIDIO, FORUS, and BOSCH datasets. The test dataset remained consistent with the experiment described in Section 3.1.1 to allow fair comparison.

The modified Attention U-Net, with DenseNet121 as the pre-trained backbone, was trained and validated on the mixed dataset. Subsequently, the network was tested on eye fundus images with masks annotated independently by each expert. Segmentation performance for both OD and OC was evaluated using Dice and IoU. This setup enabled an assessment of whether training on a mixed dataset improves the network's ability to perform effectively across annotations from different experts.

The segmentation results are presented in Table 3.15, alongside outcomes from the experiment in Section 3.1.1, where networks were

trained on datasets annotated by individual experts. This comparison highlights the influence of training on a mixed, diverse dataset versus single-expert annotations.

To further quantify the effect of expert variability on network performance, a Student's t-test was conducted on ACDR, VCDR, and HCDR values calculated from CNN-segmented OD and OC regions, compared with the ground truths provided by the experts. These analyses provide insight into how expert annotation variability impacts both segmentation accuracy and downstream glaucoma-related measurements. The evaluation was performed separately for each expert's test data, with the summarised results presented in Table 3.16. The key findings are as follows:

- When the network was trained on the mixed dataset and evaluated using individual expert datasets, VCDR values calculated from CNN-segmented OD and OC regions closely matched the assessments provided by all experts. This suggests that VCDR is an accurate and stable metric for glaucoma assessment, showing minimal sensitivity to variations in expert annotations.

- On the other hand, significant differences were observed between ACDR and HCDR values predicted by CNN and those provided by the experts. These discrepancies indicate that area-based and horizontal diameter-based metrics are more susceptible to expert annotation inconsistencies, limiting their effectiveness in automated glaucoma detection systems.

- Training the network on a mixed dataset enabled the model to better perform across the variations in expert labelling. It demonstrated an enhanced capacity to adapt to the diverse annotation styles and interpretations provided by experts with varying expertise. Although this approach helped reduce some inconsistencies observed when training on single-expert datasets, the improvement in segmentation performance, as measured by Dice and IoU, was not significantly higher than that of networks trained solely on individual expert datasets.

Table 3.16: p-value* results from a Student's t-test.

Expert	ACDR p-value	VCDR p-value	HCDR p-value
REMIDIO			
Mixed vs Expert 1	<0.001	0.334	<0.001
Mixed vs Expert 2	<0.001	0.067	<0.001
Mixed vs Expert 3	<0.001	0.067	<0.001
Mixed vs Expert 4	<0.001	0.100	<0.001
Mixed vs Expert 5	<0.001	0.057	<0.001
FORUS			
Mixed vs Expert 1	<0.001	0.025	<0.001
Mixed vs Expert 2	<0.001	0.040	<0.001
Mixed vs Expert 3	<0.001	0.136	<0.001
Mixed vs Expert 4	<0.001	0.029	<0.001
Mixed vs Expert 5	<0.001	0.018	<0.001
BOSCH			
Mixed vs Expert 1	<0.001	0.150	<0.001
Mixed vs Expert 2	<0.001	0.307	<0.001
Mixed vs Expert 3	<0.001	0.159	<0.001
Mixed vs Expert 4	<0.001	0.058	<0.001
Mixed vs Expert 5	<0.001	0.060	<0.001

* p-value at significance level α = 0.05 by Student's t-test.

3.4. Conclusions of the Chapter

Based on the addressed key research issues, the following conclusions were drawn:

- Statistical analysis using a Student's t-test (α = 0.05) demonstrated significant differences in glaucoma and normal case assessments across all datasets due to variability in expert expertise. Each expert uniquely distinguished between glaucoma and normal cases, confirming the subjective evaluation of CDR.
- Analysis of dataset-specific mean CDR values and their CIs for ACDR, VCDR, and HCDR enabled the proposal of threshold limits to classify eye health into three categories:
  - Normal: ACDR < 0.30, VCDR < 0.50, and HCDR $\leqslant 0.50$
  - Glaucoma suspect: 0.30 $\leqslant$ ACDR $\leqslant$ 0.50, 0.50 $\leqslant$ VCDR $\leqslant$ 0.65, and 0.50 < HCDR $\leqslant$ 0.65
  - Glaucoma: ACDR > 0.50, VCDR > 0.65, and HCDR > 0.65

Introducing a "glaucoma suspect" category supports the early identification of at-risk cases, enhancing diagnostic precision and enabling timely intervention.

- ANOVA testing demonstrated statistically significant differences in the performance of CNNs trained on OD and OC labels provided by different experts. However, no interaction was observed between the glaucoma and normal stages and expert assessments, indicating that the glaucoma stage itself is largely independent of specific expert annotations. A Student's t-test comparing CDR metrics (ACDR, VCDR, HCDR) derived from CNN segmentations and ground truths revealed that networks trained on mixed expert datasets reflected overall assessments well, particularly for VCDR metric. However, discrepancies persisted for ACDR and HCDR, underscoring the continued influence of expert variability on network training outcomes and the need for further refinement of segmentation models.

- Among the evaluated CDR metrics, VCDR emerged as the most accurate for identifying glaucoma. It exhibited greater consistency across datasets and was better reflected in the network outputs compared to ACDR and HCDR.

These findings highlight the need for a standardised labelling method and the cautious selection of metrics for training automated classifiers. By addressing expert variability and proposing thresholds for key metrics ACD, VCDR, and HCDR, this research contributes significantly to the development of more accurate and consistent automated glaucoma detection systems.

This thesis proposes a multi-level methodology for glaucoma identification using DL algorithms. The following conclusions summarise the key findings and contributions of this thesis:

- Different image resolutions and interpolation methods affect the segmentation results. The best results were obtained when resizing images to 512×512 px using bicubic interpolation.
- The proposed CNN ensemble method enhances segmentation accuracy across multiple datasets. Specifically, it improves Dice for OD segmentation by an average of 4% and OC segmentation by an average of 5%. Among the five voting methods applied, the majority voting approach yielded the most precise OD and OC segmentation results by effectively combining predictions from three different models.
- Statistical analysis using the Student's t-test (α = 0.05) confirmed significant differences in glaucoma and normal case assessments across most datasets due to expert subjectivity. Each expert demonstrated a unique pattern in distinguishing glaucoma cases, reinforcing the need for standardised evaluation criteria.
- Based on dataset-specific mean values and confidence intervals among different CDR metrics, threshold limits were proposed for classifying eye health into three categories:
  - Normal: ACDR < 0.30, VCDR < 0.50, and HCDR $\leqslant 0.50$
  - Glaucoma suspect: 0.30 $\leqslant$ ACDR $\leqslant$ 0.50, 0.50 $\leqslant$ VCDR $\leqslant$ 0.65, and 0.50 < HCDR $\leqslant$ 0.65
  - Glaucoma: ACDR > 0.50, VCDR > 0.65, and HCDR > 0.65

  The introduction of a "glaucoma suspect" category aids in early detection, enabling timely intervention.
- ANOVA testing revealed statistically significant differences in CNN segmentation results when trained on OD and OC labels from different experts. However, the glaucoma stage itself was found to be largely independent of specific expert annotations.

CNNs trained on mixed expert datasets demonstrated strong agreement with VCDR-based assessments but showed inconsistencies with ACDR and HCDR, highlighting the need for further refinement in segmentation models.

- Among the evaluated CDR metrics, VCDR emerged as the most reliable for glaucoma detection, exhibiting better generalisation across datasets and alignment with CNN segmentation results than ACDR or HCDR.

[1] VU ITPC, Information Technology Service Center of Vilnius University. Available at https://mif.vu.lt/lt3/en/about/structure/it-research-center.

[2] Cr-2 af digital non-myriatic retinal camera. Available at https://genop.co.za/content/canon-cr-2-af_5c2ce7bf35c7b.pdf, 2024.

[3] M. Y. S. Ali, M. Abdel-Nasser, M. Jabreel, A. Valls, and M. Baget. *Segmenting the Optic Disc Using a Deep Learning Ensemble Model Based on OWA Operators*. IOS Press, 2021. doi: 10.3233/FAIA210149.

[4] J. Bernatavičienė, G. Dzemyda, O. Kurasova, V. Barzdžiukas, D. Buteikienė, and A. Paunksnis. Rule induction for ophthalmological data classification. In *20th EURO Mini Conference "Continuous Optimization and Knowledge-Based Technologies" (EurOPT-2008)*, pages 328–334, 2008.

[5] R. Bhattacharya, R. Hussain, A. Chatterjee, D. Paul, S. Chatterjee, and D. Dey. Py-net: Rethinking segmentation frameworks with dense pyramidal operations for optic disc and cup segmentation from retinal fundus images. *Biomedical Signal Processing and Control*, 85:104895, 2023. doi: 10.1016/j.bspc.2023.104895.

[6] A. Bhuiyan, A. Govindaiah, and R. T. Smith. An artificial-intelligence- and telemedicine-based screening tool to identify glaucoma suspects from color fundus imaging. *Journal of Ophthalmology*, 2021. doi: 10.1155/2021/6694784.

[7] D. Buteikienė, A. Paunksnis, V. Barzdžiukas, J. Bernatavičienė, V. Marcinkevičius, and P. Treigys. Assessment of the optic nerve disc and excavation parameters of interactive and automated parameterization methods. *Informatica*, 23(3):335–355, 2012. doi: 10.15388/Informatica.2012.363.

[8] Carl Zeiss Meditec AG. Visucam 500. Available at http://www.knsentp.com/userfiles/file/Eye-department/New-Fundus-Camera-Visucam-500.pdf, 2010.

[9] C. Celik, Yücedağ, and H. T. Akçam. Automated retinal image analysis to detect optic nerve hypoplasia. *Information Technology and Control*, 53(2):522–541, 2024. doi: 10.5755/j01.itc.53.2.35152.

[10] J. Civit-Masot, M. J. Domínguez-Morales, S. Vicente-Díaz, and A. Civit. Dual machine-learning system to aid glaucoma diagnosis

using disc and cup feature extraction. *IEEE Access*, 8:127519–127529, 2020. doi: 10.1109/ACCESS.2020.3008539.

[11] F. Daud, R. Imtiaz, H. A. Madni, H. A. Khan, T. M. Khan, M. A. U. Khan, and S. S. Naqvi. A review on glaucoma disease detection using computerized techniques. *IEEE Access*, 9:37311–37333, 2021. doi: 10.1109/ACCESS.2021.3061451.

[12] A. Diaz, S. Morales, V. Naranjo, P. Alcocer, and A. Lanzagorta. Glaucoma diagnosis by means of optic cup feature analysis in color fundus images. pages 2055–2059. IEEE, 2016. doi: 10.1109/EUSIPCO.2016.7760610.

[13] K. Dutta, R. Mukherjee, S. Kundu, T. Biswas, and A. Sen. Automatic evaluation and predictive analysis of optic nerve head for the detection of glaucoma. *IEEE*, pages 1–7, 2018. doi: 10.1109/IEMENTECH.2018.8465169.

[14] P. Elangovan and M. K. Nath. En-convnet: A novel approach for glaucoma detection from color fundus images using ensemble of deep convolutional neural networks. *International Journal of Imaging Systems and Technology*, 32(6):2034–2048, 2022. doi: 10.1002/ima.22761.

[15] R. B. Engineering and B. Solutions. Bosch eye cam features technical specification. Available at `https://5.imimg.com/data5/SELLER/Doc/2024/1/382074543/BC/ZL/CF/25356323/hand-held-fundus-camera.pdf`, 2017.

[16] F. Francisco, T. Diaz-Aleman, J. Sigut, S. Alayon, R. Arnay, and D. Angel-Pereira. Rim-one dl: A unified retinal image database for assessing glaucoma using deep learning. *Image Anal Stereology*, 39 (3), 2020. doi: 10.5566/ias.2346.

[17] H. Fu, J. Cheng, Y. Xu, C. Zhang, D. W. K. Wong, J. Liu, and X. Cao. Disc-aware ensemble network for glaucoma screening from fundus image. *IEEE Trans Med Imaging*, 37(11):2493–2501, 2018. doi: 10.1109/TMI.2018.2837012.

[18] F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez. Rim-one: An open retinal image database for optic nerve evaluation. pages 1–6. IEEE Xplore, 2011. doi: 10.1109/CBMS.2011.5999143.

[19] F. Fumero, J. Sigut, M. Alayón, Silvia andGonzález-Hernández, and

M. González de la Rosa. Interactive tool and database for optic disc and cup segmentation of stereo and monocular retinal fundus images. Available at https://www.idiap.ch/software/bob/docs/bob/bob.db.rimoner3/stable/index.html, 2015.

[20] R. Gaidelytė and M. Garbuvienė. Health statistics of lithuania 2023. Available at https://www.hi.lt/uploads/Statistikos_leidiniai_Sveikatos_statistika/la2023.pdf, 2024.

[21] J. Gao, Y. Jiang, H. Zhang, and F. Wang. Joint disc and cup segmentation based on recurrent fully convolutional network. *PLoS ONE*, 15, 2020. doi: 10.1371/journal.pone.0238983.

[22] D. Han. Comparison of commonly used image interpolation methods. In *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*. Atlantis Press, 2013. doi: 10.2991/iccsee.2013.391.

[23] B. Harvey. Review: The kowa nonmyd wx 3d. Available at https://www.opticianonline.net/content/features/review-the-kowa-nonmyd-wx-3d/, 2017.

[24] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *arXiv preprint arXiv:1512.03385*. arxiv, 2015. doi: 10.48550/arXiv.1512.03385.

[25] L. Jiang, X. Tang, S. You, S. Liu, and Y. Ji. BEAC-Net: Boundary-enhanced adaptive context network for optic disk and optic cup segmentation. *Applied Sciences*, 13(18):10244, 2023. doi: 10.3390/app131810244.

[26] Y. Jiang, F. Wang, J. Gao, and S. Cao. Multi-path recurrent U-Net segmentation of retinal fundus image. *Applied Sciences*, 10(11):3777, 2020. doi: 10.3390/app10113777.

[27] M. Juneja, S. Singh, N. Agarwal, S. Bali, S. Gupta, N. Thakur, and P. Jindal. Automated detection of glaucoma using deep learning convolution network (g-net). *Multimedia Tools and Applications*, 79: 15531–15553, 2020. doi: 10.1007/s11042-019-7460-4.

[28] M. Juneja, S. Thakur, A. Wani, A. Uniyal, N. Thakur, and P. Jindal. Dc-gnet for detection of glaucoma in retinal fundus imaging. *Machine Vision and Applications*, 31(34), 2020. doi: 10.1007/s00138-020-01085-2.

[29] J. Kim, L. Tran, T. Peto, and E. Y. Chew. Deep learning and ensemble method for optic disc and cup segmentation. In *2022 IEEE*

Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), pages 1–8, 2022. doi: 10.1109/CIBCB55180.2022.9863022.

[30] J. R. H. Kumar, C. S. Seelamantula, J. H. Gagan, Y. S. Kamath, N. I. R. Kuzhuppilly, U. Vivekanand, P. Gupta, and S. Patil. Chákṣu: A glaucoma specific fundus image database. *Scientific Data*, 10(1), 2023. doi: 10.1038/s41597-023-01943-4.

[31] V. Kurilová, S. Rajcsányi, Z. Rábeková, J. Pavlovičová, M. Oravec, and N. Majtánová. Detecting glaucoma from fundus images using ensemble learning. *Journal of Electrical Engineering*, 74(4):328–335, 2023. doi: 10.2478/jee-2023-0040.

[32] G. Latif-Shabgahi. A novel algorithm for weighted average voting used in fault tolerant computing systems. *Microprocessors and Microsystems*, 28(7):357–361, 2004. doi: 10.1016/j.micpro.2004.02.006.

[33] F. Li, W. Xiang, L. Zhang, W. Pan, X. Zhang, M. Jiang, and H. Zou. Joint optic disk and cup segmentation for glaucoma screening using a region-based deep learning network. *Eye*, 37:1080—-1087, 2023. doi: 10.1038/s41433-022-02055-w.

[34] Z. Li, C. Zhao, Z. Han, and C. Hong. Tunet and domain adaptation based learning for joint optic disc and cup segmentation. *Computers in Biology and Medicine*, 163:107209, 2023. doi: 10.1016/j.compbiomed.2023.107209.

[35] B. Liu, D. Pan, and H. Song. Joint optic disc and cup segmentation based on densely connected depthwise separable convolution deep network. *BMC Medical Imaging*, 21(14), 2021. doi: 10.1186/s12880-020-00528-6.

[36] P. Liu and R. Fang. Regression and learning with pixel-wise attention for retinal fundus glaucoma segmentation and detection. In *arXiv preprint arXiv:2001.01815*. arxiv, 2020. doi: 10.48550/arXiv.2001.01815.

[37] W. Liu, H. Lei, H. Xie, B. Zhao, G. Yue, and B. Lei. Multi-level light U-Net and atrous spatial pyramid pooling for optic disc segmentation on fundus image. *Ophthalmic Medical Image Analysis*, pages 104–113, 2020. doi: 10.1007/978-3-030-63419-3_11.

[38] S. Lu, H. Zhao, H. Liu, H. Li, and N. Wang. Pkrt-net: Prior knowledge-based relation transformer network for optic cup and disc segmentation. *Neurocomputing*, 538:126183, 2023. doi:

10.1016/j.neucom.2023.03.044.

[39] S. Mallick, J. Paul, and J. Sil. Response fusion attention u-convnext for accurate segmentation of optic disc and optic cup. *Neurocomputing*, 559:126798, 2023. doi: 10.1016/j.neucom.2023.126798.

[40] Medzell. 3nethra classic 6.4 digital non-mydriatic fundus camera. Available at `https://www.medzell.net/product/3nethra-classic-64-retinal-camera-from-forus/`, 2024.

[41] P. Mishra, U. Singh, C. M. Pandey, P. Mishra, and G. Pandey. Application of student's t-test, analysis of variance, and covariance. *Annals of Cardiac Anaesthesia*, 22(4):407–411, 2019. doi: 10.4103/aca.ACA_94_19.

[42] J. M. Montgomery, F. M. Hollenbach, and M. D. Ward. Improving predictions using ensemble Bayesian model averaging. *Political Analysis*, 20(3):271—291, 2012. doi: 10.1093/pan/mps002.

[43] T. Nazir, A. Irtaza, and V. Starovoitov. Optic disc and optic cup segmentation for glaucoma detection from blur retinal images using improved Mask-RCNN. *International Journal of Optics*, 2021. doi: 10.1155/2021/6641980.

[44] A. Neto, J. Camera, S. Oliveira, A. Cláudia, and A. Cunha. Optic disc and cup segmentations for glaucoma assessment using cup-to-disc ratio. *Procedia Computer Science*, 196:485–492, 2022. doi: 10.1016/j.procs.2021.12.040.

[45] Z.-L. Ni, G.-B. Bian, X.-H. Zhou, Z.-G. Hou, X.-L. Xie, C. Wang, Y.-J. Zhou, R.-Q. Li, and Z. Li. Raunet: Residual attention U-Net for semantic segmentation of cataract surgical instruments. In *arXiv preprint arXiv:1909.10360*. arXiv, 2019. doi: 10.48550/arXiv.1909.10360.

[46] OphthalmologyWeb. Product detail: Kowa nonmyd wx-3d simultaneous stereoscopic retinal camera from kowa optimed, inc. Available at `https://www.ophthalmologyweb.com/5740-Digital-Retinal-Camera/55601-Kowa-nonmyd-WX-3D-Retinal-Camera/`, 2025.

[47] J. I. Orlando, H. Fu, J. Barbosa Breda, K. van Keer, D. R. Bathula, A. Diaz-Pinto, R. Fang, P.-A. Heng, J. Kim, J. Lee, J. Lee, X. Li, P. Liu, S. Lu, B. Murugesan, V. Naranjo, S. Samarth R. Phaye, S. M. Shankaranarayana, A. Sikka, J. Son, A. van den Hengel, S. Wang,

J. Wu, Z. Wu, G. Xu, Y. Xu, P. Yin, F. Li, X. Zhang, Y. Xu, and H. Bogunović. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Medical Image Analysis*, 59:101570, 2020. doi: 10.1016/j.media.2019.101570.

[48] V. Raudonis, A. Kairys, R. Verkauskiene, J. Sokolovska, G. Petrovski, V. J. Balciuniene, and V. Volke. Automatic detection of microaneurysms in fundus images using an ensemble-based segmentation method. *Sensors*, 23(7), 2023. doi: 10.3390/s23073431.

[49] Remidio. Hand held fundus camera fop nm-10. Available at `https://www.remidio.com/en-in/products/fop`, 2024.

[50] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *arXiv preprint arXiv:1505.04597*. arxiv, 2015. doi: 10.48550/arXiv.1505.04597.

[51] S. Sabzi, R. Pourdarbani, D. Kalantari, and T. Panagopoulos. Designing a fruit identification algorithm in orchard conditions to develop robots using video processing and majority voting based on hybrid artificial neural network. *Applied Sciences*, 10(1):383, 2020. doi: 10.3390/app10010383.

[52] A. Sevastopolsky. Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. *Pattern Recognition and Image Analysis*, 27(3):618—624, 2017. doi: 10.1134/S1054661817030269.

[53] S. Sharma and K. Guleria. A deep learning model for early prediction of pneumonia using VGG19 and neural networks. In *3rd International Conference on Mobile Radio Communications and 5G Networks, (MRCN-2022)*, 2023. doi: 10.1007/978-981-19-7982-8_50.

[54] T. Shyamalee and D. Meedeniya. Attention U-Net for glaucoma identification using fundus image segmentation. In *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, 2022. doi: 10.1109/DASA54658.2022.9765303.

[55] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *arXiv preprint arXiv:1409.1556*. arxiv, 2015. doi: 10.48550/arXiv.1409.1556.

[56] L. K. Singh, M. Khanna, and Pooja. A novel multimodality based dual fusion integrated approach for efficient and early prediction of glaucoma. *Biomedical Signal Processing and Control*, 73:103468,

2022. doi: 10.1016/j.bspc.2021.103468.

[57] J. Sivaswamy, S. R. Krishnadas, G. Datt Joshi, M. Jain, and A. U. Syed Tabish. Drishti-gs: Retinal image dataset for optic nerve head(onh) segmentation. pages 53–56. IEEE, 2014. doi: 10.1109/ISBI.2014.6867807.

[58] G. Stabingis, J. Bernatavičienė, G. Dzemyda, A. Paunksnis, P. Treigys, R. Vaičaitienė, and L. Stabingienė. Automatization of eye fundus vessel width measurements. volume 27, pages 787—796. Springer, 2017. doi: 10.1007/978-3-319-68195-5_85.

[59] J. D. Steinmetz and et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to vision 2020: the right to sight: an analysis for the global burden of disease study. *The Lancet Global Health*, 9 (2):144–160, 2021.

[60] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. doi: 10.1109/CVPR.2016.308.

[61] M. Tabassum, T. M. Khan, M. Arsalan, S. S. Naqvi, M. Ahmed, H. A. Madni, and J. Mirza. Cded-net: Joint segmentation of optic disc and optic cup for glaucoma screening. *IEEE Access*, 8:102733–102747, 2020. doi: 10.1109/ACCESS.2020.2998635.

[62] S. Tadisetty, R. Chodavarapu, R. Jin, R. J. Clements, and M. Yu. Identifying the edges of the optic cup and the optic disc in glaucoma patients by segmentation. *Sensors*, 23(10):4668, 2023. doi: 10.3390/s23104668.

[63] Y.-C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C.-Y. Cheng. Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis. *Ophthalmology*, 121(11):2081–2090, 2014. doi: 10.1016/j.ophtha.2014.05.013.

[64] P. Treigys, G. Dzemyda, and V. Barzdziukas. Automated positioning of overlapping eye fundus images. In *8th International Conference on Computational Science*, page 770. COMPUTATIONAL SCIENCE - ICCS 2008, PT, 2008.

[65] P. Treigys, V. Šaltenis, and G. Dzemyda. Automated optic nerve disc parameterization. *Informatica*, 19(3):403–420, 2008. doi: 10.

15388/Informatica.2008.221.

[66] P. Treigys and V. Šaltenis. Neural network as an ophthalmologic disease classifier. *Information Technology and Control*, 36(4):365–371, 2007.

[67] A. Tulsani, P. Kumar, and S. Pathan. Automated segmentation of optic disc and optic cup for glaucoma assessment using improved unet++ architecture. *Biocybernetics and Biomedical Engineering*, 41(2): 819–832, 2021. doi: 10.1016/j.bbe.2021.05.011.

[68] A. ur Rehman, A. T. Imtiaz, M. Sajid, and K. S. Karimov. An ensemble framework based on deep cnns architecture for glaucoma classification using fundus photography. *Mathematical Biosciences and Engineering*, 18(5):5321–5346, 2021. doi: 10.3934/mbe.2021270.

[69] H. Veena, A. Muruganandham, and T. Senthil Kumaran. A novel optic disc and optic cup segmentation technique to diagnose glaucoma using deep learning convolutional neural network over retinal fundus images. *Journal of King Saud University - Computer and Information Sciences*, 34(8, Part B):6187–6198, 2022. doi: 10.1016/j.jksuci.2021.02.003.

[70] G. V. Vlasceanu, R.-M. Paraschiv, C. Artene, and C.-A. oiangiu. Voting-based edge detection. pages 1–5. IEEE, 2019. doi: 10.1109/ROEDUNET.2019.8909664.

[71] J. Wang and B. Xia. CDRNet: accurate cup-to-disc ratio measurement with tight bounding box supervision in fundus photography using deep learning. *Multimed Tools Appl*, 82:16455–16477, 2023. doi: 10.1007/s11042-022-14183-2.

[72] J. Wang, X. Li, and Y. Cheng. Towards an extended efficientnet-based U-Net framework for joint optic disc and cup segmentation in the fundus image. *Biomedical Signal Processing and Control*, 85: 104906, 2023. doi: 10.1016/j.bspc.2023.104906.

[73] L. Wang, H. Liu, Y. Lu, H. Chen, J. Zhang, and J. Pu. A coarse-to-fine deep learning framework for optic disc segmentation in fundus images. *Biomedical Signal Processing and Control*, 51:82–89, 2019. doi: 10.1016/j.bspc.2019.01.022.

[74] H. Xiong, S. Liu, R. V. Sharan, E. Coiera, and S. Berkovsky. Weak label based bayesian U-Net for optic disc segmentation in fundus images. *Artificial Intelligence in Medicine*, 126:102261, 2022. doi: 10.1016/j.artmed.2022.102261.

[75] S. Yan, X. Pan, and Y. Wang. Mrsnet: Joint consistent optic disc and cup segmentation based on large kernel residual convolutional attention and self-attention. *Digital Signal Processing*, 145:104308, 2024. doi: 10.1016/j.dsp.2023.104308.

[76] J. Yi, Y. Ran, and G. Yang. Particle swarm optimization-based approach for optic disc segmentation. *Entropy*, 24(6), 2022. doi: 10.3390/e24060796.

[77] Y. Yi, Y. Jiang, B. Zhou, N. Zhang, J. Dai, X. Huang, Q. Zeng, and W. Zhou. C2ftfnet: Coarse-to-fine transformer network for joint optic disc and cup segmentation. *Computers in Biology and Medicine*, 164:107215, 2023. doi: 10.1016/j.compbiomed.2023.107215.

[78] X. Zhang, J. Song, C. Wang, and Z. Zhou. Convolutional autoencoder joint boundary and mask adversarial learning for fundus image segmentation. *Frontiers in Human Neuroscience*, 16:1–12, 2022. doi: 10.3389/fnhum.2022.1043569.

[79] A. Zhao, H. Su, C. She, X. Huang, H. Li, H. Qiu, Z. Jiang, and G. Huang. Joint optic disc and cup segmentation based on elliptical-like morphological feature and spatial geometry constraint. *Computers in Biology and Medicine*, 158:106796, 2023. doi: 10.1016/j.compbiomed.2023.106796.

[80] W. Zhou, J. Ji, Y. Jiang, J. Wang, Q. Qi, and Y. Yi. Eards: Efficientnet and attention-based residual depth-wise separable convolution for joint od and oc segmentation. *Frontiers in Neuroscience*, 17, 2023. doi: 10.3389/fnins.2023.1139181.

[81] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. UNet++: A nested U-Net architecture for medical image segmentation. In *arXiv preprint arXiv:1807.10165*. arXiv, 2018. doi: 10.48550/arXiv.1807.10165.

[82] Q. Zhu, X. Chen, Q. Meng, J. Song, G. Luo, M. Wang, F. Shi, Z. Chen, D. Xiang, L. Pan, Z. Li, and W. Zhu. Gdcseg-net: general optic disc and cup segmentation network for multi-device fundus images. *Biomedical Optics Express*, 12(10):6529–6544, 2021. doi: 10.1364/BOE.434841.

[83] J. Zilly, J. M. Buhmann, and D. Mahapatra. Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. *Computerized Medical Imaging and Graphics*, 55:28–41, 2017. doi: 10.1016/j.compmedimag.2016.07.012.

[84] V. Čekanavičius and G. Murauskas. *Statistika ir jos taikymai 1*. TEV, Vilnius, 2001.

[85] V. Čekanavičius and G. Murauskas. *Statistika ir jos taikymai 2*. TEV, Vilnius, 2002.

Articles in international research journals with a citation index in the Clarivate Web of Science (Clarivate WoS) database, other peer-reviewed scientific periodic journals, and peer-reviewed scientific conference proceedings.

Articles in Clarivate WoS journals:

[A.1] Virbukaitė, S. & Bernatavičienė, J. & Imbrasienė, D. Glaucoma Identification Using Convolutional Neural Networks Ensemble for Optic Disc and Cup Segmentation. *IEEE Access*. **12**, 82720-82729 (2024), DOI: 10.1109/ACCESS.2024.3412185

[A.2] Virbukaitė, S. & Bernatavičienė J. Impact of eye fundus image preprocessing on key objects segmentation for glaucoma identification. *Nonlinear Analysis: Modelling and Control*. **29**, 96-110 (2024), DOI: 10.15388/namc.2024.29.33669

Papers in other peer-reviewed scientific periodic journals:

[B.1] Virbukaitė, S. & Bernatavičienė, J. Deep Learning Methods for Glaucoma Identification Using Digital Fundus Images, *Baltic journal of modern computing*. University of Latvia. ISSN 2255-8942. eISSN 2255-8950. 2020, Vol. 8, p. 520–530. DOI: 10.22364/bjmc.2020.8.4.03

Papers in peer-reviewed scientific conference proceedings:

[C.1] Virbukaitė, S. & Bernatavičienė, J. Deep Neural Networks Application for Cup-to-Disc Ratio Estimation in Eye Fundus Images, *Annals of Computer Science and Information Systems. Proceedings of the 18th Conference*. IEEE, 2023. ISSN 2300-5963. Vol. 35, p. 1191–1195. DOI: 10.15439/2023F944

[C.2] Virbukaitė, S. & Bernatavičienė, J. Image Resizing Impact on Optic Disc and Optic Cup Segmentation, *WSCG 2022: full papers proceedings: 30. International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*. University of West Bohemia, 2022. p. 306–309. DOI: 10.24132/CSRN.3201.39

CURRICULUM VITAE

Sandra Virbukaitė obtained her Bachelor's degree in Informatics Engineering in 2009 and a Master's degree in Applied Statistics in 2011 from Vilnius Gediminas Technical University. Since 2020, she has been pursuing a PhD at the Vilnius University Institute of Data Science and Digital Technologies, focusing on deep learning methods for glaucoma identification. Her research interests include computer vision, medical image analysis, deep learning, and ophthalmic diagnostics. She actively participates in international conferences, summer schools, and workshops related to deep learning applications in healthcare.

Her professional career began in 2008 as an IT systems analyst at an international company. Since then, her career has spanned roles including data scientist and executive AI solutions manager. Since 2020, she has also been working as a lecturer in Database Query Languages at Vilnius University.

Glaukoma yra pagrindinė negrįžtamo aklumo priežastis visame pasaulyje, o ją diagnozuojant dažnai remiamasi akies dugno vaizdų struktūrine analize [59]. Pagrindinis rodiklis yra ekskavacijos ir regos nervo disko santykis (CDR), kurio pagalba galima kiekybiškai įvertinti regos nervo disko (OD) ir ekskavacijos (OC) santykį [10, 71]. Dažniausiai naudojami keli CDR rodikliai, tarp jų ploto (ACDR) [12], vertikalaus (VCDR) [10] ir horizontalaus (HCDR) diametrų [44] santykius. Norint tiksliai įvertinti šiuos rodiklius, reikia tiksliai segmentuoti OD ir OC. Šią užduotį tradiciškai rankiniu būdu atlieka oftalmologai, tačiau ji vis dažniau automatizuojama naudojant kompiuterinio matymo metodus, tokius kaip konvoliuciniai neuroniniai tinklai (CNN).

Nors automatizuotos sistemos yra pranašesnės greičio ir nuoseklumo atžvilgiu, CNN pagrįsti metodai susiduria su keliais iššūkiais. Pirma, mokymo duomenų anotacijoms įtakos turi anotatoriaus patirtis ir subjektyvi interpretacija, todėl OD ir OC ribos gali būti netikslios. Šis nevienodumas iškraipo etaloninius duomenis (angl. ground truth), daro tiesioginę įtaką CNN pagrįstų modelių veikimui. Antra, CDR slenksčių, kurie naudojami klasifikuojant glaukomą ir normalius atvejus, apibrėžimo ir taikymo skirtumai sukelia neatitikimų, kurie riboja automatizuotų glaukomos identifikavimo sistemų palyginamumą.

Todėl disertacijoje nagrinėjami šie iššūkiai, vertinamas kelių skirtingo lygio ekspertų anotacijų poveikis CNN pagrįstai OD ir OC segmentacijai. Be to, analizuojama skirtingų tyrimų CDR slenksčių įvairovė ir siūlomi metodai CDR metrikų standartizavimui.

Tyrimo problema

Glaukomos diagnostikos automatizavimas naudojant CNN gali pagerinti akių sveikatos vertinimo efektyvumą ir tikslumą. CNN geba analizuoti akies dugno nuotraukas ir segmentuoti pagrindines tinklainės struktūras, tokias kaip OD ir OC, leidžiant tiksliai įvertinti CDR. Tačiau, nepaisant jų potencialo, kuriant automatizuotas glaukomos identifikavimo sistemas, kyla nemažai iššūkių.

Pagrindinė tyrimo problema yra susijusi su duomenų bazių įvairove, kuri yra itin svarbus veiksnys kuriant CNN pagrįstus modelius. Dauguma giliojo mokymosi (DL) pagrindu sukurtų automatizuotų glaukomos

diagnostikos įrankių mokymui ir vertinimui naudoja tą pačią duomenų bazę. Šis būdas lemia modelių prisitaikymą prie konkrečių vaizdų tipų, tačiau jie gali veikti netiksliai su vaizdais iš kitų populiacijų ar naudojant skirtingus vaizdų gavimo įrenginius.

Antrasis tyrimo iššūkis yra susijęs su išankstiniu vaizdų apdorojimu. Skirtingomis kameromis užfiksuoti akies dugno vaizdai skiriasi raiška, o jų didelis dydis gali neatitikti CNN įvesties matmenų, todėl prieš pradedant procesą juos tenka keisti. Tiriamosios srities (ROI) išskyrimo ir dydžio keitimo procesai gali paveikti vaizdo kokybę, o tai gali turėti įtakos OD ir OC segmentavimo tikslumui, kuris yra itin svarbus tiksliam CDR matavimui.

Kita svarbi problema yra anotacijų variantiškumas akies dugno vaizdų duomenų rinkiniuose. CNN tinklų mokymas remiasi anotuotais duomenimis, tačiau OD ir OC segmentavimas gali gerokai skirtis dėl skirtingos ekspertų patirties ir interpretacijų. Šie žymėjimo neatitikimai iškraipo etaloninius duomenis, darydami įtaką CNN modelių veikimui.

Be to, CDR rodiklio ir jo slenksčių kintamumas kelia papildomų iššūkių. Skirtingi tyrimai ir klinikinės gairės taiko įvairius CDR matavimo metodus bei skirtingus slenksčius glaukomai nustatyti, todėl diagnozės gali būti nevienodos. Nesant standartizuoto CDR slenksčio, automatizuotos glaukomos identifikavimo sistemos gali pateikti prieštaringus rezultatus – jie gali skirtis priklausomai nuo naudojamos duomenų bazės ar taikomų klinikinių gairių, todėl sunku nustatyti vienodus diagnostinius kriterijus.

Disertacijoje nagrinėjami šie iššūkiai, vertinant duomenų įvairovę, išankstinio vaizdų apdorojimo metodus, ekspertų anotacijų variantiškumą bei CDR rodiklių ir slenksčių apibrėžimo neatitikimų poveikį. Šioje disertacijoje siūloma standartizuota metodika, tokiu būdu siekiama pagerinti CNN pagrindu veikiančių glaukomos identifikavimo sistemų kūrimą ir palyginamumą, efektyvinti glaukomos patikrą.

Darbo aktualumas

Akių ligos, tokios kaip glaukoma, diabetinė retinopatija, geltonosios dėmės degeneracija ir tinklainės atsiskyrimas, yra pagrindinės priežastys, lemiančios regėjimo sutrikimus ir aklumą visame pasaulyje. Ankstyva diagnostika ir savalaikis stebėjimas yra itin svarbūs siekiant išsaugoti regėjimą ir užkirsti kelią negrįžtamiems pažeidimams [11]. Iš šių ligų

būtent glaukoma kelia ypatingą iššūkį dėl to, kad ilgą laiką vystosi be jokių simptomų, o išryškėja tik vėlyvose stadijose, todėl ji yra antroji pagal dažnumą aklumo priežastis pasaulyje. Glaukoma paveikia apie 3–5 % 40-80 metų amžiaus grupės gyventojų. 2013 metais buvo paskaičiuota, kad šioje amžiaus grupėje glaukoma sirgo apie 64,3 milijono žmonių visame pasaulyje. Buvo prognozuota, kad iki 2020 metų šis skaičius išaugs iki 76 milijonų, o iki 2040 metų gali siekti net 111,8 milijono [63]. Lietuvoje glaukomos paplitimas per pastarąjį dešimtmetį taip pat didėja – nuo 37,7 atvejų 1000-čiui gyventojų 2015 metais iki 40,7 atvejų 2019 metais ir 41,5 atvejų 1000-čiui gyventojų 2023 metais [20].

Vienas pagrindinių glaukomos diagnostikos rodiklių yra CDR, kuris gaunamas analizuojant akies dugno nuotraukas. Akių gydytojai gali matuoti CDR rankiniu būdu, remdamiesi savo medicinine patirtimi arba naudotis automatizuotais kompiuteriniais metodais, tokiais kaip CNN. Automatizuotos sistemos pranašumas - didesnis darbo greitis ir patikimumas segmentuojant OD ir OC, o tai būtina tiksliam CDR skaičiavimui. Tačiau, diegiant automatizuotus segmentavimo algoritmus, kurie naudoja CDR kaip slenkstį glaukomai nustatyti, pastebimi reikšmingi šio slenksčio apibrėžimo ir taikymo skirtinguose tyrimuose skirtumai. Kuriant automatizuotas glaukomos diagnostikos sistemas, dažniausiai remiamasi anotuotomis akies dugno vaizdų duomenų rinkiniais. Viešai prieinamuose duomenų rinkiniuose pateikiami vaizdai, kuriuos aprašė ir anotavo keli ekspertai, turintys skirtingą patirtį ir diagnozavimo požiūrį. Tačiau iki šiol nebuvo įvertintas skirtingų ekspertų anotacijų poveikis CNN mokymui. Taip pat trūksta bendros nuomonės apie standartizuotą CDR slenksčio ir matavimo metodą, o tai gali apsunkinti įvairių tyrimų rezultatų palyginimą.

Tyrimo objektas

Tyrime dėmesys sutelktas į šiuos tyrimo objektus:

- Pagrindinės tinklainės struktūros, tokios kaip OD ir OC, skirtos patologiniams pokyčiams, susijusiems su įvairiomis akių ligomis, nustatyti.
- DL metodai automatiškai identifikuoti glaukomą.
- Kiekybiniai rodikliai, tokie kaip CDR, identifikuoti glaukomą.

Tyrimo tikslas ir uždaviniai

Šiuo tyrimu siekiama ištirti ir sistemingai įvertinti pagrindinius veiksnius, darančius įtaką DL modelių tikslumui diagnozuojant glaukomą, ir, remiantis gautais rezultatais, pasiūlyti bei patvirtinti daugiapakopę metodiką, kuri padės tiksliau ir efektyviau atlikti kompiuterizuotą glaukomos diagnostiką iš akių dugno vaizdų. Šio tyrimo uždaviniai yra šie:

- Atlikti analitinę DL metodų, taikomų akies dugno vaizdo segmentavimui, apžvalgą, nustatant pagrindinius veiksnius, darančius įtaką segmentavimo rezultatams.
- Įvertinti vaizdų apdorojimo metodų, tokių kaip mastelio keitimas ir apkarpymas, veiksmingumą ir pasirinkti tinkamus metodus siūlomai metodikai.
- Plėtoti DL metodus glaukomos nustatymui, sprendžiant įvairios kokybės akies dugno vaizdų apibendrinimo problemą.
- Nustatyti glaukomos rizikos grupės (stadijos) identifikavimo ribas, remiantis statistine analize ir įtraukiant įvairių ekspertų įžvalgas.
- Sukurti daugiapakopę metodiką, kuri atsižvelgtų į segmentavimo rezultatams įtaką darančius veiksnius ir prisidėtų prie tikslesnio ir patikimesnio kompiuterizuoto glaukomos diagnozavimo.

Tyrimo metodai

Šioje disertacijoje pateikiama išsami literatūros apžvalga, apimanti DL glaukomos diagnozavimo modelius, dažniausiai naudojamus akies dugno vaizdų duomenų rinkinius, duomenų apdorojimo metodus ir šioje srityje naudojamus diagnostinius rodiklius. Šiame tyrime naudojami šie tyrimo metodai:

- Išsami literatūros apžvalga apie įvairių tipų akių dugno kameras, naudojamas akių dugno vaizdams fiksuoti, ir įvairias vaizdų apdorojimo technikas bei DL architektūras, skirtas vaizdų segmentavimui.

- DL modelių, skirtų identifikuoti glaukomą, plėtojimas arba modifikavimas.
- Vaizdų išankstinio apdorojimo metodai, įskaitant ROI išskyrimą, dydžio keitimą naudojant įvairius interpoliacijos metodus, normalizavimą ir augmentavimą, siekiant užtikrinti, kad vaizdai būtų optimaliai paruošti DL modelio mokymui.
- Eksperimentiniai metodai ir viešai prieinamų duomenų rinkinių, tokių kaip Chákṣu IMAGE (toliau Chákṣu), DRISHTI-GS, REFUGE ir RIM-ONE r3 (toliau – RIM-ONE), taikant kiekybinius rodiklius, tokius kaip Dice koeficientas (Dice) ir Susikirtimas padalintas iš sąjungos (IoU), kartais vadinamas Jaccard koeficientu, ir statistiniai metodai, tokie kaip dispersijos analizė (ANOVA), Leveno testas ir Stjudento testas.

Mokslinis darbo naujumas

Šioje disertacijoje siūloma daugiapakopė metodika, skirta akies dugno vaizdų išankstiniam apdorojimui, OD ir OC segmentavimui bei glaukomos nustatymui akies dugno vaizduose. Tai pagerina automatizuotų diagnostikos sistemų tikslumą ir standartizavimą, suteikia galimybę priimti nuoseklesnius sprendimus anksti ir teisingai kliniškai identifikuojant glaukomą. Mokslinį naujumą galima apibūdinti keliais pagrindiniais punktais:

1. Nustatyta, kad vaizdų išankstinio apdorojimo metodai, ypač dydžio keitimas ir interpoliacija, daro reikšmingą įtaką pagrindinių anatominių struktūrų, būtent OD ir OC, segmentavimo tikslumui diagnozuojant glaukomą.
2. Pristatytas DL metodais pagrįstas OD ir OC segmentavimo ansamblis, leidžiantis pasiekti geresnių rezultatų pagal Dice, lyginant su vieno modelio metodais. Pažymėtina, kad OD segmentavimas pagal Dice įvairiuose duomenų rinkiniuose pagerėjo vidutiniškai 4 %, o OC segmentavimas pagal Dice – vidutiniškai 5 %.
3. Siūlomi ACDR, VCDR ir HCDR slenkstčiai, skirti akies sveikatos klasifikavimui į dvi pagrindines kategorijas: normalią ir glaukomą, papildomai įvedant naują stadiją — „įtariama glaukoma“. Ši

siūloma „įtariama glaukoma“ stadija pagerina ankstyvos glaukomos identifikavimo galimybes.

4. Išskiriamas ekspertų žymėjimo kintamumas OD ir OC segmentavime bei jo poveikis DL metodams, tokiems kaip CNN, veikimui. Statistinė analizė parodė reikšmingus ekspertų skirtumus, nustatant glaukomą pagal CDR, pabrėžiant subjektyvią rankinio vertinimo esmę.
5. Nustatyta, kad VCDR, lyginant su ACDR ir HCDR, yra patikimiausias CDR rodiklis identifikuoti glaukomą.

Praktinė darbo vertė

Šiame tyrime pristatoma nauja DL pagrindu sukurta daugiapakopė glaukomos diagnostikos metodika, ypatingą dėmesį skiriant ankstyvosios stadijos nustatymui. Pasitelkus DL algoritmus, siūloma metodika, įskaitant CNN ansamblio metodą, leidžia automatiškai ir objektyviai vertinti glaukomos stadijas, sumažinant priklausomybę nuo subjektyvių ekspertų vertinimų.

Konkrečių CDR slenksčių integravimas palengvina sistemingą akių sveikatos klasifikavimą į normalius atvejus, įtariamą glaukomą ir glaukomos stadijas. Šis sisteminis suskirstymas į kategorijas pagerina ankstyvą identifikavimą ir leidžia laiku imtis intervencijų.

Lyginamoji OD ir OC segmentavimo rezultatų analizė rodo, kad DL pagrindu atliktas segmentavimas neturi reikšmingų skirtumų lyginant su ekspertų vertinimais, o tai rodo, kad CNN modeliais galima pasitikėti kaip automatizuotos glaukomos diagnostikos įrankiu, todėl jie yra tinkami taikyti realiomis klinikinėmis sąlygomis.

Ginamieji teiginiai

- Skirtinga vaizdo skiriamoji geba ir interpoliacijos metodai turi įtakos segmentavimo rezultatui. Geriausi segmentavimo rezultatai gauti naudojant bikubinę interpoliaciją, mažinant vaizdo dydį iki 512×512 pikselių.
- Tarp trijų vertintų rodiklių (ACDR, HCDR ir VCDR) VCDR pasižymėjo geriausiais rezultatais skiriant glaukomos stadijas, todėl laikytinas tinkamiausiu glaukomos vertinimo rodikliu.

- Ekspertų žymėjimų skirtumai akies dugno vaizduose daro statistiškai reikšminę įtaką CNN segmentavimo tikslumui.
- Siūlomas CNN ansamblis rodo segmentavimo tikslumo pagerėjimus keliuose duomenų rinkiniuose. OD segmentavimo tikslumas pagal Dice įvariuose duomenų rinkiniuose vidutiniškai pagerėjo 4 %, o OC segmentavimo tikslumas pagerėjo vidutiniškai 5 %.

Darbo rezultatų aprobavimas ir publikavimas

Disertacijoje pristatomų tyrimų rezultatai buvo publikuoti 2 periodiniuose mokslo žurnaluose, indeksuotuose *Clarivate Web of Science (Clarivate WoS)* duomenų bazėje, 1 recenzuojamame mokslo žurnale ir 2 recenzuojamuose mokslinių konferencijų pranešimų rinkiniuose. Rezultatai pristatyti 2 tarptautinėse ir 5 nacionalinėse mokslinėse konferencijose.

Publikacijos, paskelbtos tarptautiniuose mokslo žurnaluose, indeksuojamuose *Clarivate WoS* citavimo indekse:

1. Sandra Virbukaitė, Jolita Bernatavičienė, Daiva Imbrasienė. Glaucoma Identification Using Convolutional Neural Networks Ensemble for Optic Disc and Cup Segmentation, *IEEE Access*. IEEE. eISSN 2169-3536. 2024, Vol. 12, p. 82720–82729. DOI: 10.1109/ACCESS.2024.3412185.
2. Sandra Virbukaitė, Jolita Bernatavičienė. Impact of Eye Fundus Image Preprocessing on Key Objects Segmentation for Glaucoma Identification, *Nonlinear Analysis: Modelling and Control*. Vilnius University Press. ISSN 1392-5113. eISSN 2335-8963. 2024, Vol. 29, p. 96–110. DOI: 10.15388/namc.2024.29.33669.

Straipsniai kituose recenzuojamuose periodiniuose mokslo žurnaluose:

1. Sandra Virbukaitė, Jolita Bernatavičienė. Deep Learning Methods for Glaucoma Identification Using Digital Fundus Images, *Baltic journal of modern computing*. University of Latvia. ISSN 2255-8942. eISSN 2255-8950. 2020, Vol. 8, p. 520–530. DOI: 10.22364/bjmc.2020.8.4.03.

Straipsniai recenzuojamuose mokslinių konferencijų leidiniuose:

1. Sandra Virbukaitė, Jolita Bernatavičienė. Deep Neural Networks Application for Cup-to-Disc Ratio Estimation in Eye Fundus Images, *Annals of Computer Science and Information Systems. Proceedings of the 18th Conference*. IEEE, 2023. ISSN 2300-5963. Vol. 35, p. 1191–1195. DOI: 10.15439/2023F944.

2. Sandra Virbukaitė, Jolita Bernatavičienė. Image Resizing Impact on Optic Disc and Optic Cup Segmentation, *WSCG 2022: full papers proceedings: 30. International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*. University of West Bohemia, 2022. p. 306–309. DOI: 10.24132/CSRN.3201.39.

Pranešimai tarptautinėse mokslinėse konferencijose:

1. Sandra Virbukaitė. Deep Neural Networks Application for Cup-to-Disc Ratio Estimation in Eye Fundus Images, *18th Conference on Computer Science and Intelligence Systems FedCSIS 2023*. Warsaw, Poland, September 17–20, 2023.

2. Sandra Virbukaitė. Image Resizing Impact on Optic Disc and Optic Cup Segmentation, *30th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*. Pilzen, Czech Republic, May 17–20, 2022.

Pranešimai nacionalinėse mokslinėse konferencijose:

1. Sandra Virbukaitė. How the Expertise of Different Experts Influences the Learning of Convolutional Neural Networks, *Artificial intelligence technologies in medicine: research and diagnostics*. Vilnius, Lithuania, October 23, 2024.

2. Sandra Virbukaitė. Application of Convolutional Neural Networks for Glaucoma Identification, *Artificial intelligence technologies in medicine: research and diagnostics*. Vilnius, Lithuania, October 20, 2023.

3. Sandra Virbukaitė. Interpolation Methods Impact on Eye Fundus Optic Disc and Optic Cup Segmentation, *Data Analysis Methods for Software Systems. 13th International Workshop*. Druskininkai, Lithuania, December 1–3, 2022.

4. Sandra Virbukaitė. Impact of Images Quality Variety and Resizing Level on Eye Fundus Optic Disc Segmentation, *Data Analysis Methods for Software Systems. 12th International Workshop*. Druskininkai, Lithuania, December 2–4, 2021.

5. Sandra Virbukaitė. Image Quality Impact on Optic Disc Segmentation Accuracy, *Computer Days*. Klaipėda, Lithuania, September 23–24, 2021.

Disertacijos strukūra

Šioje dalyje pateikiamos pagrindinės disertacijos dalys ir bendra apimtis.

Disertacija susideda iš įvado, trijų skyrių, išvadų ir santraukos lietuvių kalba. Įvade pristatomas tyrimas ir pateikiama disertacijos apžvalga. Pirmame skyriuje apžvelgiama mokslinė literatūra, antrame skyriuje nagrinėjami DL metodai, taikomos metrikos ir siūloma metodika vaizdams segmentuoti glaukomos identifikavimo tikslu, o trečiame skyriuje pristatomi atlikti DL eksperimentai bei jų rezultatai, susiję su glaukomos identifikavimu.

Bibliografinės nuorodos pateikiamos darbo pabaigoje. Disertaciją sudaro 132 puslapiai, 12 paveikslų ir 28 lentelės.

S.1. LITERATŪROS APŽVALGA

Šiame skyriuje pateikiama pagrindinių mokslinių tyrimų sričių, susijusių su automatinio glaukomos nustatymo pagal akies dugno vaizdus, apžvalga. Jame aptariami akies dugno vaizdai ir duomenų rinkiniai, DL vaizdų segmentavimo metodai, CNN ansamblio glaukomos nustatymo metodai, dažniausiai naudojamos metrikos ir slenksčiai bei susiję tyrimai, atlikti Lietuvoje.

S.1.1. Akių dugno vaizdai ir duomenų rinkiniai

Akių dugno vaizdų analizė tai procesas, kurio metu įvairių tipų akių dugno kameromis fiksuojami akių vidinio paviršiaus vaizdai. Šiuose vaizduose matomos pagrindinės anatominės struktūros, įskaitant geltonąją dėmę, foveą, neuroretinalinį kraštą, tinklainės kraujagysles ir OD, o OC yra jos centre. OD yra atsakingas už informacijos perdavimą elektrinių impulsų forma iš tinklainės į smegenis per regos nervą. OC yra nedidelis įdubimas, vadinamas fiziologine duobute, esantis OD centre. Akių dugno vaizdai, kuriuose matomos šios pagrindinės struktūros, gali būti saugomi viešai prieinamuose duomenų rinkiniuose, kad būtų galima juos naudoti moksliniams tyrimams ir diagnostikai. Toliau pateikia išsamesnė jų apžvalga.

REFUGE duomenų rinkinys [47] buvo sukurtas kaip *Retinal Fundus Glaucoma Challenge* dalis, kurį sudaro 1200 aukštos kokybės spalvotų Kinijos pacienčių akių dugno vaizdų. Iš jų 800 vaizdų buvo gauti naudojant Canon CR-2 AF su 1634×1634 pikselių (px) skiriamąja geba, o likę 400 – naudojant Zeiss Visucam 500 akies dugno kamerą su 2124×2056 px skiriamąja geba.

RIM-ONE [19] yra glaukominių ir neglaukominių tinklainės vaizdų kolekcija, gauta iš įvairių Ispanijos ligoninių naudojant nemidriatinę Kowa WX 3D stereo akies dugno kamerą. Rinkinys apima 85 normalius vaizdus, 35 vaizdus, kuriuose įtariama glaukoma, ir 39 glaukomos vaizdus, kurių skiriamoji geba yra 2144×1424 px vienam vaizdui stereo porose.

DRISHTI-GS duomenų rinkinys [57] susideda iš 70 glaukomos ir 31 sveikų spalvotų akių dugno vaizdų, kurie buvo užfiksuoti Aravind akių ligoninėje Madurai.

Chákṣu duomenų rinkinys [30] susideda iš 1345 spalvotų Indijos pacientų akių dugno vaizdų, užfiksuotų naudojant tris skirtingas nemidriatines akies dugno kameras Remidio Fundus-on-Phone (FoP), Forus 3Nethra Classic ir rankinę akių dugno kamerą Bosch. Šių kamerų užfiksuoti vaizdai atitinkamai suskirstyti į tris duomenų rinkinius:

- REMIDIO: 1074 vaizdai, kurių skiriamoji geba yra 2448×3264 px, tarp jų 150 vaizdų, klasifikuojamų kaip glaukoma, ir 924 vaizdai, klasifikuojami kaip normalūs.
- FORUS: 126 vaizdai, kurių skiriamoji geba yra 2048×1536 px, iš kurių 18 identifikuoti kaip glaukoma, o 108 – kaip normalūs.
- BOSCH: 145 vaizdai, užfiksuoti 1920×1440 px raiška, iš kurių 20 yra pažymėti kaip glaukoma, o 125 – kaip normalūs.

S.1.1.1. Giliojo mokymosi akių dugno segmentavimo metodai glaukomai identifikuoti

Diagnozuojant glaukomą tradiciniais metodais, remiamasi gydytojų profesionalumu, nes jie rankiniu būdu analizuoja akių dugno vaizdus, matuoja CDR bei vertina subtilius regos nervo pokyčius. Šis procesas dažnai yra lėtas, reikalauja daug laiko, gali pasitaikyti žmogiškųjų klaidų. Priešingai, naudojant įvairių aukštos kokybės akių dugno vaizdų duomenų rinkinius, automatizuotos pažangių technologijų sistemos, tokias kaip DL, gali greitai ir nuosekliai analizuoti ir įvertinti didelius kiekius akių dugno vaizdų. Todėl DL vaizdų segmentavimo metodai, tampa vis svarbesni oftalmologijoje, ypač tiriant ir diagnozuojant glaukomą. DL metodai leidžia automatiškai nustatyti ir analizuoti svarbias akių dugno vaizdų savybes, tokias kaip OD ir OC, kurie yra esminiai vertinant glaukomą.

Naujausi tyrimai [83], [36], [29] pabrėžia perspektyvius DL modeliais gaunamus rezultatus, sprendžiant unikalius OD ir OC segmentavimo iššūkius. Nors dabartiniai DL metodai pasiekia aukštą OD segmentavimo tikslumą, OC segmentavimo tikslumas vis dar išlieka palyginti žemas. Tačiau šie rezultatai daugiausiai buvo pasiekti modelių mokymui naudojant atskirus, tik vienos rūšies duomenų rinkinius. Šių tinklų gebėjimas apibendrinti ir efektyviai segmentuoti vaizdus iš įvairių bei skirtingų duomenų rinkinių tuo pačiu metu lieka atvira tyrimų sritis,

reikalaujanti tolesnių tyrimų. Be to, išlieka iššūkis gauti tikslius OC segmentavimo rezultatus.

S.1.2. Glaukomos rizikos nustatymo priemonės ir slenksčiai

Naudojant automatinius segmentavimo algoritmus, kurie remiasi CDR kaip slenksčiu, skiriančiu glaukomą ir normalius atvejus, pastebimi slenksčio pasirinkimo skirtumai tarp tyrimų, kuriuose naudojami CNN pagrįsti metodai, tad tinkamas slenkstis lieka neaiškus. S.1 lentelėje pateikiami CDR metrikos ir jos slenksčių verčių tyrimų skirtumai.

S.1 lentelė: CDR metrikos ir jos slenksčių verčių skirtingų tyrimų skirtumai

Literatūros šaltiniai	**Metrika**	**Slenkstis**
[75]	VCDR	> 0,7
[80], [33], [6]	VCDR	> 0,5
[71]	VCDR	$\geqslant 0,6$
[39]	VCDR	> 0,6
[56]	VCDR	> 0,4
[44]	ACDR VCDR ir HCDR	$\geqslant 0,3$ $\geqslant 0,5$
[10]	VCDR	= 0,6
[13]	ACDR	> 0,3

S.1.3. Susiję tyrimai Lietuvoje

Automatizuoto glaukomos nustatymo tyrimai Lietuvoje, vystėsi beveik du dešimtmečius.

Vienas ankstyviausių tyrimų, kurį 2007 m. atliko P. Treigys ir V. Šaltenis [66], nagrinėjo neuroninių tinklų taikymo galimybes klasifikuojant ligas, parodydamas, kad derinant šiuos modelius su dimensijų mažinimo metodais galima veiksmingai atskirti glaukomos atvejus pagal OD struktūrinius parametrus. Vėlesniame darbe P. Treigys ir kt. [65] pasiūlė automatizuotą OD lokalizacijos ir parametrizacijos spalvotų tinklainės nuotraukų algoritmą, derindami morfologinį vaizdų apdorojimą, adaptyvų kraštų detekciją ir geometrinį modeliavimą. OD centro lokalizacijos tyrime buvo taikytas iteracinis apskritiminis Hough transformacijos metodas. P. Treigio ir G. Dzemydos [64] parengtas tyrimas padėjo spręsti linijinio stebėjimo problemą, pasiūlydamas automatizuotą vaizdų registracijos metodą, leidžiantį palyginti skirtingu metu

darytas tinklainės nuotraukas ir tiksliau įvertinti ligos progresavimą. Lygiagrečiai šiems tyrimams J. Bernatavičienė ir kt. [4] nagrinėjo automatizuotus duomenų klasifikavimo taisyklių indukcijos metodus, pabrėždami diagnostinę OD ir OC skersmenų santykio reikšmę. Tolimesni D. Buteikienės ir kt. [7] darbai išplėtojo šias idėjas ir pristatė visiškai automatizuotą OD aptikimo metodą, pagrįstą elipsine kreive ir Bajeso optimizacija, kuriuo buvo pasiektas didelis (0,8) koreliacijos koeficientas tarp automatizuotų ir ekspertų įvertinimų. Vėliau G. Stabingis ir kt. [58] išplėtojo ir pagilino tinklainės tyrinėjimus ir sukūrė automatizuotą arterijų ir venų santykio matavimo sistemą.

Šie tyrimai rodo nuoseklias pastangas automatizuoti ir standartizuoti su glaukoma susijusią diagnostiką, taikant klasikinius vaizdų apdorojimo, geometrinio modeliavimo ir ankstyvuosius neuroninių tinklų metodus. Vis dėlto nors šie metodai sudarė tvirtą pagrindą, juos ribojo ypatybių išskyrimo sudėtingumas.

Pastaraisiais metais giliojo mokymosi (DL) metodų atsiradimas iš esmės pakeitė oftalmologinių vaizdų analizę, suteikdamas galimybę mokytis ypatybių nuo pradžios iki galo ir pasiekti aukštus rezultatus su nepažįstamais duomenimis iš įvairių duomenų rinkinių. 2023 m. V. Raudonio ir kt. [48] tyrime buvo pristatytas DL pagrįstas automatinio mikroaneurizmų aptikimo spalvotose tinklainės nuotraukose metodas – esminis ankstyvos diabetinės retinopatijos diagnostikos žingsnis. Siūlomas metodas naudojo trijų tinklų ansamblį, tokių kaip U-Net, ResNet34-UNet ir UNet++, siekiant pagerinti segmentavimo tikslumą. C. Celik ir kt. [9] pasiūlė automatinio regos nervo hipoplazijos aptikimo sprendimo palaikymų sistemą. Jų metodas integravo U-Net architektūrą su iš anksto apmokytu ResNet koduotoju, siekiant segmentuoti OD ir fovėją.

Šie pasiekimai rodo nuolatinę klasikinių, ypatybėmis pagrįstų duomenų vaizdų apdorojimo metodų raidą link DL grindžiamų diagnostinių sistemų. Šiame platesniame kontekste DL metodų taikymas glaukomos identifikavimui išlieka ir aktualus, ir būtinas, siekiant geresnio diagnostinio tikslumo ir praktinio pritaikomumo klinikinėje aplinkoje.

S.1.4. Skyriaus išvados

Apibendrinant galima teigti, kad akies dugno vaizdų gavimo technologija yra svarbi oftalmologijos ligų, ypač glaukomos, stebėjimo ir diag-

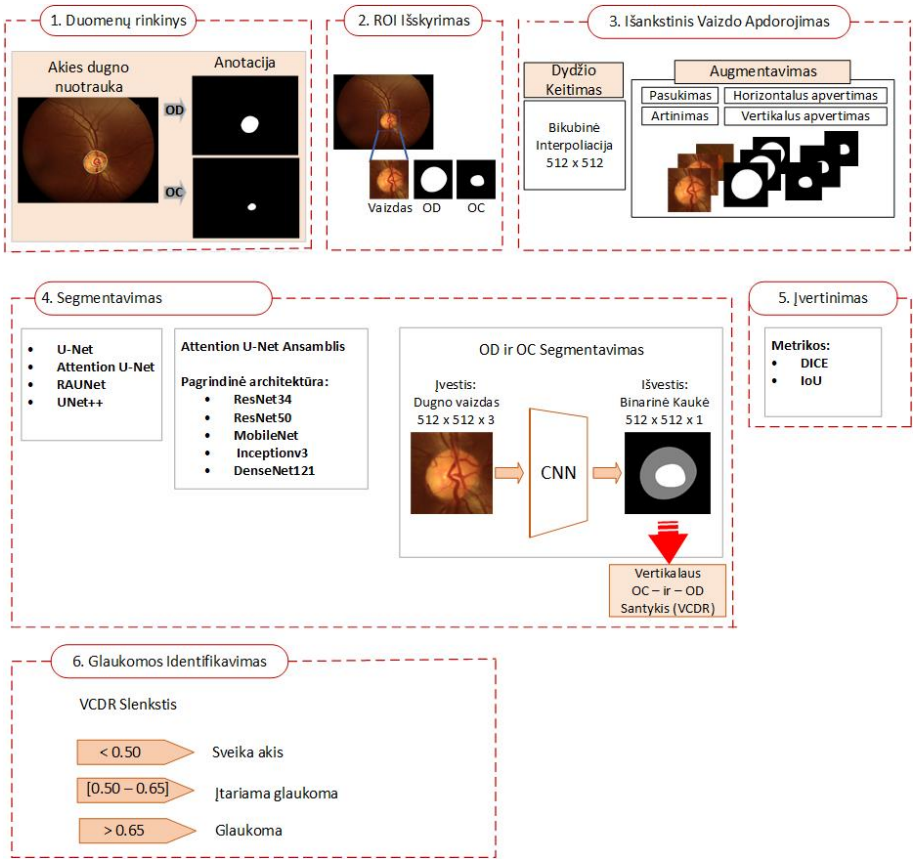
nozavimo priemonė. Šių vaizdų kokybė didžiąja dalimi priklauso nuo naudojamos kameros, o technologijos pažanga apima aukštos skiriamosios gebos sistemas, tokias kaip Forus 3Nethra Classic ir Zeiss Visucam 500, ir nešiojamus prietaisus, pavyzdžiui, Bosch30 rankines akies dugno kameras. Nepaisant šios pažangos, rankiniu būdu atliekama glaukomos diagnozė vis dar yra laiką ir išteklius eikvojantis procesas, kurio metu oftalmologai turi įvertinti daugybę parametrų.

Tyrimai automatinio glaukomos nustatymo srityje Lietuvoje vystėsi beveik du dešimtmečius, iš pradžių daugiausia dėmesio skiriant vaizdų apdorojimui ir klasikinėms ML metodams OD ir OC analizei. Pastaraisiais metais DL atsiradimas pakeitė oftalmologinių vaizdų analizę visame pasaulyje. Darbas su DL modeliais rodo didelę pažangą, ypač OD segmentacijoje, o OC segmentacija tebėra sudėtingesnė. Be to, dauguma rezultatų buvo gauti naudojant CNN, apmokytus vienu duomenų rinkiniu, todėl lieka atviras klausimas, ar šie modeliai gali pasiekti nuoseklų našumą naudojant kelis duomenų šaltinius. Todėl išlieka metodų poreikis, kuriuose naudojamas patobulintas vaizdų segmentacijos požymių išskyrimas. Šio tyrimo tikslas - prisidėti prie diagnostinės sistemos, tinkamos pirminės sveikatos priežiūros paslaugoms, kūrimo, kurią naudodami šeimos gydytojai, aprūpinti rankinėmis akies dugno kameromis, galėtų įvertinti pacientų glaukomos rizikos faktorių ir prireikus nukreipti juos pas specialistus.

Galiausiai, nors dauguma esamų tyrimų remiasi VCDR kaip pagrindiniu rodikliu, taip pat buvo tiriamos alternatyvos, tokios kaip ACDR ir HCDR. Pastebima, kad atliekant kai kuriuos tyrimus siūloma naudoti ACDR su slenksčio verte, didesne nei 0,3 glaukomos nustatymui [44], [12]. Kituose tyrimuose [44] HCDR, didesnis nei 0,5, laikomas anomalija. Taigi, nepaisant intensyvių tyrimų šioje srityje, standartizuotų slenksčių verčių tyrimų trūkumas, atsirandantis dėl duomenų rinkinių, ekspertų žymėjimo ir vertinimo metrikų skirtumų, apsunkina rezultatų palyginimą.

S.2. METODAI

Šiame skyriuje pristatoma tyrimo metu taikyta metodika (pateikta paveiksle S.1) glaukomai identifikuoti naudojant akies dugno vaizdus, apimanti išankstinio vaizdų apdorojimo metodus ir segmentavimo kokybės vertinimo metrikas.



S.1 pav.: Siūlomos metodikos darbo eiga

S.2.1. Akių dugno vaizdų išankstinis apdorojimas

Vaizdų išankstinis apdorojimas gali padėti išspręsti kai kurias problemas, su kuriomis susiduriama kuriant automatizuotas glaukomos nustatymo sistemas, ypač susijusias su skaičiavimo efektyvumu, duomenų rinkinių apribojimu, modelių pritaikomumu ir OC segmentavimu.

Išankstinio vaizdų apdorojimo technikos, tokios kaip triukšmo mažinimas ir vaizdų dydžio keitimas, padeda supaprastinti modelio įvesties duomenis, išlaikant svarbias duomenų savybes. Duome- nų trūkumo atveju duomenų papildymo technikos, tokios kaip vaizdo pasukimas, apvertimas ar mastelio keitimas, gali dirbtinai padidinti modelių mokymo duomenų rinkinio dydį ir įvairovę, taip pagerindamos jų gebėjimą apibendrinti naujus duomenis. Tokios vaizdų apdorojimo technikos, kaip kontrasto didinimas ar normalizavimas, gali

reikšmingai prisidėti prie vaizdų standartizavimo, sumažindamos skirtumus tarp skirtingų duomenų rinkinių ir užtikrindamos, kad modeliai, apmokyti vieno duomenų rinkinio vaizdais, gebėtų apibendrinti ir kitų rinkinių duomenis. ROI išskyrimas, dar viena vaizdų apdorojimo technika, taip pat gali padėti segmentuojant ekskavaciją, kadangi sutelkiamas dėmesys į svarbiausias vaizdo dalis.

Net ir po ROI išskyrimo gauti vaizdai gali būti gana didelio dydžio ir vis tiek reikalauti didelių skaičiavimo išteklių. Siekiant išspręsti šią problemą, dažnai taikoma vaizdų dydžio keitimo praktika. Sumažinus vaizdų matmenis, galima gerokai supaprastinti skaičiavimus, tuo pačiu išlaikant esmines vaizdų ypatybes, reikalingas jų analizei. Daugelio tyrimų vaizdai buvo sumažinti iki standartinių matmenų pikseliais, tokių kaip 128×128 [27, 36, 69], 256×256 [28, 73, 76], ar 512×512 [21]. Dydžio keitimo procesas paprastai apima interpolacijos technikas, kurios įvertina ir generuoja naujas pikselių reikšmes, remdamasis esamomis. Vieni iš dažniausiai naudojamų interpolacijos metodų yra dvitiesinė, dvikubinė ir artimiausio kaimyno interpolacija.

S.2.2. Segmentavimo kokybės įvertinimo metrikos

Atlikus išankstinį vaizdų apdorojimą, kitas svarbus žingsnis kuriant automatizuotas glaukomos identifikavimo sistemas yra giliojo mokymosi modelių mokymas atlikti vaizdo segmentavimo užduotį. Norint įvertinti segmentavimo algoritmų veiksmingumą, naudojamos įvairios metrikos, tokios kaip Dice ir IoU, kuriomis matauojamas prognozuojamo segmentavimo ir tikrojo žymėjimo persidengimo laipsnis.

S.2.3. Glaukomos nustatymo tyrimo metodika, pagrįsta akies dugno įvertinimu

Metodika atitinka sistemingą ir struktūrizuotą procesą, kuris apima išankstinį duomenų apdorojimą, modelio kūrimą ir jo veikimo vertinimą. Procesas prasideda nuo viešai prieinamų akių dugno vaizdų duomenų rinkimo ir integravimo. Siekiant pagerinti įvesties duomenų kokybę, atliekamas vaizdų apdorojimas. Tai apima ROI išgavimą, kai akies dugno vaizdas apkarpomas pagal OD ribą. Siekiant dar labiau standartizuoti duomenis, taikomas vaizdų dydžio keitimas, užtikrinantis nuoseklius įvesties matmenis skirtinguose duomenų rinkiniuose. Be

to, įtraukiamos duomenų papildymo technikos, pavyzdžiui, apvertimas, pasukimas ir intensyvumo kitimas. Galiausiai sukuriamas mišrus duomenų rinkinys, sujungiant kelių skirtingų šaltinių vaizdus, taip padidinant duomenų įvairovę. Procese naudotos kelios CNN architektūros, pavyzdžiui, Attention U-Net, RAUNet ir UNet++, siekiant ištirti įvairių DL modelių veiksmingumą glaukomos nustatymui. Šios architektūros yra mokomos nuosekliomis eksperimentinėmis sąlygomis, taikant hiperparametrų tikslinimą, siekiant optimizuoti mokymosi greitį ir partijų dydį, kur modifikuota Attention U-Net, su DenseNet121 kaip iš anksto apmokytu pagrindu, pasirodo esanti tinkamiausia.

S.2.4. Skyriaus išvados

Tradicinė glaukomos diagnozė priklauso nuo ekspertų atliekamo akies dugno vaizdų įvertinimo, todėl šis procesas trunka ilgai. Dėl aukštos kokybės akies dugno vaizdų duomenų rinkinių prieinamumo DL pagrįstos automatizuotos sistemos yra ypač naudingos segmentuojant OD. Tačiau OC segmentavimas tebėra iššūkis. Norint įveikti šiuos apribojimus reikia tolesnių tyrimų, kad būtų pagerintas įvairių duomenų rinkinių segmentavimo našumas.

Vaizdų apdorojimo technikos, tokios kaip triukšmo mažinimas, kontrasto didinimas, duomenų papildymas ir ROI išskyrimas, atlieka svarbų vaidmenį gerinant modelio našumą. Šie metodai standartizuoja vaizdų įvestis, gerina skirtingų duomenų rinkinių DL modelių pritaikomumą ir galiausiai užtikrina tikslesnius segmentavimo rezultatus.

Segmentavimo modelio našumui įvertinti dažniausiai naudojami tokie rodikliai kaip Dice ir IoU. Be to, statistiniai metodai, tokie kaip ANOVA ir Stjudento t-testas, suteikia vertingos informacijos, lyginant skirtingų rinkinių modelių našumą ir taikant skirtingus metodus. Pažangios DL architektūros, įskaitant Attention U-Net, RAUNet ir UNet++, užtikrina geresnį požymių išskyrimą ir segmentavimo tikslumą.

S.3. EKSPERIMENTAI IR REZULTATAI

Šio tyrimo rezultatai ir metodikos principai buvo pateikti keliuose recenzuojamuose straipsniuose [A.1], [A.2], [B.1], [C.1].

S.3.1. Akių dugno vaizdo išankstinio apdorojimo įtaka OD ir OC segmentavimui

Išankstinis vaizdų apdorojimas yra esminis kompiuterinės regos ir medicininių vaizdų analizės žingsnis, skirtas patobulinti neapdorotus akies dugno vaizdus prieš juos apdorojant DL modeliais. Pagrindiniai tikslai yra pagerinti vaizdo kokybę, sumažinti triukšmą siekiant išryškinti kliniškai svarbias struktūras, tokias kaip OD ir OC. Atsižvelgiant į tai, kad išankstinis apdorojimas taip pat apima DL modelių įvesties dydžių standartizavimą, svarbu atkreipti dėmesį, kad akies dugno vaizdų matmenys skiriasi priklausomai nuo įrenginio. Šie skirtumai yra akivaizdūs duomenų rinkiniuose, pvz., DRISHTI-GS teikia vaizdus 2896×1944 px raiška [57], RIM-ONE - 2144×1424 px [19], o REFUGE - 1634×1634 arba 2124×2056 px [47]. Nors šie didelės skiriamosios gebos vaizdai užfiksuoja daug tinklainės detalių, jie paprastai yra per dideli, kad juos būtų galima tiesiogiai apdoroti CNN, ir apima pašalinius foninius regionus, kurie nėra svarbūs vertinant glaukomą. Siekiant išspręsti šias problemas, išskiriamas ROI, apimantis OD. ROI apibrėžiamas kaip kvadratinis plotas, kurio centras yra OD, o kraštinės yra dvigubai didesnės už OD. Iškirpus ROI, skiriasi duomenų rinkinių matmenys: DRISHTI-GS rinkinyje jie svyruoja nuo 674×674 iki 1060×1060 px, RIM-ONE rinkinyje - nuo 456×456 iki 890×890 px, o REFUGE rinkinyje - nuo 408×408 iki 616×616 px. Tai užtikrina OD centrinės srities užfiksavimą kartu su diagnostikai svarbiu aplinkiniu tinklainės audiniu, o pašalinis fonas pašalinamas.

S.3.2. Vaizdų mastelio keitimo lygio įtaka

Pagrindinis šio tyrimo tikslas yra analizuoti vaizdo mastelio suderinimo technikų, tokių kaip vaizdų mastelio keitimas ir interpoliacijos metodai, įtaką segmentavimo rezultatams. Apkarpytų ROI matmenys buvo aptarti poskyriuje S.3.1. Šiame eksperimente buvo naudojami DRISHTI-GS duomenų rinkinys (su ROI dydžiais nuo 674×674 px iki 1060×1060 px) ir RIM-ONE duomenų rinkinys (su ROI dydžiais nuo 456×456 px iki 890×890 px). Siekiant suvienodinti įvairių dydžių ROI, buvo taikytas vaizdų mastelio keitimas, pritaikant juos prie bendrų dydžių 128×128 [27, 36], 256×256 [28, 76] ir 512×512 px [21, 73], naudojant plačiai taikomą dvitiesinės interpoliacijos metodą. Vaizdai buvo

normalizuoti suvienodinant pikselių reikšmes į intervalą [0, 1]. Dėl ribotо vaizdų skaičiaus DRISHTI-GS (101 vaizdas) ir RIM-ONE (159 vaizdai) duomenų rinkiniuose siekiant praplėsti mokymo aibę, duomenys buvo papildyti.

Norint įvertinti vaizdo mastelio keitimo įtaką OD ir OC segmentavimui, buvo naudojama U-Net architektūra [50], taikant kelias, žemiau pateiktas, eksperimentines konfigūracijas:

- U-Net buvo apmokytas naudojant DRISHTI-GS mokymo duomenų rinkinį ir atskirai įvertintas DRISHTI-GS bei RIM-ONE testavimo rinkiniuose.
- U-Net buvo apmokytas naudojant RIM-ONE mokymo duomenų rinkinį ir nepriklausomai išbandytas RIM-ONE bei DRISHTI-GS testavimo rinkiniuose.
- Buvo sukurtas mišrus mokymo duomenų rinkinys, sujungus visus turimus rinkinius, o apmokytas modelis atskirai įvertintas DRISHTI-GS ir RIM-ONE testavimo rinkiniuose.

Sprendimą naudoti mišrų mokymo duomenų rinkinį nulėmė pastebėjimas, kad DL modeliai dažnai yra mokomi ir tikrinami to paties duomenų rinkinio pagrindu. OD ir OC segmentavimo rezultatai, įvertinti pagal Dice, apibendrinti S.2 lentelėje eksperimentams su U-Net, apmokytu naudojant vaizdus, mastelio keitimu pritaikytus dydžiams 512×512, 256×256 ir 128×128 px iš DRISHTI-GS, RIM-ONE ir mišraus duomenų rinkinio.

S.2 lentelė: OD ir OC segmentavimo, naudojant vaizdus, mastelio keitimu pritaikytus 512×512, 256×256 ir 128×128 pikselių dydžiams iš DRISHTI-GS, RIM-ONE ir mišraus duomenų rinkinio, rezultatai pagal Dice

	512×512 px		**256×256 px**		**128×128 px**	
	OD	**OC**	**OD**	**OC**	**OD**	**OC**
Mokymo ir testavimo rinkiniai						
DRISHTI-GS prieš DRISHTI-GS	0,976	0,905	0,966	0,877	0,935	0,859
DRISHTI-GS prieš RIM-ONE	0,825	0,607	0,774	0,571	0,713	0,537
RIM-ONE prieš RIM-ONE	0,966	0,907	0,956	0,877	0,921	0,826
RIM-ONE prieš DRISHTI-GS	0,824	0,593	0,772	0,566	0,711	0,522
Mišrus prieš DRISHTI-GS	**0,953**	**0,899**	**0,939**	**0,845**	**0,912**	**0,801**
Mišrus prieš RIM-ONE	0,934	0,877	0,909	0,830	0,888	0,794

Šie eksperimentiniai rezultatai, pateikti taip pat ir [C.1], [C.2], rodo, kad didesnė vaizdų raiška pagerina OD ir OC segmentavimą. Kai U-Net buvo testuojamas vaizdais iš to paties duomenų rinkinio, kuris buvo naudojamas mokymui, segmentavimo rezultatai buvo pakankamai aukšti. Tačiau rezultatyvumas sumažėjo, kai U-Net buvo testuojamas vaizdais iš kito duomenų rinkinio. Priešingai, naudojant mišrų duomenų rinkinį, gauti rezultatai buvo perspektyvūs. Geriausias segmentavimo tikslumas pasiektas, kai tinklas buvo apmokytas mišraus duomenų rinkinio vaizdais, kurių mastelis buvo pakeistas į 512×512 px. Taikyta dvitiesinė interpoliacija keičiant mastelį lėmė OC ribų detalių praradimą. Tai pabrėžia poreikį tirti alternatyvius interpoliacijos metodus, tokius kaip artimiausio kaimyno ar bikubinę interpoliaciją, siekiant įvertinti jų įtaką segmentavimo tikslumui.

S.3.3. Interpoliacijos metodo įtaka

Ankstesnio eksperimento rezultatai (žr. poskyrį S.3.2), kuriame buvo nagrinėta vaizdų mastelio keitimo lygio įtaka, pabrėžia dvitiesinės interpoliacijos apribojimus išlaikant OC ribas. Nors didesnė raiška (512×512

S.3 lentelė: OD ir OC segmentavimo rezultatai pagal Dice, gauti taikant skirtingus interpoliacijos metodus ir vaizdų dydžius.

CNN ir duomenų rinkinys	**512×512 px**		**256×256 px**		**128×128 px**	
	OD	**OC**	**OD**	**OC**	**OD**	**OC**
Dvitiesinė						
UNet++						
REFUGE	0,963	0,855	0,954	0,844	0,935	0,829
RIM-ONE	0,962	0,811	0,938	0,728	0,927	0,715
DRISHTI-GS	0,964	0,855	0,599	0,849	0,944	0,824
RAUNet						
REFUGE	0,950	0,827	0,939	0,804	0,910	0,781
RIM-ONE	0,948	0,808	0,929	0,791	0,908	0,777
DRISHTI-GS	0,950	0,826	0,937	0,809	0,912	0,787
Artimiausio kaimyno						
UNet++						
REFUGE	0,961	0,858	0,949	0,848	0,934	0,822
RIM-ONE	0,961	0,836	0,937	0,769	0,913	0,659
DRISHTI-GS	0,965	0,867	0,958	0,839	0,952	0,800
Bikubinė						
UNet++						
REFUGE	**0,964**	**0,862**	**0,957**	**0,857**	**0,944**	**0,849**
RIM-ONE	**0,965**	**0,846**	**0,951**	**0,827**	**0,936**	**0,793**
DRISHTI-GS	**0,970**	**0,873**	**0,966**	**0,859**	**0,956**	**0,836**

px) pagerino segmentavimo rezultatus, bet OC ribų detalių praradimas rodo, kad dvitiesinė interpoliacija gali būti nepakankamai efektyvi erdvinei informacijai išsaugoti. Atliepiant tai, šiame eksperimente nagrinėjami alternatyvūs interpolacijos metodai, tokie kaip artimiausio kaimyno ir dvikubinė interpolacijos. Nors ankstesnis eksperimentas parodė, kad ROI mastelio keitimas į 512×512 px davė geriausius OD ir OC segmentavimo rezultatus, šio eksperimento tikslas - išsamiai ištirti kitus interpoliacijos metodus. Todėl šiame tyrime išlaikomi tie patys vaizdų mastelio keitimo dydžiai, t.y. 128×128, 256×256 ir 512×512 px. Kadangi mišraus duomenų rinkinio metodas davė perspektyvių rezultatų, jis taip pat taikomas ir šiame eksperimente, papildant duomenis REFUGE rinkinio vaizdais. Dėl to atskiri mokymo duomenų rinkiniai buvo sudaryti iš DRISHTI-GS, RIM-ONE ir REFUGE rinkinių vaizdų, kurių ROI buvo mastelio keitimu pritaikyti 128×128, 256×256 ir 512×512 px dydžius, naudojant skirtingus interpoliacijos metodus: dvitiesinę, artimiausio kaimyno ir bikubinę. Siekiant išplėsti CNN taikymo galimybes, į šį eksperimentą buvo įtrauktos trys U-Net architektūros modifikaci-

jos: Attention U-Net [54], Residual Attention U-Net (RAUNet) [45], ir UNet++ [81], pažangios objektų segmentavimo srityje. Šie modeliai buvo pasirinkti mokymui su mišrių vaizdų duomenų rinkiniu, siekiant segmentuoti OD ir OC. Pastarųjų segmentavimo rezultatai Dice įverčiu pateikiami S.3 lentelėje.

Šio eksperimento rezultatai, kurie taip pat buvo publikuoti [A.2], rodo, kad išankstinis vaizdų apdorojimas turi reikšmingą įtaką tiek OD, tiek ir OC segmentavimo rezultatams. Statistinė rezultatų analizė atskleidė reikšmingą Dice įverčių OC segmentavimo skirtumą, kai vaizdų mastelis buvo keičiamas naudojant skirtingus interpoliacijos metodus, nors OD segmentavimo tikslumas išliko bene nepakitęs nepriklausomai nuo taikyto interpoliacijos metodo.

S.3.4. OD ir OC segmentavimas akies dugno vaizduose taikant giliojo mokymosi ansamblio metodą

Nors išankstinio vaizdų apdorojimo metodai, tokie kaip interpoliacijos metodai ir vaizdo mastelio keitimo strategijos, turi didelę įtaką pavienio DL modelio veikimui, tolesnis segmentavimo tikslumo gerinimas gali būti tiriamas taikant modelių jungimo strategijas. Vietoje vieno tinklo naudojimo, modelių ansamblinio metodas, apjungiantis kelias CNN architektūras, gali sustiprinti modelio patikimumą ir gebėjimą apibendrinti informaciją. Toliau pateikiamas DL modelių ansamblio įgyvendinimas, vertinant jo potencialą pagerinti OD ir OC segmentavimo tikslumą. Modelių ansamblio mokymosi pagrindinė idėja yra tokia: skirtingi modeliai gali pasižymėti didesniu segmentavimo tikslumu skirtingose srityse, o sujungus jų prognozes, bendra sistema tampa tikslesnė ir patikimesnė. Galutinis rezultatas gaunamas agreguojant modelių išvestis balsavimo metodais:

- Daugumos balsavimas (MV) [51]
- Vidurkio balsavimas(AV) [42]
- Svertinio vidurkio balsavimas (WAV) [32]
- Vienbalsis balsavimas (UV) [70]
- Maksimalus balsavimas.

S.4 lentelė: Penkių balsavimo metodų veikimas testavimo aibėje.

Balsavimo metodas	Testavimo aibė	OD		OC	
		Dice	**IoU**	**Dice**	**IoU**
Svertinio vidurkio	REFUGE	0,956	0,916	0,886	0,795
	DRISHTI-GS	0,972	0,945	0,915	0,843
	RIM-ONE	0,975	0,950	0.895	0,810
Daugumos	REFUGE	**0,961**	**0,925**	**0,894**	**0,808**
	DRISHTI-GS	**0,974**	**0,950**	**0,916**	**0,845**
	RIM-ONE	**0,978**	**0,957**	**0,902**	**0,822**
Vienbalsio	REFUGE	0,952	0,908	0,877	0,781
	DRISHTI-GS	0,968	0,938	0,901	0,819
	RIM-ONE	0,973	0,948	0,886	0,796
Vidurkio	REFUGE	0,956	0,916	0,887	0,796
	DRISHTI-GS	0,972	0,945	0,914	0,842
	RIM-ONE	0,974	0,949	0,894	0,809
Maksimalus	REFUGE	0,957	0,917	0,890	0,801
	DRISHTI-GS	0,973	0,948	0,922	0,856
	RIM-ONE	0,971	0,943	0,894	0,809

Penki giliojo mokymosi modeliai buvo sukurti bendram OD ir OC segmentavimui, kiekvienas iš jų pagrįstas modifikuota Attention U-Net architektūra. Vietoj originalaus enkoderio buvo panaudoti skirtingi iš anksto apmokyti tinklai ResNet34, ResNet50, MobileNet, Inceptionv3 ir DenseNet121, kaip pagrindiniai tinklai, naudojami atskirai. Šie pagrindiniai tinklai buvo inicijuoti su svoriais, apmokytais 2012 m. ILSVRC ImageNet duomenų rinkinyje. Norint atskirti šias penkias modifikacijas, kiekvienam modifikuotam Attention U-Net buvo priskirtas atitinkamas modelio pavadinimas: Model-1 (ResNet34), Model-2 (ResNet50), Model-3 (MobileNet), Model-4 (Inceptionv3) ir Model-5 (DenseNet121). Iš jų trys modeliai, kurie pasiekė aukščiausius Dice koeficiento įverčius, buvo pasirinkti modelių ansamblio sukūrimui.

Rezultatai, apibendrinti S.4 lentelėje, rodo, kad daugumos balsavimo metodas pasiekė aukščiausius OD ir OC segmentavimo rezultatus vertinant Dice ir IoU įverčiais, leidžiančiais jį laikyti efektyviausiu metodu galutiniam segmentavimo rezultatui nustatyti.

OD ir OC segmentavimo rezultatai, pasiekti naudojant siūlomą ansamblio metodą, ir kitų tyrėjų rezultatai pateikti S.5 lentelėje. Siūlomas ansamblio metodas, kuris naudojo mišrių vaizdų duomenų rinkinį kelių giliojo mokymosi modelių mokymui, pranoko ankstesnius metodus, kurie buvo mokomi ir tikrinami naudojant tik konkrečios vienos duomenų

aibės vaizdus.

S.5 lentelė: Siūlomo metodo palyginimas su esamais metodais OD ir OC segmentavimui.

Metodas	**Mokymo aibė**	**Testavimo aibė**	**OD**		**OC**	
			Dice	**IoU**	**Dice**	**IoU**
CAE-BMAL [78]	REFUGE	REFUGE	0,963	-	0,879	-
		DRISHTI-GS	0,962	-	0,857	-
		RIM-ONE	0,898	-	0,791	-
Paired-Box RPN [79]	REFUGE	REFUGE	0,959	-	0,903	-
	ORIGA	ORIGA	0,963	-	0,893	-
TUNet [34]	REFUGE	REFUGE	0,961	-	0,901	-
	DRISHTI-GS	DRISHTI-GS	0,973	-	0,903	-
	RIM-ONE	RIM-ONE	0,969	-	0,862	-
PY-Net [5]	REFUGE	REFUGE	0,965	0,932	0,885	-
	DRISHTI-GS	DRISHTI-GS	0,971	0,944	0,876	-
	RIM-ONE	RIM-ONE	0,961	0,926	0,874	-
BEAC-Net [25]	DRISHTI-GS	DRISHTI-GS	0,861	0,839	0,809	0,763
	RIM-ONE	RIM-ONE	0,858	0,839	0,733	0,663
		66 Vision Tech	0,827	0,814	0,806	0,789
C2FTFNet [77]	REFUGE	REFUGE	0,969	-	0,908	-
	DRISHTI-GS	DRISHTI-GS	0,976	0,954	0,920	0,854
	DRIONS-DB	DRIONS-DB	0,968	0,936	-	-
EARDS [80]	REFUGE	REFUGE	0.955	0.915	0.887	0.802
	DRISHTI-GS	DRISHTI-GS	0.974	0.950	0.916	0.849
EE-Unet [72]	REFUGE	DRISHTI-GS	0,962	0,885	0,923	0,815
	GAMMA	RIM-ONE	0,956	0,880	0,864	0,762
Ansamblis **Pasiūlytas** [A.1]	Mišrus iš RIM-ONE REFUGE DRISHTI-GS	RIM-ONE	0,961	0,925	0,894	0,808
		REFUGE	0,974	0,950	0,916	0,845
		DRISHTI-GS	0,978	0,957	0,902	0,822

Eksperimentiniai rezultatai, pateikti [A.1], gali būti apibendrinti taip:

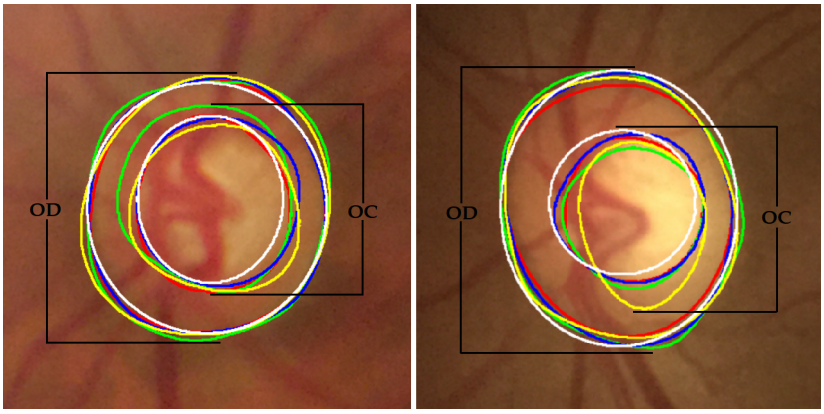
- Pasiūlytas CNN ansamblinis metodas, lyginant su vienu CNN modeliu, pagerino OD segmentavimo tikslumą pagal Dice atitinkamai 2 %, 2 % ir 7 % REFUGE, DRISHTI-GS ir RIM-ONE duomenų rinkiniuose. OC segmentavimo tikslumas pagal Dice šiuose rinkiniuose pagerėjo 3 %, 2 % ir 9 %.

- Iš penkių taikytų balsavimo metodų, sujungus trijų skirtingų

modelių prognozes, daugumos balsavimo metodas pateikė tiksliausius OD ir OC segmentavimo rezultatus.

S.3.5. Skirtingų ekspertų akių dugno vaizdų žymėjimo įtaka CNN mokymui

Viešai prieinamas akių dugno vaizdų duomenų rinkinys Chákṣu apima vaizdus, kuriuos žymėjo keli skirtingą patirtį turintys ekspertai. Pastebimi OD ir OC ribų žymėjimų skirtumai gali sukelti glaukomos identifikavimo netikslumų, o tai yra svarbu mokant CNN. Šie skirtumai gali neigiamai paveikti automatizuotų segmentavimo algoritmų veikimą. S.2 paveiksle pateikiami glaukomos ir sveikos akies atvejai, kuriuos žymėjo penki skirtingi ekspertai. Nors OS ribos paprastai yra gerai pažymėtos, pastebimi reikšmingi OC ribų žymėjimo skirtumai.



S.2 pav.: Iš kairės į dešinę: Glaukomos ir normalūs atvejai, kuriuos žymėjo penki ekspertai. OD ir OC žymėjimai išskiriami šiomis spalvomis: raudona (Ekspertas 1), žalia (Ekspertas 2), mėlyna (Ekspertas 3), geltona (Ekspertas 4) ir balta (Ekspertas 5).

Todėl šis tyrimas siekia atsakyti į kelis svarbius klausimus:

- Ar skirtingų ekspertų vertinimai statistiškai reikšmingai skiriasi?
- Ar galima nustatyti skirtingų CDR metrikų slenksčius, kurie galėtų būti naudojami kuriant automatizuotą klasifikatorių glaukomos ir sveikų akių atvejams?
- Kaip palyginti CNN, apmokytus naudojant vaizdus, žymėtus skirtingų ekspertų?

- Kuri metrika yra tinkamiausia glaukomos identifikavimui?
- Kokia skirtingų ekspertų sužymėtų vaizdų įtaka CNN mokymui?

Norint atsakyti į kiekvieną iš aukščiau pateiktų klausimų, eksperimentas buvo atliktas naudojant pavienį CNN. Kadangi ankstesniame tyrime modifikuotas Attention U-Net su iš anksto apmokytu DenseNet121 enkoderyje parodė geriausią segmentavimo rezultatą, jis buvo pasirinktas šiam tyrimui. Trys akių dugno vaizdų duomenų rinkiniai REMIDIO, FORUS ir BOSCH buvo iš anksto apdoroti ir naudojami atskirų CNN mokymui ir testavimui. Kiekvienam duomenų rinkiniui buvo sugeneruotos tikrosios anotacijos, pagrįstos penkių skirtingų ekspertų žymėjimais.

Statistinis akių dugno vaizdų žymėjimo skirtumų tarp skirtingų ekspertų reikšmingumas, taip pat jų įtaka CNN veikimui OD ir OC segmentavime įvertinama naudojant ANOVA testą ir Stjudento t-testą.

S.3.6. Skirtingų metrikų ACDR, VCDR ir HCDR slenksčių nustatymas

Šiam eksperimentui kiekvieno duomenų rinkinio akies dugno vaizdams buvo priskirtos glaukomos arba normalaus atvejo etiketės, remiantis daugumos balsavimo metodu, sumažinant subjektyvaus ekspertų vertinimų poveikį. Pagal ekspertų pateiktas ACDR, VCDR ir HCDR metrikų vertes kiekviename duomenų rinkinyje buvo įvertintas akių sveikatos stadijų vidurkis ir pasikliovimo intervalas. Remiantis statistiniais rezultatais, kiekvienos iš trijų metrikų ACDR, VCDR ir HCDR slenksčiai pasiūlyti S.6 lentelėje.

S.6 lentelė: Siūlomi akių sveikatos stadijų vertinimo slenksčiai ACDR, VCDR, ir HCDR metrikoms.

Stadija	**ACDR slenkstis**	**VCDR slenkstis**	**HCDR slenkstis**
Sveika akis	< 0,30	< 0,50	$\leqslant 0,50$
Įtariama glaukoma	[0,30 - 0,50]	[0,50 - 0,65]	(0,50 - 0,65]
Glaukoma	> 0,50	> 0,65	> 0,65

S.3.7. Metodas, skirtas CNN, apmokytų ant skirtingų ekspertų pateiktų OD ir OC anotacijų, palyginimui

Norint įvertinti skirtingų ekspertų anotacijų įtaką CNN veikimui, buvo atliktas ANOVA testas, siekiant palyginti CDR metrikas, kurių paskaičiavimui naudojami tinklų, apmokytų su kiekvieno eksperto žymėtais duomenimis, segmentuoti OD ir OC. Ši statistinė analizė buvo gauta atliekant tris atskirus testus:

- **Testas 1:** Vertinama, ar egzistuoja statistiškai reikšmingi skirtumai tarp glaukomos ir normalių atvejų.
- **Testas 2:** Tikrinama, ar tarp skirtingų ekspertų vertinimų yra reikšmingų skirtumų.
- **Testas 3:** Vertinama sąveika tarp glaukomos ir normalių atvejų, ir ekspertų anotacijų.

Rezultatai, pateikti S.7 lentelėje, naudojant reikšmingumo lygmenį $\alpha = 0.05$.

S.7 lentelė: ANOVA testo p-reikšmės* rezultatai.

Testas	**Aibė**	**ACDR p-reikšmė**	**VCDR p-reikšmė**	**HCDR p-reikšmė**
Testas 1	REMIDIO	<0,001	<0,001	<0,001
	FORUS	<0,001	<0,001	<0,001
	BOSCH	<0,001	<0,001	<0,001
Testas 2	REMIDIO	<0,001	<0,001	<0,001
	FORUS	<0,001	<0,001	<0,001
	BOSCH	<0,001	<0,001	<0,001
Testas 3	REMIDIO	0,045	0,226	0,040
	FORUS	0,203	0,149	0,298
	BOSCH	<0,001	<0,001	<0,001

* p-reikšmė, reikšmingumo lygmuo $\alpha = 0.05$ pagal ANOVA testą.

S.3.8. Ekspertų žymėjimų skirtumų įtakos tinklo mokymui įvertinimas ir tinkamiausios metrikos glaukomos identifikavimui parinkimas

Šiame eksperimente buvo įvertinta ekspertų anotacijų skirtumų įtaka CNN mokymui ir ištirtas skirtingų CDR metrikų, apskaičiuojamų pagal pastarųjų plotų santykį (ACDR) ir jų vertikalaus bei horizontalaus

S.8 lentelė: p-reikšmės* rezultatai atlikus Stjudento t-testą.

Ekspertas	ACDR p-reikšmė	VCDR p-reikšmė	HCDR p-reikšmė
REMIDIO			
Mišrus vs Ekspertas 1	<0,001	0,334	<0,001
Mišrus vs Ekspertas 2	<0,001	0,067	<0,001
Mišrus vs Ekspertas 3	<0,001	0,067	<0,001
Mišrus vs Ekspertas 4	<0,001	0,100	<0,001
Mišrus vs Ekspertas 5	<0,001	0,057	<0,001
FORUS			
Mišrus vs Ekspertas 1	<0,001	0,025	<0,001
Mišrus vs Ekspertas 2	<0,001	0,040	<0,001
Mišrus vs Ekspertas 3	<0,001	0,136	<0,001
Mišrus vs Ekspertas 4	<0,001	0,029	<0,001
Mišrus vs Ekspertas 5	<0,001	0,018	<0,001
BOSCH			
Mišrus vs Ekspertas 1	<0,001	0,150	<0,001
Mišrus vs Ekspertas 2	<0,001	0,307	<0,001
Mišrus vs Ekspertas 3	<0,001	0,159	<0,001
Mišrus vs Ekspertas 4	<0,001	0,058	<0,001
Mišrus vs Ekspertas 5	<0,001	0,060	<0,001

* p-reikšmė, reikšmingumo lygmuo α = 0.05 pagal Stjudento t-testą.

diametrų santykį (VCDR ir HCDR atitinkamai), tinkamumas glaukomos nustatymui. Siekiant spręsti skirtumus, atsirandančius dėl skirtingų ekspertų skirtingų žymėjimų, buvo sukurtas mišrus mokymo duomenų rinkinys, atsitiktinai parenkant akių dugno vaizdus kartu su atitinkamois OD ir OC anotacijomis, parengtomis skirtingų ekspertų REMIDIO, FORUS ir BOSCH duomenų rinkiniuose. Buvo atlikta kiekvieno eksperto parengtų testavimo duomenų analizė, o apibendrinti rezultatai pateikiami S.8 lentelėje. Rezultatai parodė, kad ekspertų skirtumai daro reikšmingą įtaką tinklo mokymui, ypač kai naudojamos metrikos ACDR ir HCDR. Tuo tarpu VCDR mažiausiai priklauso nuo subjektyvumo, todėl tinkamiausia glaukomos identifikavimui.

Šioje disertacijoje siūloma daugiapakopė glaukomai nustatyti nustatymui naudojant DL algoritmus. Toliau pateiktos išvados apibendrina pagrindinius šios disertacijos rezultatus ir indėlį:

- Skirtinga vaizdo skiriamoji geba ir interpoliacijos metodai turi įtakos segmentavimo rezultatui. Geriausi segmentavimo rezultatai gauti naudojant bikubinę interpoliaciją, mažinant vaizdo dydį iki 512×512 pikselių.

- Pasiūlytas CNN ansamblis pagerina segmentavimo tikslumą naudojant keletą skirtingų duomenų rinkinių vaizdų. Tai yra, OD segmentavimas pagerėja vidutiniškai 4 %, o OC vidutiniškai 5 %, vertinant Dice koeficientu. Tarp penkių taikytų balsavimo metodų, kurie efektyviai sujungia trijų skirtingų modelių prognozes, tiksliausi OD ir OC segmentavimo rezultatai gaunami daugumos balsavimo metodu.

- Statistinė analizė, naudojant Stjudento t-testą (α = 0,05), patvirtino reikšmingus glaukomos ir sveikos akies atvejų įvertinimo skirtumus dėl ekspertų subjektyvumo daugumoje duomenų rinkinių. Pastebėtas unikalus kiekvieno eksperto glaukomos ir sveikos akies atvejų vertinimas, o tai ir patvirtina standartizuotų vertinimo kriterijų poreikį.

- Remiantis konkrečiam duomenų rinkiniui būdingais skirtingų CDR metrikų vidurkiais ir pasikliautinaisiais intervalais, buvo pasiūlyti slenksčiai akių sveikatos būklei klasifikuoti į tris kategorijas:

  - Sveika akis: ACDR < 0.30, VCDR < 0.50, o HCDR $\leqslant 0.50$
  - Įtariama glaukoma: 0.30 $\leqslant$ ACDR $\leqslant$ 0.50, 0.50 $\leqslant$ VCDR $\leqslant$ 0.65, o 0.50 < HCDR $\leqslant$ 0.65
  - Glaukoma: ACDR > 0.50, VCDR > 0.65, o HCDR > 0.65

  Kategorijos „įtariama glaukoma“ įvedimas padeda anksti nustatyti ligą ir laiku atlikti intervenciją.

- ANOVA testas atskleidė statistiškai reikšmingus CNN segmentavimo rezultatų skirtumus, kai tinklai buvo apmokyti naudojant

skirtingų ekspertų OD ir OC žymėjimus. Tačiau paaiškėjo, kad glaukomos stadija iš esmės yra nepriklausoma nuo konkrečių ekspertų anotacijų. CNN tinklais, apmokytais naudojant mišrų skirtingų ekspertų duomenų rinkinį, gauti segmentavimo rezultatai VCDR vertinimu gana tiksliai sutapo su ekspertų vertintu VCDR, tačiau skyrėsi ACDR ir HCDR vertinimai, gauti tinklo ir ekspertų, o tai rodo, kad reikia toliau tobulinti segmentavimo modelius.

- Tarp įvertintų CDR rodiklių VCDR pasirodė esantis patikimiausias glaukomai nustatyti, rodantis geresnį skirtingų duomenų rinkinių rezultatų apibendrinamumą ir didesnį suderinamumą su CNN segmentacijos rezultatais nei ACDR ar HCDR.

NOTES

Sandra Virbukaitė
Methodology for Applying Deep Learning Algorithms for Glaucoma Identification
Doctoral Dissertation
Technological Sciences
Informatics Engineering (T 007)
Thesis Editor: Zuzana Šiušaitė

Sandra Virbukaitė
Giliojo mokymosi algoritmų taikymo metodika glaukomai identifikuoti
Daktaro disertacija
Technologijos mokslai
Informatikos inžinerija (T 007)
Santraukos redaktorė: Zenta Mačiulaitienė

Vilnius University Press
9 Saulėtekio al., Building III, LT-10222 Vilnius
Email: info@leidykla.vu.lt, www.leidykla.vu.lt
bookshop.vu.lt, journals.vu.lt
Print run of 20 copies