Computer Vision and Machine Learning for Glaucoma Detection

Mohammad Norouzifard

A thesis submitted to
Auckland University of Technology
in fulfilment of the requirements for the degree of
Doctor of Philosophy (PhD)

April 2020
School of Engineering, Computer and Mathematical Sciences

Abstract

Deep learning based on computer vision and machine learning is an emerging technology in both the medical imaging industry and academia. Despite the existence of some commercial glaucoma detection systems such as retinal imaging, OCT scans, and ocular tonometry, we are at the beginning of a long research pathway toward the future generation of intelligent glaucoma detection systems.

Early glaucoma diagnosis prevents permanent structural optic nerve damage and consequent irreversible vision impairment. Longitudinal studies have described both baselines structural and functional factors that predict the development of glaucomatous change in ocular hypertensive and glaucoma suspects. Although there is neither a gold standard for disease diagnosis nor progression, photographic assessment of the optic nerve head remains a mainstay in the diagnosis and management of glaucoma suspects and glaucoma patients.

This thesis discusses several image processing techniques comprising disparity map, superpixel and noise removal for pre-processing.

A stack of traditional classifiers was utilized as a hybrid model based on the ensemble method to generalise and boost the performance of the proposed model to detect glaucoma through the thickness of the retina.

A method was needed aiming at both detecting pathologic changes characteristic of glaucomatous optic neuropathy in optic disc images, and classification of images into categories of glaucomatous/suspect or normal optic discs. Therefore, different machine learning algorithms were used including transfer learning, deep convolutional neural networks, and deep multilayer neural networks that extract features automatically based on clinically relevant optic disc features. Meanwhile, biomarkers were demonstrated with the proposed deep learning model to interpret which parts of the retina had been affected by glaucoma.

Finally, this research proposes methods based on evolving deep pre-trained learning architecture, stereo matching with the usage of disparity maps, hybrid models with statistical analysis to *retinal nerve fibre layer* (RNFL) classification, and visualization of biomarkers with deep learning to detect glaucoma in early stages based on fundus images. Besides, in Appendix A; we discuss a hypothesis of glaucoma detection through detecting specified pattern with signal processing and video processing to achieve glaucoma detection at its early stages. Thus, we are going to specify the OKN pattern of eye movement to detect glaucoma at its initial stage.

Keywords

Biomedical engineering, Glaucoma detection, Deep learning, Machine learning, Transfer learning, Convolutional neural network, Traditional classifier, Ensemble model, Pattern recognition, Supervised and unsupervised classifier, Hybrid classifier, RNFL classification, Fundus image classification, Optic disc and optic cup segmentation, Vision testing, Optokinetic Nystagmus, Eye tracking.

Acknowledgement

I have often asked myself: What is a doctor is for, helping patients? Today, I finally understand that it is a degree and people who own this degree have responsibilities to contribute their knowledge to make this world go further. Also, "the only way to do great work is to love what you do." Steve Jobs.

First, I gratefully acknowledge the cast of thousands who supported my journey of discovery, in particular, my primary supervisor, **Professor Reinhard Klette**. Completing this journey would not have been possible without his professional guidance. He always supported me, not only in academic aspects to deserve a PhD degree but also in leading me to be a good person and live a better life.

I am always very thankful to my secondary supervisor **Associated Professor Hamid GholamHosseini**, for his significant contributions to all joint publications and guiding me very well. I also sincerely thank:

> Dr Anmar Abdul-Rahman (MD – ophthalmologist at Counties Manukau DHB, Auckland),
>
> Dr Jason Turuwhenua (Auckland Bioengineering Institute – University of Auckland),
>
> Dr Siamak Yousefi (Assistant Professor, University of Tennessee Health Science Centre, Memphis, USA), and
>
> Dr Kouros Nouri-Mahdavi (MD – Associate professor of ophthalmology at UCLA Medical Centre, California)

for being admirable captains guiding me through countless challenges on the knowledge pursuing adventure.

Also, this thesis would not have been done without the unconditional love of my dear mum, dad, brother, and the rest of my family. Besides, I much appreciate the support of Reinhard's research group (CeRV) at AUT. Last but most, I would like to thank my friends for the great discussions and contributions to enrich my research quality.

By the way – thank God, for his truth, mercy, grace, wisdom, and blessings.

Mohammad Norouzifard
Auckland
April 28, 2020

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning.

Signed: Mohammad Norouzifard

Date: April 28, 2020

Publications

M. Norouzifard, A. Abdollahi Dehkordi, M. Naderi Dehkordi, H.GholamHosseini, and R. Klette, "Unsupervised optic cup and optic disc segmentation for glaucoma detection by ICICA", I-SPAN, China, 2018.

M. Norouzifard, A. Nemati, H. GholamHosseini, R. Klette, K. Nouri-Mahdavi, and S. Yousefi, "Automated glaucoma diagnosis using deep and transfer learning: Proposal of a system for clinical testing", IVCNZ, New Zealand, 2018.

M. Norouzifard, A. Dawda, A. Abdul-Rahman, H. GholamHosseini, and R. Klette, "Superpixel segmentation methods on stereo fundus images and disparity map for glaucoma detection", IVCNZ, New Zealand, 2018.

M. Norouzifard, A. Nemati, A. Abdul-Rahman, H. GholamHosseini, and R. Klette, "A comparison of transfer learning techniques, deep convolutional neural network and multilayer neural network methods for the diagnosis of glaucomatous optic neuropathy", ICS, Taiwan, 2019.

L. Ding, H. Zhang, J. Xiao, B. Li, S. Lu, and M. Norouzifard, "An improved image mixed noise removal algorithm based on super-resolution algorithm and CNN", J. Neural Computing and Applications, 2019.

M. Norouzifard, A. Nemati, R. Klette, H. GholamHosseini, K. Nouri-Mahdavi, and S. Yousefi, "A hybrid machine learning model to detect glaucoma using retinal nerve fiber layer thickness measurements", Investigative Ophthalmology & Visual Science, 60(9):3924, 2019.

M. Norouzifard, A. Nemati, R. Klette, H. GholamHosseini, K. Nouri-Mahdavi, and S. Yousefi, "Identification of clinically relevant glaucoma biomarkers on fundus images using deep learning", Investigative Ophthalmology & Visual Science, 60(11):PB090, 2019.

M. Norouzifard, A. Nemati, A. Abdul-Rahman, H. GholamHosseini, and R. Klette, "A Fused Pattern Recognition Model to Detect Glaucoma Using Retinal Nerve Fiber Layer Thickness Measurements" Accepted to Conf. PSIVT, Australia, 2019.

M. Norouzifard, J. Black, B. Thompson, R. Klette, and J. Turuwhenua "A state machine for real-time eye tracking to detect optokinetic nystagmus" Accepted to Conf. ACPR, Auckland, 2019.

Contents

List of Figures

List of Tables

List of Symbols

General Notation	Description
$\star$	Denotes the convolution operator
$\cdot$	Multiplication of two scalars where it is read as "times" or "multiplied by"
A	The area of an image region is identified with the number of pixels contained in this region
$f(\cdot)$	Denotes a function
H	Horizontal diameter is measured in number of pixels
V	Vertical diameter is measured in number of pixels

Deep Learning Variables

$b_k^{(n)}$	Bias element for the k-th output-feature-map of the n-th layer
$\mathbf{G}^{(n)}$	The input and output matrices for the n-th layer of a CNN which are composed of a finite number of feature maps
$\mathbf{I}$	The unit matrix
$M_i^{(n)} \times M_i^{(n)}$	Size of the input of the n-th layer
$M_o^{(n)} \times M_o^{(n)}$	Size of the output of the n-th layer
$P_i^{(n)}, P_o^{(n)}$	The number of input and output maps, respectively, for the n-th layer
$\mathbf{w}_{ik}^{(n)}$	Kernel linking the i-th input map to the k-th output map

Clustering Variables

μ_i	Centre of the i-th cluster
$\|x_j - \mu_i\|$	Euclidean distance of the j-th pattern from the centre of the i-th cluster
$\{x\} = \{x_1, ..., x_n\}$	Given data set of n patterns
k	Number of clusters

Chaotic Imperialist Competitive Variables

a, b, β, d	Scalars
$\mathbf{A}$	Matrix

$\{V_1\}$	Vector
$\{cm_k\}$	k-th chaotic vector
$\otimes$	Denotes an element-by-element multiplication
$\{\theta_k\}$	The k-th vector of θ values that will be randomly in the range of $(0, 1)$ at each iteration

Stereo Vision and Superpixel Variables

d	Disparity map
$p = (x, y)$	The base image is displayed by pixel p
$q = (x + d, y)$	q is the corresponding pixel of pixel p
$\mathbf{c}_i$	Initial cluster centre
s	Roughly equally sized superpixels
E	Residual error

Statistical Analysis Variables

$\mathbf{x}$	Datapoint from the data set
z	Z-score is a numerical measurement
μ	Mean of data
σ	Standard deviation of data

List of Abbreviations

Acronym	Definition
ACC	*accuracy*
AI	*artificial intelligence*
AMD	*age-related macular degeneration*
AUC	*area under the receiver operating characteristic curve*
CDR	*cup-to-disc ratio*
CPU	*central processing unit*
CSLO	*confocal scanning laser ophthalmoscopy*
CNN	*convolutional neural network*
DC	*dice coefficient*
DDLS	*disc damage likelihood scale*
FDT	*frequency-doubling technology*
FN	*false negative*
FP	*false positive*
FPGA	*field programmable gate array*
GA	*genetic algorithm*
GCIPL	*ganglion cell-inner plexiform layer*
GCL	*ganglion cell layer*
GPU	*graphics processing units*
GRI	*glaucoma risk index*
ICA	*imperialist competitive algorithm*
ICICA	*improved chaotic imperialist competitive algorithm*
INL	*inner nuclear layer*
ISNT	*inferior superior nasal temporal*
IOP	*intraocular pressure*
IPL	*inner nuclear layer*
IU (IOU)	*intersection over union*
JC	*Jaccard coefficient*
KDE	*kernel density estimation*
KNN	*k-nearest neighbours*
MSE	*mean squared error*
NVCom	*Nerian Vision Technologies*
OCT	*optical coherence tomography*
OC	*optic cup*

OD	*optic disc*
OKN	*optokinetic nystagmus*
ONH	*outer nuclear layer*
ONL	*optic never ahead*
OPL	*outer plexiform layer*
PCA	*principal component analysis*
PP	*positive predictive*
RMSE	*root-mean-square error*
RNFL	*retinal nerve fibre layer*
ROI	*region of interest*
SAP	*standard automated perimetry*
SE	*sensitivity*
SGM	*semi-global matching*
SLIC	*simple linear-iterative clustering*
SLP	*scanning laser polarimetry*
SMO	*sequential minimal optimization*
SNIC	*simple non-iterative clustering*
SODI	*stereo OD image*
SP	*specificity*
SVD	*singular-value decomposition*
SVM	*support vector machine*
SWAP	*short-wavelength automated perimetry*
TC	*traditional classifier*
TN	*true negative*
TP	*true positive*
TPU	*tensor processing unit*
UCLA	*University of California Los Angeles*
VF	*visual field*

Chapter 1

Introduction

This chapter provides a brief background of image processing and machine learning on glaucoma detection and motivations. The subject and the structure of the thesis are also specified in this chapter. Additionally, the problem statements and research questions are presented in this chapter of the thesis. The flowchart in Fig. 1.1 represents the structure of the thesis on glaucoma detection.

Moreover, we applied machine learning techniques to detect glaucoma with three different aspects; we employed transfer learning for image classification to detect glaucoma, and a stack of traditional classifiers to make a robust and generalise model for glaucoma detection with the thickness of the retina. Also, we utilized OKN pattern recognition as a specific pattern identification via pupil tracking and video streaming technologies. Besides, the schematic view for glaucoma classification of our research is illustrated in Fig. 1.2.

1.1 Background

In this section, we[1] present a brief discussion of glaucoma vision impairment and the reason why it occurs. Also, the usage of image processing and machine learning to predict glaucoma including a brief theoretical background of those machine learning techniques that we will use in the following chapters of the thesis.

Glaucoma

Glaucoma is an optic neuropathy resulting in progressive vision loss [279]. It is the leading cause of global irreversible blindness [256]. In 2013, the population aged 40–80 years with glaucoma worldwide was estimated to be 64.3 million; this is projected to increase to 76.0 million by 2020 and 111.8 million by 2040 [240]. It is also estimated that approximately 1.3 billion people live with some form of vision

[1]The use of "we" throughout this thesis is purposeful. It is used to involve the reader with the thesis as recommended by Knuth [140].

impairment. Glaucomatous optic neuropathy is listed as the fourth major cause of vision impairment by the *world health organisation* (WHO). In 2015, an estimated three million people went blind due to this disease.

In New Zealand, glaucoma is responsible for 7% (1,192 patients) of cases of bilateral blindness, ranking the third most common reported etiology [279]. The reported prevalence among the population in New Zealand is 2% over the age of 40 years. About 10% of those over 70 years are diagnosed with this disease [279]. Early detection is vital to reduce the burden of unnecessary blindness due to glaucoma. The Royal Australian and New Zealand College of Ophthalmologists recommends a biannual ophthalmic examination, and the New Zealand Association of Optometrists recommends a regular examination every 2–5 years for healthy adults. However, like other developed countries, New Zealand does not have a formal screening programme for glaucoma [144]. Diagnosis and subtype classification are based on intraocular pressure measurement, gonioscopy, and the presence of both structural and functional evidence of glaucomatous optic neuropathy [81, 219]. Intra- and inter-observer settlement in the detection of *optic disc* (OD) abnormalities could be improved with stereo fundus images [248].

The retina is a light-sensitive neural tissue that forms the innermost part of the posterior eyewall and consists of a thin multi-layered tissue. It maps received light

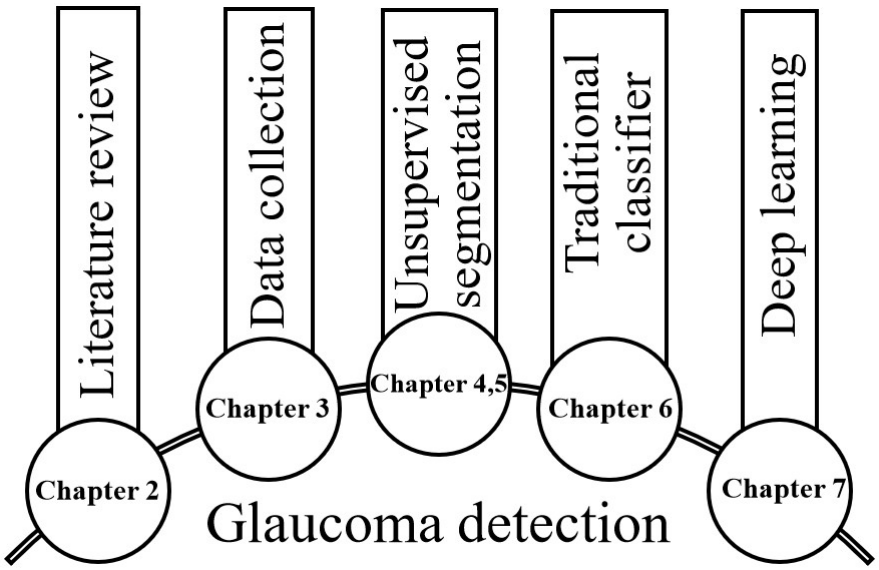


Figure 1.1: Flow chart of the thesis.

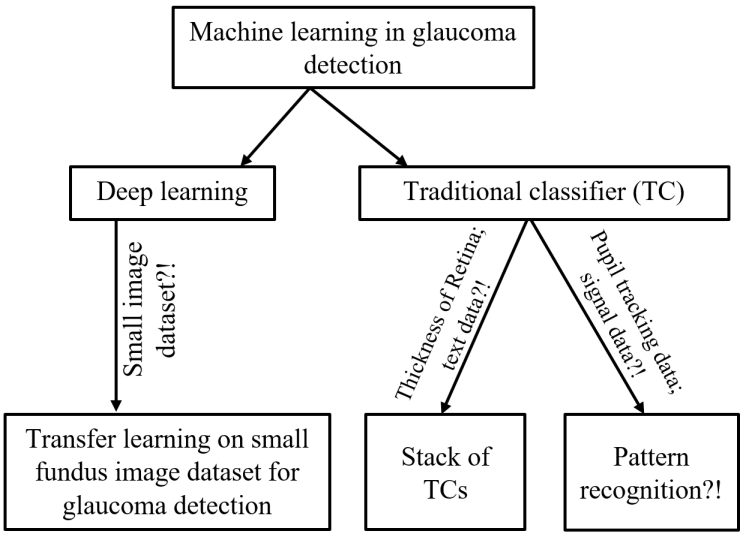


Figure 1.2: Schematic view for glaucoma classification of our research.

into neural signals to the brain, thus being the initiator of visual perception [183]. Glaucoma refers to an eye disease that manifests as slow damage to the optic nerve. The optic nerve is in charge of transmitting visual information to the brain and optic nerve damage gradually causes reduced vision. If glaucoma is not treated, the disease progresses slowly and results in varying degrees of irreversible visual disability and, in some cases, blindness [258]. Loss of vision begins with side vision loss and approaches towards the centre of the eye, as illustrated in Fig. 1.3.

Glaucoma is a disease that damages the eye's optic nerve, and it has no warning signs at the beginning, thus often being named the "silent thief of sight".

Also, glaucoma is a chronic neurodegenerative disease characterised by loss of

Figure 1.3: Illustration of decreased side vision due to a gradual progression of glaucoma.

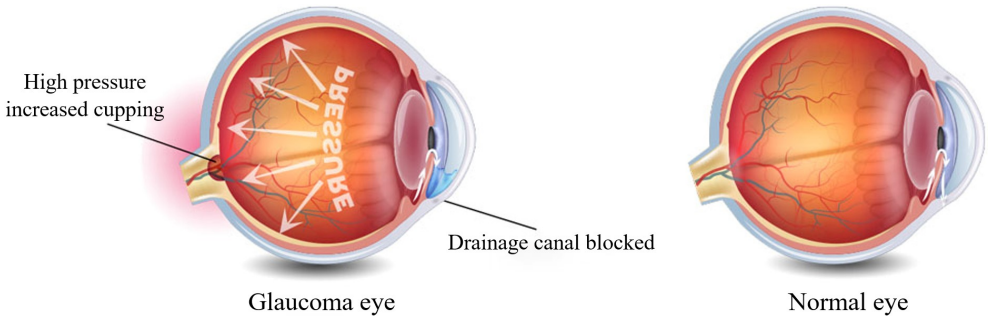


Figure 1.4: Glaucoma eye vs. healthy one [277].

retinal ganglion cells, resulting in distinctive changes in the *optic nerve head* (ONH) and *retinal nerve fibre layer* (RNFL) [219]. It is the leading cause of global irreversible blindness.According to Glaucoma New Zealand [287] statistics, approximately 91,000 New Zealanders aged 40 or older currently suffer from glaucoma while about 50% of them are unaware of having the disease.

The high intra- and inter-observer variability in clinical examination [90] demands the use of ancillary techniques to assess the structure of the ONH and RNFL such as optic disc photography, confocal scanning laser ophthalmoscopy, scanning laser polarimetry, and optical coherence tomography. Similarly, selective perimetry techniques, including *short-wavelength automated perimetry* (SWAP) and *frequency-doubling technology* (FDT) perimetry, are being explored as replacements to *standard automated perimetry* (SAP) to provide earlier detection of visual field deficits [219].

Glaucoma usually happens when fluid builds up intraocularly. That extra fluid increases the intraocular pressure, damaging the eye's optic nerve. The high intraocular pressure is caused by the increasing amount of aqueous humor, and it is one of the reasons for glaucoma disease (see Fig. 1.4). In a healthy eye, a balance is maintained as the amount of liquid produced is equal to the amount of liquid discharged by the eye. However, in glaucoma the liquid does not flow out of the eye and increases stress on the eye resulting in damage to the optic nerve which is responsible for brain-eye communication. Increase in pressure over time results in severe destruction to the optic nerve and may end up with irreversible blindness [210].

As we can see in Fig. 1.4, the eye produces aqueous humor or vitreous fluid that constantly feeds the ganglion cells of the eye, and this fluid is constantly moving through the eyeball and the excess is removed from the angel of the eye. If the amount of aqueous humor gets higher than the healthy eye, eventually the person

becomes blind. Thus, the blocked drainage canal causes too much fluid in the eye and increases pressure. This section is described in Chapter 2.

Machine Learning and Image Processing

Research has been done using state-of-the-art supervised and unsupervised machine learning techniques such as deep pre-trained models, convolutional neural networks and traditional machine learning models such as *support vector machine* (SVM), decision tree, *K-nearest neighbours* (KNN) and K-means clustering.

Most of the image processing techniques for glaucoma detection try to detect the structural features such as the cup-to-disc ratio (CDR) and/or non-structural features such as texture and intensity of fundus images. There are also some favourable techniques to detect the CDR morphological operations based on multi-threshold techniques, active contour models, region growing segmentation, and disparity map techniques on stereo fundus images. This section is discussed in Chapter 2.

Motivations

This research aims to develop an autonomous glaucoma detection tool to detect glaucoma at its early stages and assist ophthalmologists in the diagnosis of glaucoma using different data modalities such as monocular, stereo fundus images, and thickness of the retina. Moreover, by employing the computational retina biomarker modelling with the latest transfer learning techniques, ophthalmologists will be able to develop a new way of predicting clinical outcomes.

This PhD project was motivated by the limitations in the glaucoma classifications used in recent academic publications (2013-2019). One of the constraints is accessing medical annotated data that is extremely costly, consequently we attempt to find a solution to have a deep learning model to classify glaucoma patients at its early stages with a tiny dataset.

In addition, technically, there is a motivation behind why it is imperative to detect glaucoma in the initial stages. The reason is finding a model to choose more precise in time and performance than ophthalmologist's choice with the equivalent data. In this manner, we chose a deep pre-trained model as a tool since it imitates the human brain. We started the development of a computer aid application to cope with this problem with the inspiration of brain function. Besides, apart from supporting the global concern and motivation to deal with this disease, the author has a personal motivation to contribute to this field as his father has suffered from this disease for a decade.

Goals and Objective

The main goal of the thesis was to develop a computer-aided diagnostic model for glaucoma detection and propose a decision support system for ophthalmologists. Moreover, we aimed to detect glaucoma at its early stages using the same dataset. After employing deep learning techniques, the model was tested on both private and public annotated datasets. Besides, the objectives of this study were to perform a reliable segmentation on fundus images, improve the performance of glaucoma image classification on limited data when annotated fundus images are limited by utilising deep transfer learning models.

Another outcome of the project was the innovative OKN extraction through eye-tracking with data streaming to identify eye disorders in the early stages, such as glaucoma and lazy eyes for kids [70, 176].

1.2 Problem Statements and Research Questions

According to the census of the Glaucoma New Zealand website, glaucoma is the leading cause of blindness in New Zealand, and it is estimated that approximately 91,000 New Zealanders are unaware of having the disease [287]. It is estimated that there will be approximately 80 million people worldwide affected by glaucoma by 2020 [7]. In 2010, glaucoma affected more than 2.7 million Americans aged 40 and older, which is approximately 2% of the population [314]. Glaucoma is the third cause of blindness in New Zealand [279].

The research questions of this study are listed as follows:

> How can we segment the OD and OC on fundus images as a step of unsupervised annotation? (discussed in Chapter 4)
>
> How can we detect changes on the retina in rim section of ONH with a stereo image? (discussed in Chapter 5)
>
> How can we find a method for glaucoma screening to detect glaucoma in the early stages? (discussed in Chapter 6)
>
> How can a convolutional machine learning model be optimized for glaucoma detection? How can we cope with the lack of annotated data to have a high-performance machine learning classifier? How can we develop a more robust deep learning model for glaucoma detection? How can reliability be optimized for the proposed model? How can we validate the classification performance rate of the implemented classifier for glaucoma detection? (discussed in Chapter 7)

How can we use the eye-tracking to detect OKN pattern? (discussed in Appendix A)

1.3 Structure of the Thesis

This thesis is organised in eight chapters. Chapter 2 provides an analytical and critical survey of the literature using a systematic review approach and presents the theoretical concepts of deep transfer learning and traditional classifiers. Chapter 3 describes the data collection and difficulties in accessing medical data and discusses the public datasets as a benchmark for our research. Chapters 4 to 7 present the author's contribution to the field. These chapters include three main contributions. Chapters 4 and 5 are about fundus image segmentation with unsupervised models for monocular and apply disparity map on fundus stereo images to magnify the rim area. Chapter 6 presents a novel optimized hybrid model with traditional machine learning classifiers based on an ensemble model. Chapter 7 discusses deep learning models on deep neural networks, deep convolutional neural networks and deep pre-trained models. The last Chapter 8 presents discussions, conclusions and future works and gives an overall conclusion of the thesis as well as highlighting the original and significant contributions. Finally, some key future research directions are mentioned. Moreover, Appendix A is the assumption of using eye-tracking for glaucoma detection through detecting a specific pattern which is named OKN.

1.4 Original Contributions

The objective of this research is to design, implement, and evaluate a novel computer-aid application based on machine learning algorithms that can be applied as a decision support system for glaucoma specialists. We propose a unique multimodal dataset to be employed for an optimized, robust, and intelligent classifier.

This work is aimed to address the shortcoming of deep learning technique in dealing with small datasets by proposing a model which optimizes the time and memory usage. After pre-processing, noise reduction and data cleaning, boosting and ensemble techniques are employed to discover the correlation between multi-modal data and optimization of the classifier. The main contributions of this study are:

- Conducting a comprehensive survey of the literature covering both traditional

and deep machine learning technologies as well as the implementations of deep pre-trained classifiers.

- Development of an optimized deep learning model, as a decision support system, for assisting eye specialists.
- Optimization of the proposed classifier using the ensemble technique.
- Proposing an unsupervised segmentation model for fundus images.
- Implementation of a novel superpixel method on stereo fundus images to detect the rim area with a disparity map model.

Chapter 2

Review and Theoretical Background

A survey has been done on detecting glaucoma using traditional classifiers, conventional machine learning and emerging deep learning. Deep learning methods have outperformed machine-learning techniques in different domains in recent years. This chapter discusses the past and current impact of machine learning on glaucoma detection. We describe how glaucoma symptoms can be identified by machine learning and provide a brief overview of deep learning approaches, in particular of convolutional neural networks (CNNs). The emerging CNN approaches have been used widely for solving different computer-vision problems, and especially, for analysing retinal images. We summarise pre-trained CNN architectures for glaucoma detection. We then outline the effectiveness of different types of deep CNN architectures, some of which perform well for automated glaucoma classification. We present a brief history of deep learning techniques and provide critical analysis of methods and applications in glaucoma detection followed by current challenges and future research directions.

2.1 Artificial Intelligence and Glaucoma Research

Artificial intelligence (AI), particularly deep learning, is transforming medicine. The depth and impact of this transformation are likely to increase rapidly, and many physicians and health care organisations are unaware of the degree to which medicine will be fundamentally reshaped over the next generation.

Due to this advancement and growth in available databases, identification of ocular disorders including glaucoma have been enhanced significantly [241].

Deep learning is a machine learning approach that makes use of several neural networks, being subsequently used for transforming extensive input data into short characteristics [56]. The technique can be modelled by a graph that has several *processing layers*, each defined by many neurons (mathematically defining a linear or non-linear layer function). *Convolutional neural networks* (CNNs) are particularly effective for solving image analysis tasks [72]. In this chapter, we

review advances in CNN architectures [146] in the context of glaucoma detection [249].

Although there are already some dedicated surveys on applications of deep learning for glaucoma detection [103, 156, 239], covering a considerable amount of work, essential areas in the field, such as *distinct deep learning frameworks*, their advantages and disadvantages have not yet been included. This chapter aims at providing a comprehensive survey of the latest developments in the application of deep pre-trained conventional models in medicine, particularly for glaucoma detection. This review also describes publicly available and labelled glaucoma fundus image databases, highlights different performance evaluation measures commonly used within the field, and outlines current approaches in feature-based fundus image classification. The motivation for our review is to offer a comprehensive overview of techniques for glaucoma detection. We also include our own experience with the application of combinations of traditional models and deep-transfer-leaning architectures to detect glaucoma [67, 177, 178, 180–182], and with an unsupervised learning method for optic cup and optic disc segmentation for glaucoma detection [175].

We chose preferred reporting items for systematic reviews and meta-analyses (PRISMA) as the systematic review methodology [168]. An extensive search was conducted using the most reliable scientific databases, including Google Scholar, Scopus, ACM and PubMed. The keywords used in the search were "Review/survey" or "Convolutional neural network" or "Deep learning", and "Glaucoma detection" or "Glaucoma classification". Additionally, all selected papers were graded on the basis of glaucoma classification with deep learning approaches.

Article Selection and Exclusion Criteria. After conducting an initial screening of titles and abstracts of the retrieved papers, we eliminated duplicate articles that were cited with different databases, such as Scopus, arXiv, IEEE, and Springer. Moreover, additional records were excluded after reviewing the included studies and an evaluation of full-text articles for eligibility. The eligibility criteria for inclusion in the review were as follows:

- Original articles published as a journal article or in conference proceedings
- publication or reporting year between (inclusive) 2015 and 2019
- deep pre-trained conventional models are the primary subject of the reported study

- focusing on glaucoma classification with deep learning approaches only
- written and published in English

We excluded articles that did not report original research, such as letters to an editor, comments or reviews. Initially, over 500 studies were identified through database searching. After excluding duplicated records, 450 records were eligible for screening. A total of 150 records did not meet our inclusion criteria based on the initial screening. Finally, 300 studies were included to be evaluated for eligibility. After excluding irrelevant studies, 180 articles were selected for the final review. Note that the references at the end of this thesis also include sources pointing to techniques or definitions considered of relevance while reviewing the selected articles.

2.2 Structural Biomarker within Glaucoma Disease

Glaucoma is the second leading cause of blindness worldwide, with almost 60 million cases reported worldwide in 2010, and an increase of 20 million is expected by 2020 [17, 197]. It is also the third-ranked cause of blindness in New Zealand [279]. If glaucoma remains unnoticed, it can cause irremediable damage to the optic nerve leading to blindness [217]. Therefore, diagnosing glaucoma at its early stages is essential [17]. According to the census of the "Glaucoma New Zealand" website, glaucoma is the leading cause of blindness in New Zealand and it is estimated that 91,000 New Zealanders have the disease but are not aware of it [287].

Because of the rapid ageing of the human population, accurate diagnosis is critical for establishing a treatment to preserve vision and maintain quality of life [123, 126, 240]. *Intraocular pressure* (IOP) is a major risk factor and is the fluid pressure inside the eye [5, 99]. Typically, a high IOP (greater than 24 mmHg) leads to glaucoma.

The optic cup and disc are major retinal areas for glaucoma detection. The *optic disc*, often identified (see more detailed comments below) with the *optic nerve head* (ONH), is the beginning of the optic nerve, i.e. the area where the nerves of retinal cells come close to each other. It is the entering point of blood vessels that pass blood to the retina. The *optic cup* is the central depression of variable size present on the optic disc.

See Fig. 2.1 for an optic cup and optic disc example on a fundus image. The optic disc region in Fig. 2.1 is divided into superior, nasal, inferior, and temporal

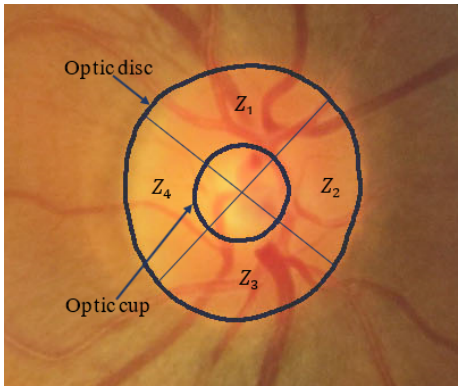


Figure 2.1: Example of the optic cup and optic disc regions in a fundus image; Z_1 to Z_4 denote the superior, nasal, inferior and temporal regions of the ONH, respectively [18].

quadrants [161] for the right eye (*oculus dextrus*) in a clockwise orientation. In glaucomatous eyes, most early signs occur in the superior and inferior sectors; close examination of these regions facilitates diagnosis of glaucoma by ophthalmologists [55].

The structural biomarker *cup-to-disc ratio* (CDR), has been used as a critical outcome measure for glaucoma screening and detection [17, 86]; there are also other essential features of the optic nerve (besides CDR) that could be used for accurate glaucoma detection, see [78]. The CDR can be defined by the area quotient of the optic cup and optic disc:

$$A_{\mathrm{CDR}} = \frac{A_{cup}}{A_{disc}} \tag{2.1}$$

The area of an image region is identified by the number of pixels contained in this region [139].

The CDR has previously been measured by estimating the ratio of the horizontal or vertical optic cup and optic disc diameters, formally expressed, respectively, by

$$H_{\mathrm{CDR}} = \frac{H_{cup}}{H_{disc}} \qquad \text{and} \qquad V_{\mathrm{CDR}} = \frac{V_{cup}}{V_{disc}} \tag{2.2}$$

For combined use of V_{CDR} and H_{CDR}, see [79, 97]. Horizontal or vertical diameters H_{cup}, ..., V_{disc} are also measured in numbers of pixels, but best along digitised lines [139].

The measure A_{CDR} is a 2-dimensional geometric property, therefore more robust for expressing the CDR, especially in the cases of "non-circular" optic cups

or optic discs. In the case of non-circular (such as elliptical) optic cups or optic discs, a combined use of vertical and horizontal ratios, or, even best-fitted ellipses may be preferable [155, 260].

Accurate CDR measurements require precise optic cup and optic disc segmentation in a fundus image [84]. To be precise, the ONH is defined by the region where the retinal ganglion cell axons and blood vessels leave the eye; clinically, it has been characterised by a very narrow rim of the sclera, called the *scleral rim*. Recent work by Burgoyne and Chauhan has demonstrated that what clinicians identify as the clinical optic disc margin is not a unique and uniform anatomic structure. Depending on the eye, the 'clinical' disc margin is formed by the Bruch's membrane opening or border tissue of Elschnig [50]. The *optic cup* is the central region in the ONH with no nerve fibres, usually appearing as being whitish or very bright.

This common definition of optic disc and optic cup is not identifying those pixels uniquely in a fundus image which belong to optic disc or optic cup. Thus, a "precise segmentation" is already a somehow vague intention.

The presence of glaucoma can be identified by an enlargement of *optic nerve cupping* (i.e., an optic cup that is increased in size), which is secondary to thinning the neuroretinal rim; see [217]. An eye with $A_{CDR} > 0.6$ is generally declared as being a glaucoma suspect in clinical practice [14, 217]. Other methods for detection of glaucoma include assessment of *optical coherence tomography* (OCT) and *perimetry*, the latter being the standard test of visual function for glaucoma analysis [80, 215].

Retinal nerve fibres are typically represented by the annular region between optic disc and optic cup, known as the *neuroretinal rim* [40, 171, 197]. This loss of optic nerve fibres results in thinning of the *retinal nerve fibre layer* (RNFL), which is characterised by increased cupping. Progressive cupping (or enlargement of the cup) is a frequent sign of glaucoma progression [40, 45]. A decrease in healthy neuroretinal tissue can be easily noticed by measuring the CDR, which is an indication of glaucomatous change [45]. Typically, for a healthy eye the area CDR value of Eq. (2.1) is ≈ 0.3 or less; see [83, 99]. Figure 2.2 shows variances in the optic disc, neuroretinal rim, and optic cup.

The CDR measures remain essential biomarkers for the diagnosis of glaucoma; ONH imaging is carried out for estimating the CDR. Other parameters such as the *inferior superior nasal temporal* (ISNT) rule, the *disc damage likelihood scale* (DDLS), and the *glaucoma risk index* (GRI) have been used widely for the diagnosis of glaucoma, also using ONH imaging [169, 239]. A pallid disc, from orange to pink in the disc, is another biomarker of glaucoma [239].

A stepwise approach for systematic identification of glaucoma on fundus images has been described for clinicians in [79]. First, the clinician identifies the scleral ring

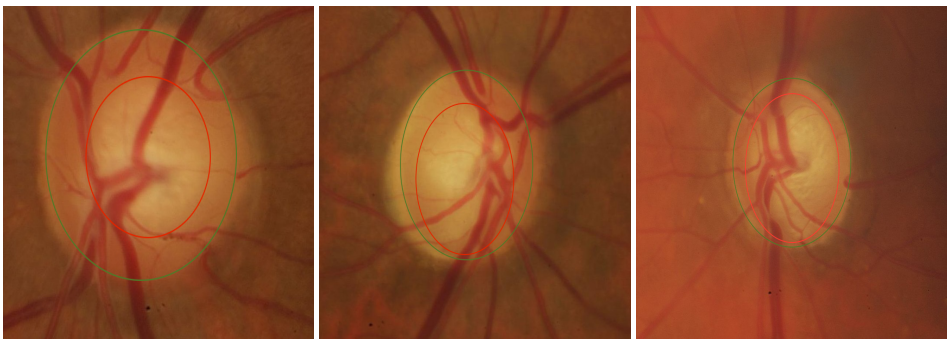


Figure 2.2: Healthy disc in comparison with two stages of glaucoma on fundus images. *Left:* Healthy disc. *Middle:* Moderate glaucoma. *Right:* Advanced glaucoma [88, 292].

to determine the outer boundary of the optic disc and its size; second, estimation of the size of the neuroretinal rim; third, the examination of the retinal nerve fibre layer; fourth, the examination of the region of para-papillary atrophy and finally, detection of optic disc haemorrhages. However, this process is subjective and prone to bias.

2.3 Current Clinical Methods for Glaucoma Detection

There is a variety of diagnostic tools for determining whether a patient has glaucoma or not [192, 307].[1] Currently, glaucoma is often diagnosed by *standard automated perimetry* (SAP) assessment of the *visual function* (VF) [124, 173, 255]. SAP is a subjective psychophysical test which requires patient interaction to complete the VF testing [12]. The result is a map of local retinal sensitivities relative to fixation [20, 41]. Detecting glaucoma at an early stage using VFs remains challenging because VFs present a high degree of variability [167].

There is some evidence that suggests that detectable glaucomatous axonal loss precedes detectable VF impairment [198, 211]. Several technologies such as *scanning laser polarimetry* (SLP), *confocal scanning laser ophthalmoscopy* (CSLO), and *optical coherence tomography* (OCT) are used to detect axonal loss [227].

OCT is a non-invasive, non- contact method of generating high-resolution images using the coherence property of light reflected from a scanned area [142]. The working principle of OCT is like ultrasound, but OCT has a higher resolution

[1] The *Glaucoma Foundation* [307] was established in 1984 by Dr R. Ritch.

that discriminates quite well between healthy and glaucomatous eyes. OCT has received more interest in glaucoma diagnosis currently. Nevertheless, its role in detecting any glaucomatous progression is not well established [33].

Fundus photography is another simple and non-expensive technology appropriate for recording the optic nerve, independent of a specialised viewing instrument. Fundus photos remain one of the most widely used methods for documentation of the ONH [38, 171, 239]. Since qualitative observation of the ONH is time-consuming and subject to inter-observer variability, from a clinical perspective, access to a more objective analysis for glaucoma diagnosis is preferred. Recent advances in deep learning and transfer learning, and the significant growth in available fundus photographs allows identifying ocular pathologies due to glaucoma manifesting in ONH fundus photographs and has led to enhanced glaucoma diagnosis which is discussed in the following sections; see also [6, 52, 53, 200].

2.4 General Background of Machine Learning

Supervised learning and unsupervised learning are two primary machine learning categories. Supervised learning employs identified input and output variables with a formula that maps process from the input to the output data. The aim of supervised learning is therefore, to able to classify and predict the mapping function so that any new input data can be classified according to the output variables for that data. Likewise, supervised learning can be utilised to classification and regression.

Unsupervised learning involves modelling the underlying structure and the hidden data distribution to be told extra concerning the data. Unsupervised learning just requires input data and no corresponding output data, and we can use it for clustering and association.

Overview of Deep Learning and CNNs

Deep neural networks employ supervised learning in their tasks, and they are adequately similar to ordinary neural networks. The name of neural networks conveys brain-simulation learning. Besides, deep neural networks comprise of neurons that have learnable weights and biases.

The name of "CNN" indicates that the network utilises a mathematical operation called convolution as a linear operation. CNNs are simply neural networks, and they use convolution operation in place of general matrix multiplication in at least one of their layers [99]. CNNs also employ an activation

function (e.g. relu, softmax) on the last fully connected layer. Moreover, all regularisation steps including drop-out, deducting layers, early stopping, and data augmentation utilise to improve the learning rate and avoid over-fitting.

An assumption for CNN models with input images is that; these models can extract features automatically by convolution operator and multiple filters within convolutional layers. There are three positive points of this assumption, first, implement a model more accurate than normal neural networks, second, do not need to extract features manually, third, vastly reduce the number of hyper-parameters in the network. This will be explained in greater detail in the next section.

2.5 Machine Learning for Glaucoma Diagnosis

Machine learning can determine markers for diagnosis of different diseases by combining input features in a linear or non-linear fashion [21, 32, 36, 39, 108, 203, 262, 263, 265]. These methods provided an objective approach for glaucoma diagnosis [33, 37, 44, 63, 264, 266]. Conventional machine-learning approaches typically use structured data such as demographic parameters, ocular measurements, VF parameters, or RNFL thickness measurements for glaucoma diagnosis or prognosis [21, 166]. Some conventional machine learning techniques have shown acceptable precision in glaucoma diagnosis [21, 37, 199].

However, emerging deep learning approaches have advanced automated diagnosis significantly. A substantial breakthrough in machine learning, introduced in 2006 [107], has been the *deep belief network*, which incorporates multiple layers of restricted Boltzmann machines to learn highly-complex non-linear models for discriminating features from different classes. These can train one layer of a neural network at a time in an unsupervised way. The main principle behind these series of developments was guiding and training intermediate levels of representation using unsupervised learning, performed locally at each level. This idea has revolutionised deep architectures and deep learning algorithms in the last decade. Deep learning methods can learn from unstructured data such as fundus images without requiring a separate feature extraction step. In the following section, we described traditional, conventional, and deep learning models that applied to glaucoma diagnosis in more detail.

Selected Traditional Supervised Classifiers for Detecting Glaucoma

There are multiple studies on supervised classifiers that apply for glaucoma detection [22, 66]. In this section, we discuss selected traditional classifiers which employ in our research to deal with glaucoma detection's problem.

Principal component analysis. PCA is an unsupervised method for understanding global attributes of a dataset consisting of vectors. A covariance matrix of data points is analysed to understand what dimensions or data points are more important, for instance, data points having high variance amongst themselves, but low covariance with others. One way to calculate the top principal components of a matrix is by using eigenvectors with the highest eigenvalues. *Singular-value decomposition* (SVD) is essentially a way of calculating ordered components also, but one does not need to acquire the covariance matrix of points [220].

K-nearest neighbour. KNN is an instance-based supervised learning method that identifies the class (label) of an unknown sample based on the known labels of its neighbours in the feature space. Euclidean distance utilises in KNN algorithm as the distance measurement metric. It is the typical distance function used for measuring the distance between samples. The main advantage of KNN is its simplicity and, in some cases, its efficiency [117]. KNN employed for glaucoma diagnosis in several studies [93, 224]. Besides, the performance of different classifiers including naíve-Bayes, KNN, SVM, and random forests were compared with KNN algorithm's performance as a benchmark [93].

Support vector machine. SVM is, besides naíve-Bayes, another widely-used conventional classifier. In a typical diagnosis problem, the SVM finds a hyperplane that maximizes the distance between the hyperplane and the nearest samples of the healthy and the disease class in the feature space. The SVM is a supervised classifier in which the labeled samples are employed in the training step. Different kernels such as linear and Gaussian can apply for the SVM to feature engineering [132, 133, 242]. *Sequential minimal optimization* (SMO) has used in SVM for glaucoma diagnosis. SMO is an optimization method that can help us to obtain a noteworthy accuracy.

In addition, *support vector classier* (SVC)/ SVM is a machine learning model that generates splits (hyperplanes) in the data to acquire a better interpret in the new unknown data. In Fig. 2.3, the hyperplane splits the data optimally into two, and then new data can be properly classified-it would be blue above and red below. The ideal split would not be a line in higher dimensions (not 2D), but hyperplanes.

NuSVC classifier. Support Vector Machine with Nu-algorithm classification is similar to SVC, but a parameter is used to control the number of support vectors.

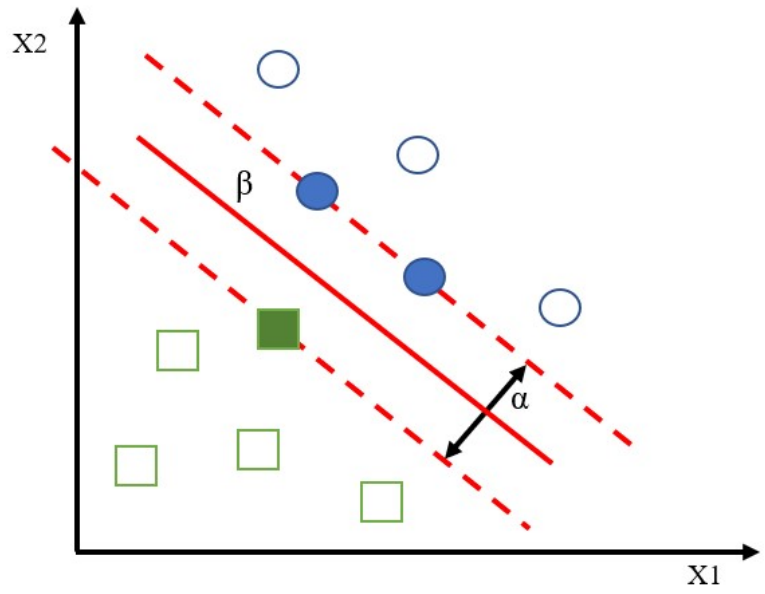


Figure 2.3: Support vectors classifier, adopted from [290]; α is the largest margin and β is an optimal hyperplane.

"The difference of NuSVC classifier with the standard SVC routine is the use and interpretation of the regularisation parameter Nu. It is an upper bound for the expected classification error [311]."

In other words, not only NuSVC is similar to SVC but also there is a parameter to control the number of support vectors in each class. The support vectors are the points where the hyperplane (split) exists. By defining the hyperplane, the SVC algorithm works while maximizing the margin or distance between classes [290] (see Fig. 2.4).

Grid search CV classifier. Grid search *cross-validation* (CV) is a hyper-parameter tuning classier or approach that, for each combination of algorithm parameters defined in a grid, will methodically construct and evaluate a model [275]. Let us take an example as follow: Assume that a machine learning system X takes hyper-parameters $b1$, $b2$ and $b3$. We first determine the range of values for each of the $b1$, $b2$ and $b3$ hyper-parameters in grid searching. For each of the hyper-parameters, we can think of as a set of values. Now the grid search technique will build many versions of X with all possible combinations of hyper-parameter values ($b1$, $b2$ and $b3$) that we first defined. This range of values

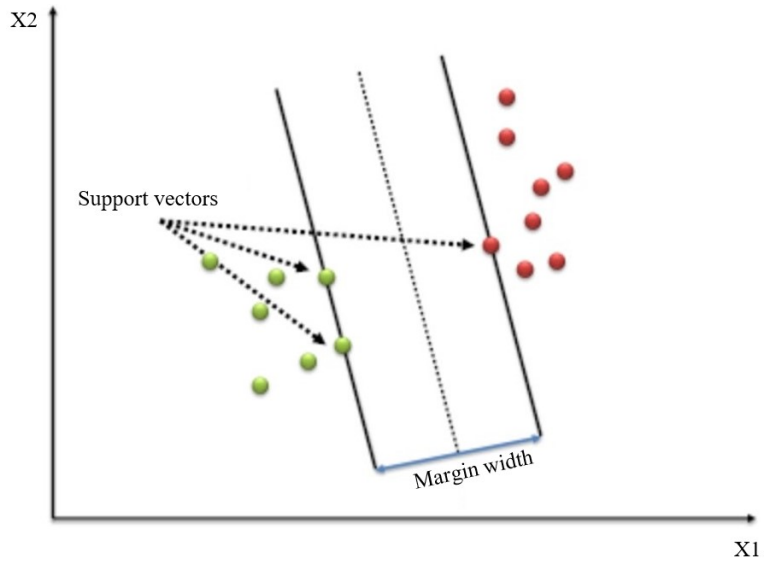


Figure 2.4: NuSVC- support vectors classifier, adopted from [290].

for hyper-parameters is called the grid. Suppose the grid is described as:

- $b1 = [1, 2, 9, 4, 5, 6]$
- $b2 = [16, 23, 33, 44, 5, 65]$
- $b3 = [115, 115, 130, 105, 155, 185]$

It will begin with the $[1, 16, 115]$ combination and end with $[6, 65, 185]$. It will go through all the intermediate variations between these two, making a grid search very costly computationally.

Random forest. A random forest classifier, as its name implies, consists of a large number of individual decision trees that function as an ensemble. Each individual tree in the random forest spits out a class prediction and the most-voted class become the prediction of our model. A simple but powerful one – the wisdom of crowds – is the fundamental concept behind the random forest. In data science, the reason why the model of random forest works well is a large number of fairly uncorrelated models (trees) acting as a committee can surpass any of the individual models of constituents [310].

AdaBoost classifier. AdaBoost is short for adaptive boosting and it is a sort of iterative ensemble learning algorithm [82, 94, 208, 275]. We can use the AdaBoost

algorithm to combine multiple classifiers to increase the performance of weak classifiers. The fundamental concept behind AdaBoost is to set the weights of classifiers and train the data sample in each iteration to ensure that difficult observations are classified accurately.

AdaBoost should meet two conditions:

- The classifier should train on various weighted learning examples interactively.
- In each iteration, by reducing learning error, it aims to attain an excellent fit for these instances.

Also, AdaBoost works in the following steps:

1. At first, AdaBoost randomly selects a learning subset.
2. It trains the AdaBoost machine learning system iteratively by choosing the training set based on the last precise training's prediction.
3. This applies the higher weight to incorrect classified results so that the high probability of classification will achieve in the next iteration.
4. It also assigns the weight in each iteration to the qualified classifier according to the classifier's accuracy. High weight will be given to more accurate classifier.
5. This method iterates until the complete training data suits without error or exceeds the maximum number of estimators defined.
6. To classify, make a vote across all the learning algorithms we have made [275].

XGB classifier. XGBoost (extreme gradient boosting) is a decision-tree-based ensemble machine learning algorithm that uses a gradient boosting framework to handle structured data [316]. XGBoost is a scalable tree boosting system that is widely used by data scientists to achieve an appropriate result and deal with problems by using a minimal amount of resources. It is presented by a novel sparsity aware algorithm for handling sparse data.

The XGBoost model works only with numeric features and can lead to over-fitting easily if hyper-parameters are not properly tuned, so it requires to do regularisation appropriately. On the other hand, the XGBoost algorithm has some advantages [51]. For example, it is:

- Extremely fast with parallel computation.

- Highly scalable end-to-end tree boosting system.
- Introduces a novel sparsity-aware algorithm for parallel tree learning.
- An open-source software library that provides a gradient boosting framework with different programming languages such as C++, Java, Python, R, Julia, Perl, and Scala.

Moreover, XGBoost algorithm has an option to penalize the cost function through L1 and L2 regularisation which helps prevent over-fitting. There are two common regularisation techniques called L1 and L2 regularisation for supervised machine learning algorithms. These two techniques can reduce the effects of over-fitting [174]. Both regularisation techniques are implemented by adding the following errors to minimize error in the cost function. L1 *(Lasso)* and L2 *(Ridge)* are objectives to minimize [288]:

$$E_1 = \alpha \Sigma_w |w| \tag{2.3}$$

$$E_2 = \alpha \Sigma_w w^2 \tag{2.4}$$

Regularization algorithms are attached to the training algorithm by adding objective [288]. These two algorithms work with adding a weight penalty to the supervised machine learning model's training section. Additionally, L1 and L2 regularization algorithms can be utilised to rank the selected features [285, 288].

Transfer Learning for Detecting Glaucoma

In this section, we have a brief discussion about traditional learning that we discussed in the earlier section versus the transfer learning technique.

We train a new model with a given data set and task or domain in traditional learning. Therefore, the learned prototype is called isolated in the sense that without external information it is trained. All the knowledge in the model extract from the data sets. To solve more complex problems, transfer learning employs accumulated knowledge from pre-trained models and moves this knowledge to a separate task.

There is a sequential system in the transfer learning technique that can conduct information from $task\ A$ to $task\ B$. Besides, a multi-task learning algorithm is identical to a neural network which is able to do multiple tasks simultaneously. And then, possibly, each of these tasks would improve the other task [312]. We can utilise transfer learning as a multi-task classification to detect glaucoma in various fields such as detecting healthy eyes, early, moderate, and severe glaucoma.

Accordingly, we can apply transfer learning for binary glaucoma detect to classify glaucoma and healthy eyes, it is likewise our research.

Using an example for the multi-task transfer-learning classifier is the best way to prove practical concept.Thus, building an autonomous vehicle is a respectable instance. Besides, the self-driving car would need to detect manifold things including pedestrians, other vehicles, and stop signs.

The second sample of usage of transfer learning is speech recognition. Likewise, the "Alexa" is a tool in Amazon. We will not train a system from scratch, however, but instead, move information from the existing model of speech recognition to our new model. The result is a new model based on a pre-trained system to trigger word detection. The third example of usage of learning from multiple tasks is training a neural network to detect cat then we can use part or all of this knowledge to help to detect an eye disease within X-ray scans.

Therefore, we can use it for glaucoma image recognition. To train the model, we have a trained neural network model to detect cat, dog, bird and so on, we can adapt this model for glaucoma diagnosis. Thus, we have (X, Y) in that X is an image and Y is the labelled healthy eye and glaucoma eye. Figure 2.5 shows a schematic view of transfer learning. We can drop the output layer of the neural network with related wights, then, create a new set of randomly initialized wights just for the last layer to detect glaucoma. We also freeze all prior layers and wights. Then, the last one or two layers need to train with the glaucoma fundus image dataset. Our fundus dataset is not that big to train more than two or three last layers that we can name it pre-training. If we have a big annotated glaucoma image dataset, we can also re-train all parameters on all layers on the rest of the neural network, so we can call in fine-tuning.

Moreover, there is (1) in Fig. 2.5 on the left as the first section of a pre-trained network for image binary classification. Output layer (2) in the network needs to be eliminated. Next (3), we can add a new layer/ layers to train by fundus images to the pre-trained template the network (4) to cause glaucoma detection structure. As

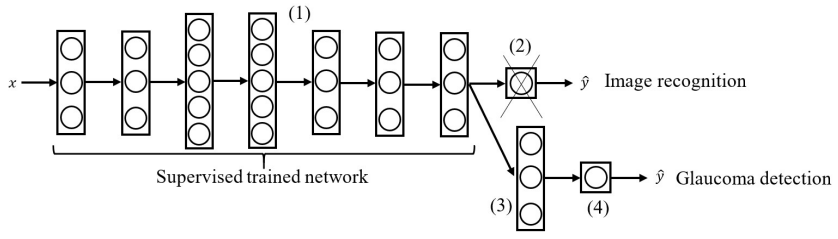


Figure 2.5: Schematic view of transfer learning for glaucoma detection.

illustrated, the combination of both networks results in one network including nine layers.

In addition, transfer learning is applicable and precious in various image-classification tasks. Transfer learning technique employs a lot of low-level image features including lines, dots, edges, curves, and pattern of the objects which come from an exceptionally large image dataset, such as COCO and ImageNet [1]. Transfer learning makes sense when $task\ A$ and $task\ B$ have the same input type of data. It is also possible to retrain the model in a new domain with the same type of data. If there is a trained model for $task\ A$ including a huge number of images, transfer learning technique can employ for $task\ B$ with a tiny dataset. Thus, low-level features from $task\ A$ could be beneficial for learning $task\ B$. The goal of transfer learning is to deal with the lack of data, improve performance, and reduce the overall time required to train a new model.

Deep Pre-trained Conventional Machine Learning for Glaucoma Detection

Conventional machine learning models have been used for classification and segmentation of images in the past few years. The classical pipeline typically includes a pre-processing step for extracting useful features from unstructured data such as images (feature engineering) or transforming data from a high-dimensional space to a lower-dimensional space. The feature-engineering step is most tedious yet critical for conducting a successful machine-learning task. Texture-feature extractions based on Fourier transforms, histogram models, B-spline coefficients, or *principal component analysis* (PCA) are typical for applied pre-processing [152, 153]. Based on these features, a *glaucoma risk index* (GRI) was introduced and then applied in machine learning [34, 35, 164, 271]. For example, wavelet-based energy features were extracted from fundus images and used for glaucoma diagnosis in [71].

There are varieties of supervised and unsupervised conventional classification models, using strategies such as *support vector machines* (SVMs) or *K-nearest neighbour* (KNN) classifiers. Those are explained briefly above as selected classical models [133, 291]. *Naíve-Bayes* is a widely used, yet effective, conventional classifier. Naíve-Bayes is considered a statistical method which incorporates the Bayes rule in predicting the class with the highest likelihood.

After extracting appropriate features, different machine-learning classifiers can be applied. We sketch major conventional machine-learning classifiers that have been used for glaucoma diagnosis.

In this section, we also outline the CNN architecture and deep learning

frameworks; different variants of those can apply for glaucoma detection.

Deep convolutional neural networks. CNNs have been developed especially in the context of image classification. A major disadvantage of CNNs is that they require for training a huge amount of data as input with respect to scale, translation or rotation, for supporting feature learning or possible generalisations [249].

Deep learning algorithms are composed of different models in comparison to machine-learning algorithms. This is due to the flexibility that neural networks provide. One can create any complicated structure with deep learning algorithms. Ultramodern architecture can employ to classify images accurately, which have been proven to be successful based on different reviews.

State-of-the-art CNN architectures are characterised by a minor difference that separates them from prevalent models and takes them one step further in solving problems. For instance, in the "ImageNet" (a large collection of annotated images on the web that is presented in "ImageSteam" for solving computer vision problems) challenge, CNNs successfully solve complex problems of identifying object classes in thousands of visually recorded different classes [64].

These CNN architectures fall into the category of "deep models" which work better in comparison to their shallow-model counterparts.

CNN Structures. We look at those CNN architectures in some detail for medical image classification and object detection. As mentioned before, CNNs are invariant to transformations which is one of the essential properties of CNNs, especially for computer-vision problems in image classification and segmentation. This feature allows one to extract an object's identity or category independent of the specifics of the visual input, such as relative positions or directions of the camera and the object. This enables a network to effectively recognise a given object in cases where actual pixel values within the object's region significantly differ [249]. Unlike conventional machine learning, CNNs can automatically extract features from unstructured data such as images. In the basic layer, these are low-level features such as edges in the image, while in subsequent layers they extract the higher-level features such as disease signs. Features at different levels are calculated by convolving filters with images at different locations.

Let $\mathbf{G}^{(n-1)}$ and $\mathbf{G}^{(n)}$ denote the input and output for the n-th layer of a CNN, respectively, both composed of a finite number of *feature maps*. By using bold letters, we identify matrices. Let $\mathbf{G}^{(0)}$ be the 2D input image, and $\mathbf{G}^{(N)}$ be the output of layer N, the final layer. Also, $M_i^{(n)} \times M_i^{(n)}$ and $M_o^{(n)} \times M_o^{(n)}$ are the size of the input and the output, respectively, of the n-th layer. Let $P_i^{(n)}$ and $P_o^{(n)}$ be the number of input and output maps, respectively, for the n-th layer. The input to the n-th layer is the output of the $(n-1)$-th layer, that means $P_i^{(n)} = P_o^{(n-1)}$ and $M_i^{(n)} = M_o^{(n-1)}$.

Let $\mathbf{G}_k^{(n)}$ be the k-th output-feature-map of layer n. Let $\mathbf{w}_{ik}^{(n)}$ be the kernel linking the i-th input map to the k-th output map, $\star$ denotes the convolution, and $b_k^{(n)}$ is the bias element for the k-th output-feature-map of the n-th layer. We obtain the output of the n-th convolutional layer as follows:

$$\mathbf{G}_k^{(n)} = f\left(\sum_i \mathbf{G}_i^{(n-1)} \star \mathbf{w}_{ik}^{(n)} + b_k^{(n)}\mathbf{I}\right) \tag{2.5}$$

where $0 \leq i \leq P_i^{n-1}$, $0 \leq k \leq P_o^{n-1}$, and $\mathbf{I}$ is the unit matrix of size $M_o^{(n)} \times M_o^{(n)}$; see [52, 154].

Although the current CNN models work very well for various applications, there is not yet a mature theory of CNNs. More efforts are required to investigate the fundamental principles of CNNs. Meanwhile, it is also worth exploring how to leverage the natural visual perception mechanism to progress the design of CNNs further [102]. CNN architectures are applied for classification of glaucomatous or non-glaucomatous fundus images, typically using transfer learning with weights trained on ImageNet.

CNN Layers. Convolutional networks are often now at the core of state-of-the-art computer-vision solutions, for a wide variety of tasks. Since 2014, very deep convolutional networks started to define the detection and recognition field, yielding substantial gains in various benchmark tests. Although the increased model size and computational cost tend to translate to immediate quality gains for many tasks (assuming that sufficient labelled data is provided for the training), computational efficiency and low parameter count are still the enabling factors for various use cases such as mobile vision or big-data scenarios.

Generally, there are five different layers in CNN models which include input, convolutional, pooling, fully connected and output layers:

- The convolutional layer computes the output at vertices of the network (also called *neurons*) that are connected to local regions in the input layer, each computing a local convolution (also called *dot product*) between specified weights at a neuron and a small region of the corresponding size to the weights [309].

- The pooling layer is a building block of a CNN. Its function is to progressively reduce the spatial size of the representation to reduce the number of parameters and the computation in the network. The pooling layer operates on each feature map independently. Common approaches used in pooling are max-pooling, average-pooling, and sum-pooling.

- The fully connected layers of the network address transfer learning, it starts with "flattening" (i.e. mapping the 2D data array into a linear vector) and ends with the "output layer". Different kinds of activation functions are used in the fully-connected layer such as the *softmax function* (to classify an object with probabilistic values between 0 and 1 for multi-object classification) or the *relu function* for binary classification in a final step of the fully connected layer.

Deep Learning Algorithms.

Deep learning algorithms are a subclass of machine learning algorithms that use a cascade of multiple layers of non-linear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input [65].

In a clinical context, deep learning can already be performed desktop-based or even just on a mobile platform [298]. Some frameworks in deep learning, their advantages and disadvantages, are summarised in Table 2.1; all these deep learning frameworks are the open-source [298].

As an example, the Spark MLlib framework [165] is valuable for large-scale data, but Scikit-learn is more efficient for medium or small-size datasets. One of the strengths of MLlib is the ability to be developed in Python, R, Java, and Scala. Bender [278], Quantized-CNN [257], and Apple's Core ML [276] were offered for machine learning on the iOS operating system and allow iPhone users to run machine learning algorithms from cloud GPU, although they are still mainly used for running pre-trained data. Tensorflow-lite [304] was designed as a software platform for machine learning on IOT *internet of things* and the Android operating system.

Selected Deep Transfer-learning Architectures. There are deep learning models for segmentation, classification, or visualisation. We list and briefly describe deep learning models that are appropriate for glaucoma detection.

AlexNet [143]. AlexNet was introduced in 2012. It is available as a pre-trained model in Matlab and Python. It is the first deep learning architecture introduced by one of the early developers of deep learning; the model was trained on ImageNet data set at that time (in 2010), which contains over 15 million annotated images from a total of over a thousand categories.

AlexNet has eight layers in total; five layers are convolutional, and three layers are fully-connected. Dropout and data augmentation techniques are applied which use image translations. The activation function is "Relu", to decrease the training time. A technique that randomly selects neurons for being ignored during the training of the neural network is called *dropout*, and *data augmentation* is a

Table 2.1: Desktop and mobile platform-based deep learning frameworks and libraries.

Framework/ library	Language	Processor	Benefits	Drawbacks
TensorFlow [2]	Python, C++, CUDA	CPUs/ GPUs/ TPUs	High performance library, operates at large scale datasets, really compatible with industry	Only Nvidia GPUs are supported
Keras [281]	Python	CPUs/GPUs	Fast prototyping and easily extensible library, works seamlessly with CNTK, Theano, TensorFlow and Auto-Keras [121]; easily accessible deep learning tools	Can not be used as an independent framework, not good enough for data processing
Torch / Pytorch [282, 296]	LuaJIT, Python C / CUDA	CPUs/GPUs	Very flexible library, high level of speed and efficiency, lots of pre-trained models available	Unclear documentation, lack of plug-and-play code for immediate use, based on Lua that is not so popular
CNTK [293]	C++	CPUs/GPUs	easily realize framework, combine popular deep learning model	Available only on Windows and Linux
Theano [11]	Python	CPUs/GPUs	Efficient framework for numerical tasks	Needs to be used with other frameworks to gain an efficient deep learning architecture
Caffe [122]	C++	CPUs/GPUs	This framework is available for Python and MATLAB, training of models without writing code	Not great with new architectures
Accord. NET [300]	C#	CPUs/GPUs	Very well-documented framework, good quality in visualization	Slow compared to TensorFlow
Scikit-learn [194]	Python	CPUs	Really efficacious for statistical modelling techniques such as classification and regression, clustering on medium-scale data based on supervised and unsupervised learning algorithms, this library is efficient for data mining	Not efficient with GPU
MLPack [61]	C++	CPUs/GPUs	Very scalable library, C++ and Python bindings available	Do not have the best documentation

technique to generate additional data without imposing additional labelling costs.

ZF-Net [269]. ZF–Net model was introduced in 2014 and achieved approximately 89% accuracy on ImageNet. It is more fine-tuned than AlexNet. ZF-Net provides a visualisation of filters and weights, therefore, it generates more interpretable outcomes. Slight modifications are made to the AlexNet model as a novel way for visualising feature maps.

VGG-Net [223]. VGG-Net was introduced in 2014 and is available in two different versions: VGG16 and VGG19. Both are pre-trained on ImageNet and available in Matlab and Python. This is an appropriate architecture to approach a specific task. The use of only 3×3 sized filters is quite different from AlexNet's 11×11 filters in the first layer. One of the benefits is a decrease in the number of parameters which is more appropriate for problems with a relatively small number of samples.

GoogLeNet [234]. GoogLeNet was introduced in 2015 and is provided in Matlab as a pre-trained model on ImageNet. There are nine inception modules [234] in the whole architecture, with over 22 layers in total without any fully connected layers. Average-pool is used to go from a 7 volume to a $1 \times 1 \times 1024$ volume, and twelve times fewer parameters than in AlexNet are used. During testing, multiple crops of the same image are created, fed into the network, and the softmax probabilities are averaged to provide a final solution. It uses concepts from R-CNN for their detection model. It is one of the first models that introduced the idea that CNN layers do not always have to be stacked up sequentially.

Microsoft ResNet [105]. ResNet50 was introduced in 2015 and is available as a pre-trained model (on ImageNet) in Keras. Other variations such as Resnet50 and Resnet101 are available in Matlab. It has a residual block and 152 layers, and it is eight times deeper than the VGG architecture. After only the first two layers, the spatial size gets compressed from an input volume of 224×224 into a 56×56 volume. A naíve increase [105] of layers in plain nets results in higher training and testing errors. The error rate was 3.6% on ImageNet dataset.

Generative adversarial networks (GAN) [98]. GAN was introduced in 2014. GANs are new types of generative models, which correspond to unsupervised learning and are used to create images. They can be used as feature extractors in a CNN model. GANs can generate artificial images that look rather natural. GANs can also be useful as a part of CNNs for data augmentation. They represent an entirely different kind of neural network architecture in which a neural network is used to generate an entirely new image, not present in the training dataset, but realistic enough to be in the dataset.

You only look once (YOLO) [205] YOLO was introduced in 2016. This is the current state-of-the-art model built on deep learning for solving object detection problems in real-time situations. All processes are done in parallel, so it can run in real-time, processing up to forty images in a second by YOLO. Although it gives reduced performance compared to its R-CNN counterpart for object detection, it still has the advantage of being real-time to be viable for use in day-to-day problems.

SqueezeNet [145]. SqueezNet, introduced in 2016, is available in Matlab to classify objects. It is a highly successful model with a small usage space (4.9 MB) compared to Inception which occupies about 100 MB. This drastic change is brought up by a specialised structure called the *fire module*.

Inception [235]. InceptionV3, introduced in 2016, is available as a pre-trained model in both Keras and Matlab. RMSProp Optimizer, factorised 7×7 convolution, and batchNorm in auxiliary classifiers are used for the InceptionV3 model. It uses

Table 2.2: Selected publications on convolutional machine-learning and transfer-learning techniques for glaucoma detection.

First author/ year	Methods/ features	Activation Function in Classifier	Dataset name/ Number of images	Performance(%)
Mohammad 2019 [178]	Deep-MNN, Deep-CNN, VGG19 Inception-ResNet-V2, Xception **NasNet-large (best performance)**	**Softmax**	**Public: 455 (Rim-one V2)**	**ACC: 98.16** **SE: 98.32** **SP: 98.48** **PP: 98.82**
Mohammad 2018 [181]	VGG19 Inception-ResNet-V2	Softmax	Healthy private–UCLA dataset: 277 Glaucoma private–UCLA dataset: 170 Retest public: 30 (HRF)	ACC: 92.13 SE: 92.39 SP: 91.26 PP: 94.95
Christopher 2018 [55]	VGG16 InceptionV3 ResNet50	Softmax	Healthy private: 9189 Glaucoma private: 5633	ROC: 0.83 (Native VGG16) ROC: 0.89 (Transfer learning VGG16) ROC:0.83 (Native InceptionV3) ROC: 0.90 (Transfer learning InceptionV3) ROC: 0.89 (Native ResNet50) ROC: 0.91 (Transfer learning ResNet50)
Al-Bander 2018 [7]	DenseNet	-	Training private: 650 (Orega) Testing public: 110 (DRIONS-DB), 101 (Drishti-GS), 99 (ONHSD), 159 (RIM-ONE).	OD segmentation DC: 0.9653 JC: 0.9334 ACC: 0.9989 SE: 0.9609 SP: 0.9995 optic cup segmentation DC: 0.8659 JC: 0.7688 ACC: 0.9985 SE: 0.9195
Raghavendra 2018 [200]	CNN-18	LDA	Private: 1426	ACC: 98.13 SE: 98.00 SP: 98.30 PP: 98.79
Orlando 2017 [185]	OverFeat [216] VGG-S	-	Public: 101 (Drishti-GS)	Without vessels AUC= 0.7212 AUC= 0.6655 With vessels AUC= 0.7626 AUC= 0.7180
Sevastopolsky 2017 [217]	U-Net	-	Public: 110 (Drions-DB) Public: 159 (Rim-one V3) Public: 101 (Drishti-GS)	IU = 0.89, DC = 0.94 (OD) IU = 0.89, DC = 0.95 (OD) IU = 0.69, DC = 0.82 (OC) IU = 0.75, DC = 0.85 (OC)
Chen 2015 [52]	CNN	-	Private: 1676 (Orega:650 and SCES: 1026)	AUC: 0.887
Chen 2015 [53]	CNN	Softmax	Private: 1676 (Orega:650 SCES: 1026)	ACC: 0.838 ACC: 0.898

added computations as efficiently as possible by suitably factorising convolutions and aggressive regularisation.

Xception [59]. Xception was introduced in 2016 and is available as a pre-trained model in Keras. It uses an intermediate step between regular convolution and a depth-wise separable convolution operation (a depth-wise convolution followed by a point-wise convolution). A depth-wise separable convolution can be understood as an Inception module with a maximally large number of towers. This

observation led to the proposal of a novel deep convolutional neural network architecture inspired by Inception, where Inception modules have been replaced by depth-wise separable convolutions. It is slightly improved compared to Inception V3 on ImageNet and has the same number of parameters as Inception V3. The performance gains are not due to increased capacity but rather to more efficient use of model parameters.

ResNeXt [259]. Introduced in 2016, ResNetXt is one of the top-level techniques for image classification. The model name, ResNeXt, contains Next. It means the next dimension, on top of the ResNet. This next dimension is called the"cardinality" dimension. ResNeXt model, in comparison with ResNet produces a 3.03% top-5 error rate, which is a large relative improvement of about 15%.

InceptionResNet [233]. InceptionResNet V2 was introduced in 2017 and is available as a pre-trained model in Keras and Matlab. It provides clear empirical evidence that training with residual connections accelerates the training of Inception-V3 networks significantly, in comparison with the previous version of an Inception-V3 network. Proper activation scaling stabilises the training of extensive residual Inception networks.

MobileNet [113]. MobileNet was introduced in 2017 and is accessible as a pre-trained model in Keras for image classification. It is a model for mobile and embedded vision applications and is based on a streamlined architecture that uses depth-wise separable convolutions to build light weight deep neural networks.

DenseNet [114]. DenseNet was introduced in 2017 and is available as a pre-trained model in Keras. It has been used for glaucoma diagnosis [7]. Deeper convolutional networks are used in this model to provide more accurate and efficient training by containing shorter connections between layers close to the input and those close to the output.

NASNet [274]. NASNet was introduced in 2017 and is available as a pre-trained model in Keras for image classification. It is a method to learn the model architectures directly on the dataset of interest. As this approach is expensive when the dataset is large, we propose to search for an architectural building block on a small dataset and then transfer the block to a larger dataset. The main contribution of this work is the design of a new search space, and finding the best convolutional architecture, then applying a defined cell to the ImageNet dataset by stacking together more copies of this cell, each with their parameters to design a convolutional architecture. Also, it introduces a new regularisation technique called *Scheduled-Drop-Path* [274] that significantly improves generalisation. A large version of NASNet also achieves 82.5% accuracy, which is 2.2% better than

equivalently-sized, state-of-the-art models, and the image features learned from ImageNet classification are generically useful and can be transferred to other computer vision problems.

Some of the listed architectures can be used for transfer learning or native modes such as NASNetLarge, InceptionResNetV2, or ResNet50.

The considerable boost of deep learning in recent years is also due to the progress in computing hardware, from *central processing units* (CPUs) via *tensor processing units* (TPUs) to *graphics processing units* (GPUs), and to the progress in the availability of extensive labelled (i.e. ground truth) image datasets. GPU-based training allowed researchers a significant acceleration in the development of deep learning models, such as Google's Colab.

Discussion and Evaluation of Deep CNNs and Traditional classifiers for Glaucoma Detection

Discussion of Deep CNNs for Glaucoma Detection

We summarised selected papers using CNN for glaucoma detection in Table 2.2. In this table, **-** indicates that a result is not reported. LDA stands for *linear discriminant*

Table 2.3: Selected publications on traditional machine-learning techniques for glaucoma (progression*) detection with the thickness of the retina.

First author/ year	Methods (best)	Data features in Classifiers	Private dataset (Number of images)	Performance
Mohammad 2019 [179]	Ensemble 26 classifiers (hybrid model)	RNFL, GCL and GCL++	H-eyes: 107 G-eyes: 68 Patients: 87	F1: 0.82 ACC: 0.82
Mohammad 2019 [180]	Five classifiers individually (AdaBoost)	RNFL	H-eyes: 154 G-eyes: 97	SE: 0.92 SP: 0.98 AUC: 0.97
Stefan 2019 [160]	Eight classifiers individually (logistic regression)	RNFL and GCIPL	H-eyes: 137 G-eyes: 432	AUC:0.89
	(CNN)	OCT images		AUC:0.94
Seong 2017 [134]	Four classifiers individually (random forest)	RNFL and VF	H-eyes: 202 G-eyes: 297	ACC: 0.98 SE: 0.983 SP: 0.975 AUC: 0.979
Siamak* 2014 [262]	Seven classifiers individually (random forest)	RNFL and SAP	G-eyes: 73 Patients: 39	SE: 0.82 SP: 0.80 ROC: 0.88 F1: 0.80

analysis. Weights in a *native model* are randomly initialised, and training was performed using only the fundus ONH data described here. Weights for *transfer learning* are initialised based on pre-training on a general image dataset (ImageNet), except for the final, fully connected layers which were randomly initialised. Table 2.2 shows the datasets that we used in my research are public and private datasets. We got the best result with the proposed model based on the NasNet-large pre-trained model. Also, we need to mention our research in comparison with other research got better performance with approximately 98%. Even though we need to mention the datasets are the same as other research, we achieved a sustainable and conceivable result to detect glaucoma which was approved by ophthalmologists.

Discussion of Traditional Classifiers with RNFL for Glaucoma Detection

Table 2.3 has selected publications on traditional machine learning techniques for glaucoma detection with the thickness of the retina in two different types of data; text (csv file) and OCT images. Although all datasets on the retinal nerve fibre layer are private, we can describe that usage of adding GCL and GCL++ to RNFL data can not help to classify with high performance in comparison with just RNFL data. Also, all studies in this area try to add more data to RNFL that show us RNFL data is really helpful to classify glaucoma versus healthy. Moreover, as we can see in both of our studies, we achieved a convincible result, and the classifier's performance were dependent on the complexity of the data.

Evaluation of Deep CNNs and Traditional classifiers for Glaucoma Detection

The listed papers use different metrics to evaluate the quality of learning. The *positive predictive* (PP) value or precision, the *area under the receiver operating characteristic curve* (AUC), and the *intersection over union* (IU or IOU) are more specific measures. Generic measures, as widely used in classification theory, are the *dice coefficient* (DC), the *Jaccard coefficient* (JC), *accuracy* (ACC), *sensitivity* (SE), *specificity* (SP), *true positive* or *true negative* (TP or TN), *false positive* or *false negative* (FP or FN) can be used [7] to evaluate new methods in glaucoma detection.

We state a few common definitions:

$$\mathrm{DC} = \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{2.6}$$

$$\mathrm{JC} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{2.7}$$

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{2.8}$$

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2.9}$$

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{2.10}$$

$$\mathrm{PP} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2.11}$$

$$\mathrm{F1} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{2.12}$$

$$\mathrm{AUC} = \frac{1 + \mathrm{TP} + \mathrm{TN}}{2} \tag{2.13}$$

The goal of using AUC in data classification is to deal with situations where one has unbalanced samples from different classes to avoid over-fitting to one class [31]. In other words, AUC provides a single metric of a classifier's performance for the evaluation of whether which model has gotten a better result than the average [94]. The F1 score represents the harmonic mean of precision and recall as in Eq. (2.12). The value of the F1-score is ranged from zero to one, and high values of F1-score specify high classification performance [229, 239]. There are many other evaluation measures such as *mean squared error* (MSE) and *root-mean-square error* (RMSE). The *confusion matrix* can also be useful in interpreting the quality of learning. These measures are typically evaluated using 10-fold cross-validation.

Challenges and Limitations

We discuss some of the common challenges of using machine learning libraries to solve glaucoma diagnosis and prognosis problems.

Challenges of Frameworks and Libraries.
TensorFlow is an open-source software library that is broadly utilised as a back-end for implementing different machine learning architectures. One of the TensorFlow constraints is time optimization. Although TensorFlow has been boosted with quantum computing algorithms recently [42], it still lies in constructing a new architecture because it is comparatively slow and should be streamlined. As a result, it is challenging to develop sophisticated deep architectures in TensorFlow (e.g., architectures that change their structure dynamically). For instance, developing TreeLSTM is not a trivial task in TensorFlow [201]. Other limitations related to deep learning frameworks and libraries include:

- Lack of general intelligence and multiple domain knowledge integrations. The intelligence of human civilisation evolved due to connectivity between people and the exchange of knowledge while current deep learning networks lack exchanging knowledge. Moreover, deep learning models fed with inaccurate or incomplete data will produce inaccurate results that make this process more complicated.

- Unable to learn from relatively small sets of examples. A framework intelligence highly depends on the training datasets that have been used, and in most applications, only relatively small sets of samples exist (as it is typically the case for glaucoma studies).

- Less powerful beyond classification problems. Most current deep learning algorithms have focused on classification tasks. They are generally less powerful for higher-level intelligence tasks or long-term planning. They (still) lack creativity or imagination.

Deep learning models suffer from global generalisation. While humans can predict different potential problems and their causes, and provide a solution to each problem, current deep learning models are unable to solve any problem besides what they have learned.

Another limitation of deep learning is the shortage of understanding about exact distinguished rules within the input data. On the basis of training and validation phases and input data, deep learning algorithms can only estimate the output, without guarantee of its correctness. These models produce only approximated outputs [297].

As biological neurons are (much) more complex than any of the artificial variants, the brain research continues toward a better understanding of the function of biological neurons. Any success in this area can lead to more intelligent

deep learning models. For instance, [191] indicates that spiking neural networks may constitute the next generation of deep learning models.

In summary, the impact of deep learning is certainly limited by the currently used models, and almost all of its successful applications use supervised learning with human-annotated data. Deep learning cannot deal with complex decisions beyond any previous training. However, *Deep Q learning* algorithms represent a "small step" towards that goal. Q learning is a model-free reinforcement learning algorithm [253] that seeks to locate the best action to take given the current state. It does not require a model of the environment, and it can confront issues with stochastic transitions and rewards, without requiring adaptations. Deep Q learning is a combination of deep learning and reinforcement learning [306].

Moreover, reinforcement learning is the problem faced by an agent that learns behaviour through trial-and-error interactions with a dynamic environment. It is an area of machine learning that focuses on training agents. Reinforcement learning algorithms take certain actions at certain states from within an environment to maximize rewards. This learning behaviour accrues via trial-and-error interactions with a dynamic environment [129]

Limitations of Datasets.
There are many difficulties involved in glaucoma data collection for developing deep learning models. One needs to comply with ethical considerations, particularly in the biomedical area. While medical research is subject to ethical standards that promote and ensure respect for all human subjects to protect their rights, it creates more challenges towards dataset generation for deep learning problems. Moreover, clearing and annotating biomedical data, which are tedious tasks, adds other challenging dimensionalities to the deep learning problems.

Other Aspects.
Another major fundamental limitation of glaucoma detection is the lack of annotated multi-modal data and fusion models to develop more robust systems, with high-accuracy models.

2.6 Critical Analysis

There are some diagnosis parameters reported for glaucoma detection [239]. Fundus images from healthy eyes have normal discs, and the CDR value is smaller in healthy eyes compared to eyes with moderate or advanced stages of glaucoma. CDR increases gradually with glaucoma progression, becoming close to one in advanced glaucoma cases (see Fig. 2.2). The area of the neuroretinal rim often decreases with the progression of the disease. Inequalities (2.14) specify the

relationship for neuroretinal rims for healthy eyes, and for eyes with moderate or advanced stages of glaucoma. This equation can be well adapted for glaucoma prognosis:

$$0 \leq \mathrm{CDR}_{\mathrm{Healthy}} < \mathrm{CDR}_{\mathrm{Moderate}} < \mathrm{CDR}_{\mathrm{Advanced}} \leq 1 \tag{2.14}$$

In addition to (2.14), one can supplement the analysis using image processing. For instance, we developed an image clustering method to describe the image contents using the *improved chaotic imperialist competitive algorithm* (ICICA) [175] and colour conversion techniques. This approach can segment retina images to support the extraction of disc and cup parameters for glaucoma diagnosis and prognosis beyond deep learning models.

Glaucoma Detection Challenges

Conventional machine learning problems deal mainly at first with feature extraction and engineering, and secondly with machine learning algorithms. Considering the tedious static feature extraction task, deep learning methods could play an important role in glaucoma detection because they do not require feature engineering prior to learning.

The first major challenge for developing a generalisable deep learning framework is a large well-annotated dataset. Collecting, annotating, cleaning, and pre-processing fundus datasets poses a significant challenge. To keep up with extensive well-annotated ImageNet datasets, currently available to computer scientists, the glaucoma research community must take similar steps.

The second challenge will be selecting an appropriate back-end for implementation and an appropriate deep learning architecture. The appropriate deep learning architecture is highly dependent on the number and quality of the fundus images. It is recommended to use architectures with a reduced number of parameters to optimize while working with small datasets. It is also recommended to use transfer learning (models initially trained on the large ImageNet dataset) techniques on deep learning algorithms to mitigate small sample size and potentially avoid over-fitting and achieving higher accuracy.

We have investigated several deep learning architectures and datasets to identify glaucoma from fundus images. We observed that image concatenation typically improves the outcome. We also observed that the NASNet architecture, developed in TensorFlow and Python, is appropriate for glaucoma detection because it does not require large datasets and generates relatively high accuracy.

For future work, we propose a combination of pre-trained deep learning architectures with unsupervised learning in a single (unified) framework to further improve the accuracy of glaucoma detection.

2.7 Summary

Original methods of medical image classifications deal with extracting appropriate features based on physicians' advice or using feature extraction algorithms, which is often time-consuming and an error-prone process. We summarised the result of this literature review as follows;

- Usage of transfer learning techniques for deep learning algorithms can achieve an appropriate accuracy and produce a faster outcome while coping with small size datasets.
- The most recent pre-trained CNN architectures for glaucoma detection were reviewed. Moreover, some frameworks in deep learning, their advantages and disadvantages were investigated.
- Selected papers using CNN for glaucoma detection as well as the impact of deep learning and limitations of currently used models were discussed. It was found that several successful applications of deep learning use supervised learning with human-annotated data.

We also categorised fundus images into two (healthy and glaucoma) groups. It was found that using current algorithms, we were able to achieve improved image classification results.

For detecting glaucoma, patients should follow the advice of the hospital's ophthalmology department or an ophthalmic clinic. But, by remotely monitoring the retina fundus and patient clinical data, an AI system might be able to detect the glaucoma disease automatically with high accuracy.

As the first step for such an AI approach, it is necessary to collect fundus images with high accuracy and reliability. Secondly, the extraction of features from the collected images is a major step, which requires accurate identification of glaucoma by ophthalmologists.

We expect that the use of transfer learning techniques for deep learning algorithms achieve high accuracy results, and the fastest and most qualitative ones are summarised in this chapter.

Overall, pre-processing and normalisation on fundus data help to promote the deep learning process. Retinal images coming from mass screening may have various image resolutions, illumination effects, or contrast. It will be important to resize input images and apply some filters on different channels and different colour spaces to enhance the reliability of results.

In some cases, we faced data constraints in the area of identifying a disease, so we had a look at different architectures and frameworks. One of the most suitable

ones, which does not require huge quantities of data, is NASNet. It is a powerful neural network based on TensorFlow in Python that achieves state-of-the-art performance for mobile platforms on many datasets. We also suppose that a feature concatenation approach might be very promising and may also enhance other areas of machine learning.

We hope that this survey not only provides a better understanding of deep CNN architectures but also facilitates future research activities and application developments in the field of traditional classifier and transfer learning for glaucoma detection.

Chapter 3

Data Collection

We present all fundus public datasets and two private datasets which utilised to our research. Data collocating is one of the most challenging parts in a medical study, in particular to classify eye diseases. It is also costly and time-consuming. One of the difficulties in my research was lack of data. Besides, getting ethics approval in New Zealand for medical research is tough and takes a long time. In this chapter, we present multi-modal data which ophthalmologists use to diagnosis glaucoma.

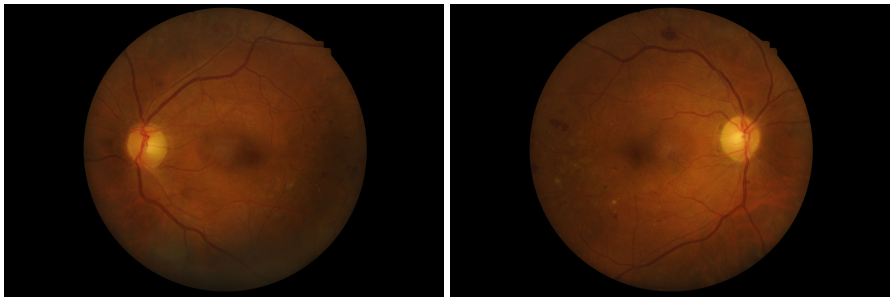
3.1 Different Data Modality

There are three common data types for glaucoma detection presented in this section: fundus image, stereo image, and OCT image including retina thickness.

Fundus Image

Fundus photography is an image of the eye-bottom. Fundus imaging is defined as the process whereby reflected light is used to obtain a *two-dimensional* (2D) representation of the *three-dimensional* (3D), semitransparent, retinal tissues projected on to the imaging plane [209]. Any process that results in a 2D image where the image intensities represent the amount of a reflected quantity of light is fundus imaging. Consequently, the purpose of *optical coherence tomography* (OCT) imaging (to be discussed below) is different from fundus imaging (see Fig. 3.1).

Fundus photographs are routinely ordered for a wide variety of ophthalmic conditions. For instance, AMD, diabetic retinopathy and glaucoma can affect the macula, fovea, blood vessels and the optic nerve head over time. Ophthalmologists' studies use serial photographs to detect subtle changes in the optic nerve and then recommend the proper therapy [19].



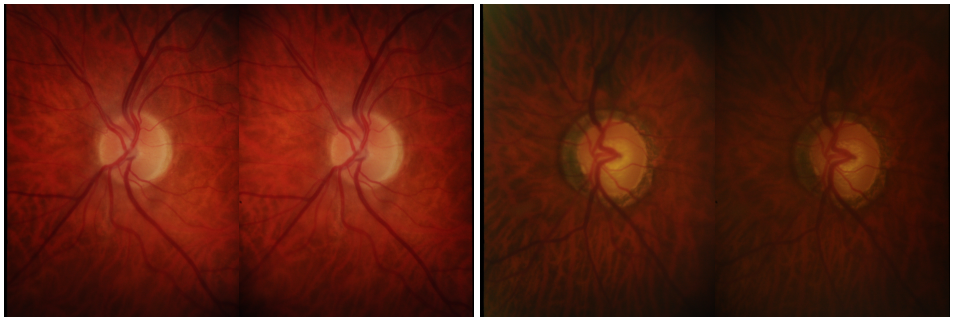
(a) Fundus photographs of the left eye (b) Fundus photographs of the right eye

Figure 3.1: Fundus photographs of both eyes of a glaucoma patient in the private SuperClinic dataset.

Stereo Image

In stereo images, the optical systems are kept parallel to each other and perpendicular to the plane of the subject. This introduces the least amount of distortion in the image [244]. The stereo fundus image is used to inspect anomalies associated with diseases that affect the eye, and to monitor their progression to identify glaucoma (see Fig. 3.2). Figure 3.2a shows a stereo image pair of healthy optic discs, and Fig. 3.2b shows optic discs with glaucomatous optic neuropathy. Even though, in this case, the glaucoma background is darker than the healthy fundus image, the rim is narrow, and one can see some changes on the optic disc area because of high IOP.

There is a potential to detect the depth with and length of changes on rim area



(a) Stereo fundus images of a healthy (b) Stereo fundus images of a glaucoma

Figure 3.2: Stereo fundus images; courtesy of [88, 292].

on fundus stereo images by stereo matching technique. Stereo matching is the search for corresponding pixels in a pair of stereo images, and it is an example of the labelling approach [138].

The OCT Image and Retinal Thickness

Generally, glaucoma has particular symptoms which aid ophthalmologists to detect it in the early stages: the retina may get thinner gradually on the ONH area because of ganglion cell death, and the rim area may get slimmer.

OCT Imaging

OCT imaging is a non-invasive imaging test which uses light beams to take cross-section pictures of the retina. It is also a sort of optic disc photography in high resolution. A cross-sectional image is constructed by scanning the light waves in the transitive direction on the tissue [85].

Thickness measurements can be extracted with a blue circle at the centre of the ONH and a diameter of 3.4 mm that covers the optic disc area (see Fig. 3.3 – A and B) [91]. With OCT, ophthalmologists can see each of the retina's distinctive layers. This allows specialists to map and measure their thickness and aids in diagnosis. They also provide treatment guidance for glaucoma and other diseases of the retina such as *age-related macular degeneration* (AMD) and diabetic eye disease [295].

The Retina's Thickness

The retinal nerve fibre layer (RNFL) or nerve fibre layer, stratum opticum, is formed by the expansion of the fibres of the optic nerve; it is thickest near the optic disc, gradually diminishing toward the ora Serrata (see Fig. 3.3 – C and D).

Attributes of the retina's thickness on the collected dataset include unique patient ID, gender, age, left or right eye, scan size 6×6 mm, scan resolution 512×256 pixel, fixation is disc, OCT focus mode is vitreous, contents are as follow; RNFL, GCL+ (*ganglion cell layer + IPL (inner nuclear layer)*), and GCL++ (RNFL+GCL+IPL) with 1024 thickness features. These three layers have been affected by glaucoma disease and the most changes have been accrued on these three layers (RNFL, GCL and IPL) (see Fig. 3.4).

3.2 Publicly Available Glaucoma-related Fundus Datasets

We list seven annotated datasets that offer fundus images with diagnostic labels which are publicly available for machine learning experiments to identify glaucoma (see Table 3.1). Public image databases can also be used as a benchmark to improve the quality of the proposed model and the patients' experience; however, the model includes traditional and deep machine learning. A brief explanation is presented as follows:

RIM-ONE. This database includes fundus photographs of ONHs with two diagnostic labels: healthy and glaucoma [88]. This dataset is exclusively focused on ONH segmentation. It was released in three different online versions [195]. It has a Matlab interactive tool in all three versions for optic disc and optic cup detection [89]. It can also illustrate the accurate gold standard, the ONH, by professionals in this field. Three hospitals, Hospital Clìnico San Carlos, Hospital Universitario Miguel Servet, and Hospital Universitario de Canarias, contributed

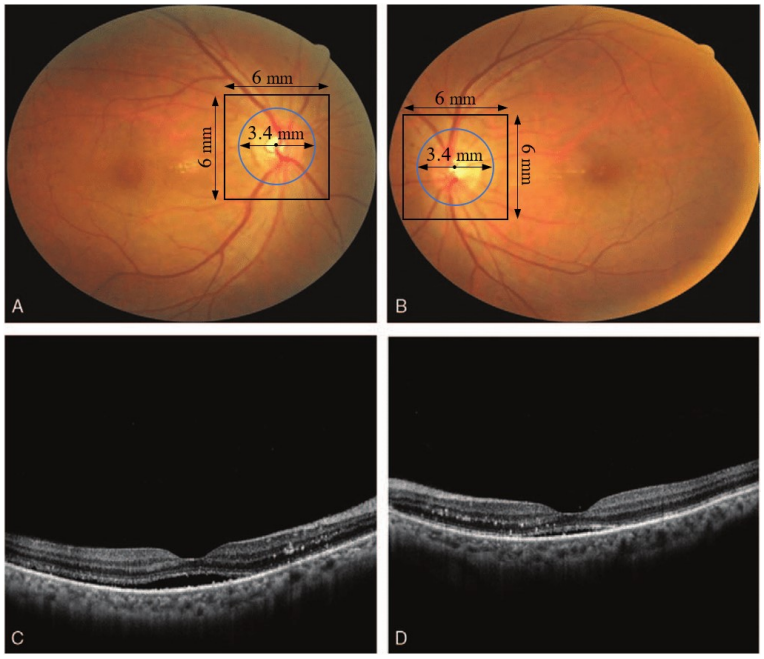


Figure 3.3: Schematic of OCT images for both eyes [115].

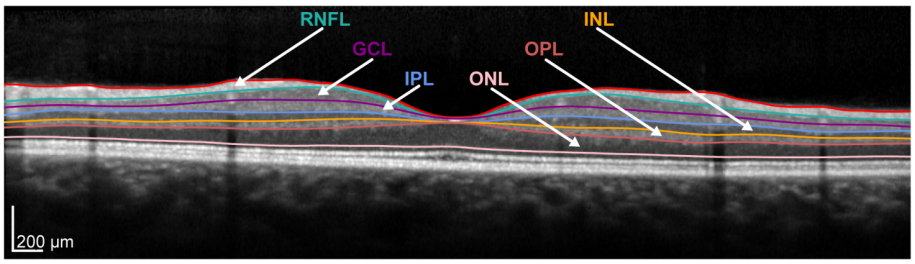


Figure 3.4: Example of an OCT with manually reviewed and corrected segmentation into the following retinal layers: retinal nerve fibre layer (RNFL), ganglion cell layer (GCL), inner plexiform layer (IPL), inner nuclear layer (INL), outer plexiform layer (OPL) and outer nuclear layer (ONL) [184].

to the development of this database.

The first version (V1) has five different diagnostic labels as identified by glaucoma experts [292]: A total of 118 fundus images from *healthy* (H) eyes, 12 images from eyes with *early glaucoma* (EG), 14 images from eyes with *moderate glaucoma* (MG), 14 images from eyes with *severe glaucoma* (SG), and 11 images from eyes with *ocular hypertension* (OHT).

The second version (V2) includes two subsets. All images are cropped to include only the ONH. This version has three diagnostic labels; healthy, glaucoma-suspect, and glaucoma. Glaucoma suspects consist of 200 images and 255 images from healthy eyes. The resolution of the images in this dataset are not consistent. This dataset is classified into two subsets: glaucoma and glaucoma suspects with a total of 200 images, and a healthy subset including 255 images at different resolutions (see Fig. 3.5).

The third version (V3) was released in 2015 with 159 stereo images of size $(2,144 \times 1,424)$ that consists of 85 healthy and 74 glaucoma stereo images. It includes five manual reference segmentations and a gold standard for each image [175]. The large number of expert segmentations enable the creation of solid gold standards and the development of highly accurate segmentation algorithms [48]. It includes images from 85 healthy eyes as well as 74 images from eyes with glaucoma at various stages. These fundus images have been captured in three hospitals in Spain. Compiling images from different medical sources guarantees acquisition of a representative and heterogeneous image set.

HRF. This dataset includes 15 high-resolution fundus images from healthy eyes, 15 from eyes with diabetic retinopathy, and 15 from eyes with glaucoma [43]. The dataset is provided by the "Pattern Recognition Lab" (CS5) of the "Department of

Ophthalmology", Friedrich-Alexander University Erlangen-Nuremberg (Germany), and the Brno University of Technology. The images are of size $(3,504 \times 2,336)$ pixels.

GlaucomaDB. This dataset includes 120 eyes with glaucoma with different sizes (e.g., (200×300) [133]. The original dataset includes 462 images captured using a Topcon TRC 50EX camera with a resolution of $1,504 \times 1,000$. The publicly available subset includes only 120 annotated images from eyes with glaucoma.

DRION-DB. The *digital retinal images for optic-nerve segmentation* (DRION) database [48] was originally created by a joint collaboration of three Spanish organisations in 2008. It has 110 retinal images of size 600×400 pixels that were annotated by two glaucoma experts with 36 landmarks. Although labels are not available for types of eye diseases, 23.1% of the images are from eyes with chronic

Table 3.1: Labelled public glaucoma datasets.

Dataset	**Number of images**	**Resolution**
RIM-ONE V1 [88, 292]	H: 118 EG: 12 MG: 14 SG: 14 OHT: 11	$\sim 400 \times 500$
RIM-ONE V2 [88, 292]	H: 255 G & GS: 200	$\sim 500 \times 600$
RIM-ONE V3 [88, 292]	H: 85 G & GS: 74	$2,144 \times 1,424$
HRF [43]	H: 15 G: 15 D: 15	$3,504 \times 2,336$
DRIONS-DB [48]	110	600×400
Drishti-GS [228]	H: 31 G: 70	$\sim 2,050 \times 1,750$
GlaucomaDB [133]	G: 120	$\sim 200 \times 300$

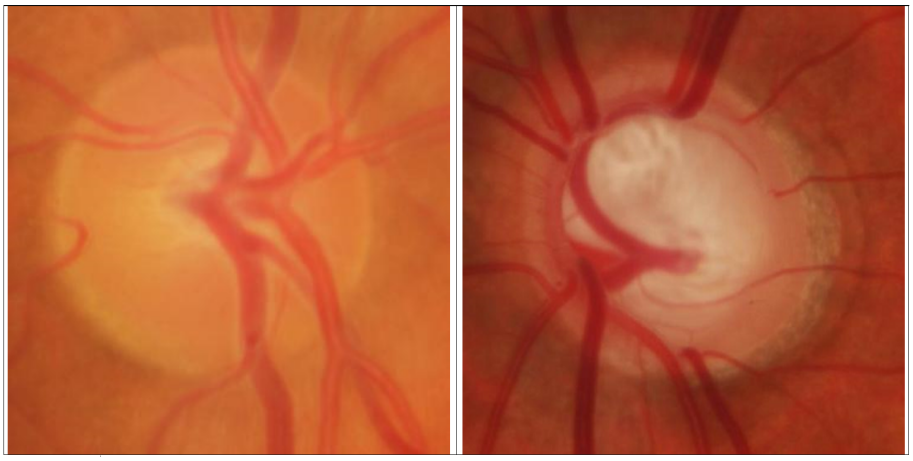


Figure 3.5: Healthy *(left)* and glaucoma *(right)* optic disc critical region on the RIM-ONE dataset [88].

glaucoma and 76.9% are from eyes with hypertension. The images were acquired with a colour analog fundus camera, approximately centred on the ONH and were stored in slide format. The images were digitized using an HP-Photo-smart-S20 high-resolution scanner in *red-green-blue* (RGB) format and 8-bits/pixel. Independent contours from two medical experts were collected by using a software tool provided for image annotation. Technicians with medical education and solid experience in ophthalmology acted as experts. To generate annotated images, in each image, each expert traced the contour by selecting the most significant papillary contour points and the annotation tool connected automatically adjacent points by a curve [103].

Drishti-GS. This database [226, 228] has two subsets: a training subset and a testing subset. A total of 50 training images are provided with ground truth for optic disc and optic cup segmentation and notching information. The testing subset includes 51 images for which ground truth is available only with permission through registration to their website. There are 31 fundus images from healthy eyes and 70 images from eyes with glaucoma at different resolutions (approximately $2,050 \times 1,750$). Selected participants were between 40-80 years of age with roughly equal numbers of males and females.

Although there are some other public fundus images datasets are available such as RIGA [9], none are annotated for glaucoma diseases.

The RIGA [18] dataset has been used in this research because of OD and OC

detection. The RIGA dataset includes three different types of files:

- Some part of the MESSIDOR [62] database that contains 460 original fundus images with a resolution of $2,240 \times 1,488$.
- The *Bin Rushd Ophthalmic Center* dataset which contains 195 original images with a resolution of $2,376 \times 1,584$.
- The *Magrabi Eye Center* database that has 95 retinal images with a resolution of $2,743 \times 1,936$.

In total, there are 750 fundus images that are labelled manually by six ophthalmologists; thus, 4,500 manually marked images are available in RIGA. All images are saved in JPG and TIFF formats.

Additionally, most of the fundus data that are used on many researches all around the world are stored locally and are not accessible to researchers. Therefore, a dataset has been collected to do this research on retinal data that present in the private dataset section.

3.3 Private Dataset Description

This section presents private datasets for our research in classification techniques with reliable outcomes [9].

UCLA Dataset

ONH photographs in the UCLA dataset have a diagnostic label of either healthy or glaucoma. The collection of these ONH photographs followed the tenets of the Declaration of Helsinki, Health Insurance Portability and Accountability Act guidelines; the Human Research Protection Program approved these studies. Written informed consent was obtained from all study participants.

The UCLA dataset includes 447 fundus images: 170 fundus images from eyes with glaucoma and 277 images from healthy eyes. Eyes were defined as glaucomatous or glaucoma suspect if there was evidence of localized or diffuse neuroretinal rim loss of retinal nerve fibre layer loss based on a review of the ONH photographs by a glaucoma specialist (KNM) regardless of the visual field findings. Otherwise, eyes were considered healthy (see Fig. 3.6). The IRB at UCLA approved the original study and all the patients consented prospectively. All procedures adhered to the tenets of the Declaration of Helsinki.

Moreover, the second category of UCLA dataset is the thickness of retina includes 251 recodes. Retinal nerve fibre layer (RNFL) thickness measurements

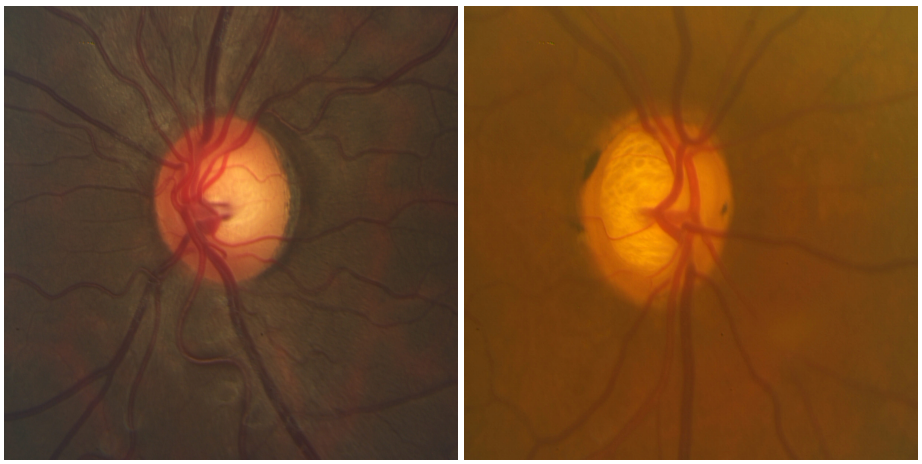


Figure 3.6: ONH photographs from the UCLA dataset. *Left:* ONH photograph of an abnormal eye, *Right:* ONH photograph of an eye with glaucoma.

from 154 normal eyes and 97 eyes with glaucoma were acquired with Spectralis OCT. We used a total of 733 RNFL measurements as input to the machine learning classifiers (described in Chapter 6) consisting of RNFL thickness in six sectors, average RNFL thickness, and 726 peripapillary A-scans.

SuperClinic Dataset

A SuperClinic dataset has been collected in Manukau, Auckland, New Zealand, which contains 3,605 patients. There are monocular and stereo fundus images obtained. These images have not been annotated by any ophthalmologists yet. It could be a potential to analyse and do more research for the future to apply this outstanding dataset in upcoming research. It is an adequately large dataset to use for training and testing of any deep learning and developed systems. Such reference would be useful to researchers developing algorithms as well to detect the optic disc and cup boundaries.

For data collecting, we were in touch with ToPcon Ltd. to fetch data from the eye devices and update the licence. We thereupon tried to de-identify and develop the scripts to generate a unique ID for patients (see Fig. 3.7). Then, data should be annotated which takes a long time. Thank to Dr Anmar Abdul-Rahman, who guided me well in all medical aspects of the project.

Adhering to the tenets of the Declaration of Helsinki the study was considered minimally observational; therefore, it did not require ethics approval as indicated by

(a) De-identified fundus and RNFL datasets (b) Data collection in the SuperClinic

Figure 3.7: Data collection and de-identified big dataset.

Table 3.2: Demographic characteristics of study participants (L: left, R: right, G: glaucoma, H: healthy).

	H	G	L eye	R eye	L eye G	L eye H	R eye G	R eye H
Male (52%)	62%	38%	50%	50%	34%	76%	38%	62%
Female (48%)	60%	40%	49%	51%	44%	66%	35%	75%

recommendations from the New Zealand Health and Disability Ethics Committee.

There are two categories on this dataset which include 3,605 patients: fundus and stereo images. Stereo and fundus images for left and right eyes are available for most patients. The resolution of fundus and stereo images is $4,288 \times 2,848$. These fundus images were chosen randomly with regard to glaucomatous and healthy patients. Moreover, the second category, the thickness of the retina, includes 525 recodes, of three to each patient.

In our annotated RNFL dataset, there are 175 RNFL data of 87 (42 female and 45 male) healthy and glaucoma patients, including a few patients with glaucoma without cupping. Some patients also have two RNFL data which were collected at different dates. About 39% of RNFL data are from patients with glaucoma and the rest of the data belongs to healthy eyes.

Also, we show the participants' demographic distribution in Table 3.2. There are three age ranges in our non-public data set. There are 34 patients who are 8–60 years old, 87 patients are of age 61–80, and the remaining 59 people are over 80 years old. Thus, our database indicates that the majority of people are in the 60–80 years range.

The dataset has a healthy and a glaucoma/suspension patients class defined by a binary classification (healthy or glaucoma label) which was annotated by an

ophthalmologist. Furthermore, we can see in Table 3.2 that there are 52% male and 48% female patients. This data set illustrates that we have approximately a balance in male and female patients.

3.4 Pupil Tracking Dataset

Use of an eye-tracking dataset as a hypothesis can be applied for the purpose of glaucoma detection with calculating changes on visual filed via evaluating the special patterns on pupil movement and quality of the vision.

Public Eye Movements Dataset

There is one freely available dataset with pupil tracking that was published in 2018 [24]. This dataset is collected in the school of Health Science, City, University of London, UK to test the hypothesis that age-related neurodegenerative eye disease can be detected in a person's spontaneous eye-movements while watching three separate video clips. Data was recorded in 32 healthy vision, and 44 glaucoma patients. The dataset contains raw gaze data, processed eye movement data, clinical vision test results, with demographic information in Table 3.3 [284].

Table 3.3: Demographic information of public eye movements glaucoma dataset [24]; "H" is healthy, and "G" is glaucoma.

Total patients	H	G	Male	Female	$Age <= 60$	$60 < Age < 80$	$Age >= 80$
76	32	44	36	40	14	59	3

Private Eye Movements Dataset

Our dataset was approved by the Auckland University ethics committee and complied with all Helsinki declarations. All participants gave full, written informed consent. This study and data collection have been continuing on a joint project with AUT and Auckland University with partially MedTech CoRE funding (see Fig. 3.8). We used the Pupil Labs device [305] as an open-source eye-tracking platform to capture the video streams then developed a model to detect the special pattern which is called *optokinetic nystagmus* (OKN)to find out the quality of the vision which can then can be applied for glaucoma detection. This model is discussed in Appendix A.

Healthy participants (n = 6 adult volunteers) were recruited for this study. OKN was elicited for each participant using an array of drifting disks (see Fig. A.5)

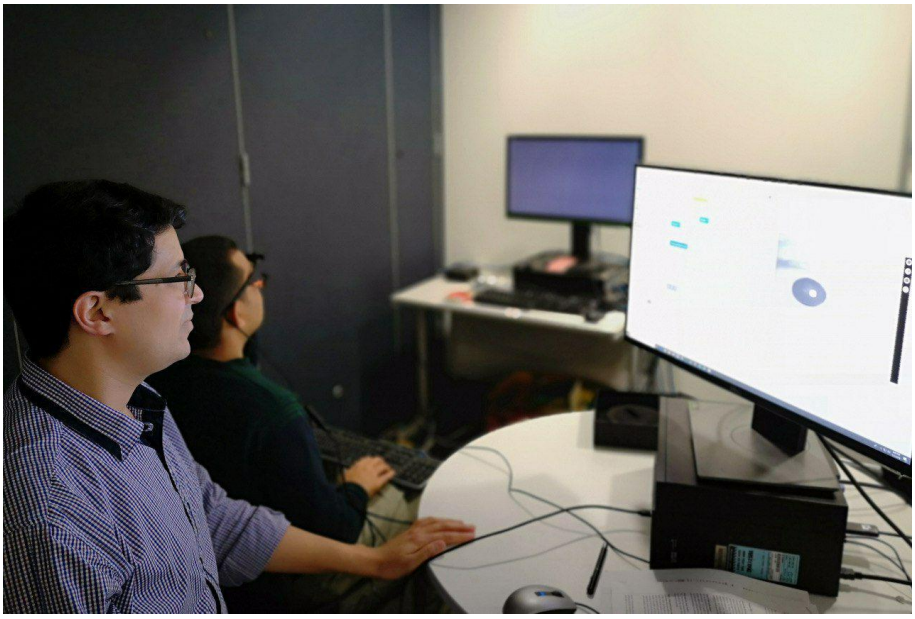


Figure 3.8: Data collection using head-mounted eye-tracking system.

presented on a 24" LCD display (AOC G2460PG with Ultra Low Motion Blur). The arrays comprised of disks with a central disk diameter chosen to ensure that the stimulus was easily seen (0.7 logMAR, 5 minutes of arc). The central intensity of the disks was 35 cd/m^2, the peripheral intensity of the disks was 9 cd/m^2, and the background intensity of the screen was 13 cd/m^2. The arrays were shown over intervals of five seconds, during which time it was shown either stationary (0 deg/sec) or drifting horizontally ($\pm$5 deg/sec) in a random left/right direction. Each state (left, right, stationary) was shown to the participant five times.

3.5 Summary

This chapter reviewed the public and private multimodal datasets on glaucoma detection, and three different type of data modality including Fundus images, RNFL data and eye movement video were discussed. In addition, seven glaucoma annotated datasets were identified among the over ten public datasets including resolution and number of images. Besides, we introduced two private fundus datasets to be able to compare the results with public datasets as a benchmark. We also selected thickness of the retina in two private datasets of a generic OCT

imaging that led to both a faster and more robust classifier for glaucoma screening.

Moreover, the eye movement dataset was utilised to recognise the OKN pattern that aimed to detect clinically significant deficits in *visual acuity* (VA), the (self-reported) ability of the eye to see fine detail that applied for OKN detection.

Chapter 4

Unsupervised Fundus Image Segmentation

Measuring the cup-to-disc ratio (CDR) is a common approach for glaucoma detection. Glaucoma can be specified by thinning the rim area that identifies the CDR value. Clustering and image segmentation can simply divide fundus images into distinct areas to estimate the optic disc (OD) and the optic cup (OC). This chapter is based on a robust method, using the improved chaotic imperialistic competition algorithm (ICICA) for determining the position of the OD and OC on colour fundus images for glaucoma detection. The predicted OD and OC boundaries are then used to estimate the CDR for glaucoma diagnosis. The performance of the proposed method was evaluated by using the publicly available RIGA dataset. It was found that some of the common problems of the K-means clustering algorithm can be addressed by the proposed method for achieving better results. Moreover, the OC and OD regions can be precisely separated from the colour image so that ophthalmologists can measure OC and OD areas more accurately. Material discussed in this chapter has been published in my publication [175]

4.1 Unsupervised Learning and Glaucoma Risk Factor

In this chapter, different types of glaucoma are investigated which are defined by the appearance of changes in fundus images due to various causes. One of the most important factors is the *intra-ocular pressure* (IOP) to resize the *optic disc* (OD) and *optic cup* (OC). A localisation of the OD and OC (often the central parts of the optic disc), and finding their borders, are performed by eye specialists where optic nerve tests include eye fundus examination. The presence of glaucoma can be identified by noticing optic nerve cupping, i.e. an increase of the OC in size [217]. One of the main indicators of the disease is the *cup-to-disc ratio* (CDR) between the sizes of the cup and the disc [10]. It is considered to be one of the most representative features of OD and OC areas for glaucoma detection, and, according to [14], an eye with a CDR of at least 0.65 is generally declared as being glaucomatous in clinical practice [217].

Clustering divides data into different clusters which are considered (e.g. by the applied algorithm) useful for object detection. Cluster analysis is the task of

grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups [151, 159]. It is a common technique for statistical data analysis, used in many fields such as machine learning, image analysis, pattern recognition, or medical imaging.

Clustering does not use category labels that tag objects with prior class labels; thus, it is named *unsupervised learning* [119].

In this chapter, we focus on partitional clustering and in particular, a popular partitional clustering method known as K-means clustering. The well-known K-means algorithm is one of the widely used algorithms due to its efficiency and simplicity in data clustering where it measures the distance between cluster representatives (centroids) and data points to partition data into K clusters. In most cases, the Euclidean distance is used as the dissimilarity measure. To find the best position of the representatives, the K-means algorithm minimizes the cost function of data variations around the centroids. However, the initial state may cause the algorithm to trap into a local optimum and as a result, affect the quality of the final solution.

Recently, the use of meta-heuristic algorithms for solving the clustering problem has been successful in attracting more attention [170]. From the perspective of optimization problems, clustering can be considered as a specific type of NP-hard problems [74]. These types of algorithms involve a search to find an optimal solution for an optimization problem, with reducing the risk of being stuck in a local optimal region.

Motivated by the success of the *imperialist competitive algorithm* (ICA) with variant optimization problems, this chapter proposes a novel data clustering algorithm based on an improved chaotic ICA for cluster analysis. The ICICA aims to enhance the capability of ICA in exploration without constraining its exploitation capabilities. In this chapter, we integrate positive benefits of the chaos into ICA to enhance its performance. The proposed ICICA differs from the standard ICA in two important aspects: First, a special equation for assimilation mode is used to discard the local optima which provide better exploration opportunities for colonies. Second, in most of ICA implementations, an empire is collapsed and eliminated when it loses all of its colonies. Nevertheless, in this research, we replace one of the weakest colonies of the best empire (low cost) with this imperialist.

4.2 Improved Chaotic Imperialist Competitive Algorithm

The flow chart in Fig. 4.1 represents our suggested approach for OC and OD detection of fundus images based on the ICICA for glaucoma detection; the historic development of the ICICA is discussed below. The following sections explain the flowchart.

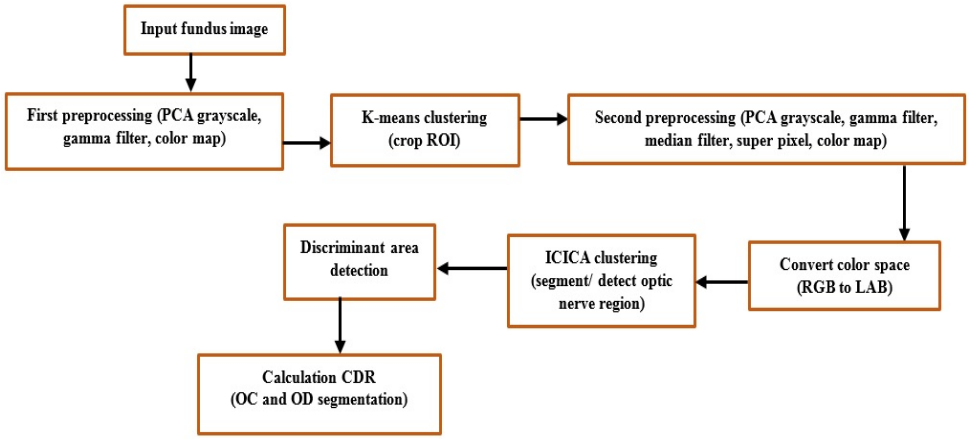


Figure 4.1: Overall approach for detecting OC and OD in the fundus image.

Pre-processing

First of all, detecting a *region of interest* (ROI) is an important part because this is the area which should include OD and OC. In a pre-processing stage, we consider it as being helpful to increase the quality (resolution) of the fundus images; this should precede the segmentation step.

In general, the pre-processing and enhancement stage is known as being essential for processing medical images. This stage is used to decrease image noise, highlight edges, and visualize digital images where needed. Examples of techniques also used in medical image pre-processing include *principal component analysis* (PCA) and superpixels before image segmentation. A subsequent enhancement stage includes an increase in image resolution or contrast enhancement for removing image noise.

A PCA greyscale method [58] has been used to enhance the quality of brightness when converting the given red-green-blue (RGB) input images into greyscale. A

gamma filter is applied to give the best possible detection. Ideally, a given medical image should be sharp and balanced in brightness.

In this research, noise is removed by using a median filter of size 11×11; due to this filter, the intensity variance of light decreases in the fundus images, and this filter is also appropriate to maintain the shape of edges and edge locations.

Subsequent superpixel partitioning provides accurate boundaries between different tissues; after that, image features are extracted from each superpixel [230]. Using superpixels leads to improved accuracy and increases the speed of computation.

In this chapter, superpixel segmentation is used to separate background regions in an image from the foreground, and for reducing the negative impacts of vessels in the segmentation process. The RGB colour space of the images, given prior to the PCA-greyscale transform, is mapped into a LAB space for better clustering. The use of the colour-opponent LAB colour space simplifies the separation of colours more approximate to human visual perception. We convert all fundus images from RGB into LAB colour space.

Passing these stages, the fundus images are ready to be used in the clustering step.

K-means Clustering Algorithm

K-means is a well-known clustering method that was primarily proposed by MacQueen and further developed by Hartigan and Wong [158]. K-means is an unsupervised learning algorithm in which the user divides the data (here we consider each data item as being a *pattern*) into a set of predetermined clusters. The clusters are defined by the use of patterns that identify the related clusters and are regarded as being the centroids of the clusters.

For the given data set $\{x\} = \{x_1, ..., x_n\}$ of n patterns, a value $k > 0$ is specified, and k centres of k clusters (called the *centroids*) are, at first, determined randomly. These are the initial *seeds*.

Then, the set $\{x\}$ is divided into k clusters (i.e. subsets of patterns) such that each pattern is devoted to that cluster where the centre of this cluster has the closest distance to the considered pattern.

Then, considering all patterns devoted to the clusters, the position of the centres will be re-calculated, and the algorithm will continue this trend till no change (or "nearly no change") is made in the position of the centres. This means that all the patterns have converged into an "appropriate" cluster, and there is no further change in the clusters.

In this case, the objective function in Eq. (4.1) is a square-error function that

defines the performance of the algorithm; it is applied for minimizing the following error term:

$$f = \sum_{i=1}^{k} \sum_{j=1}^{n} \|x_j^i - \mu_i\|^2 \tag{4.1}$$

where μ_i represents the centre of the i-th cluster and $\|x_j - \mu_i\|$ is the Euclidean distance of the j-th pattern from the centre of the i-th cluster, for $j = 1, \ldots, n$ and $i = 1, \ldots, k$.

Imperialist Competitive Algorithm

The *imperialist competitive algorithm* (ICA) is one of the most recently developed meta-heuristic optimization algorithms that have been developed based on a socio-politically motivated strategy. The main idea behind this algorithm is to divide countries into two types: imperialistic countries and colonies [25–28, 74].

Imperialistic competition and assimilation policies result in the convergence of colonies towards an optimal position. The efficiency of ICA as an excellent optimization method has been proven in various fields such as data clustering [172], hybrid flow scheduling problems [30], travelling salesman problems [267], skin colour detection [204], multilayer perceptron weight optimization, and artificial neural networks [238].

The ICA simulates a socio-political process of imperialism and imperialistic competition. This algorithm contains a population of agents or countries. Like other evolutionary algorithms, the ICA begins with a random primitive population in which each individual of the population represents a country. Countries in the ICA are similar to the chromosomes in a *genetic algorithm* (GA). At the initial stage, some of the best countries (less costly) are selected as imperial countries, and survivors are considered to be colonies of the imperialists [111].

Then colonies are divided between imperialists with regard to the power of the imperialists. After dividing all the colonies between the imperialist countries, the colonists move towards their related imperialists with regard to the cultural space. A collection of imperialist countries and some colonies make up an empire.

These empires compete with each other and sometimes replace each other. The survival of an empire depends on its power and the way it controls its colonies against other rivals. The power of the larger empires increases with the collapse of smaller empires. As a result of the repetition of this competition among empires, the colonies' power comes close to the power of their imperialists, which is an indication of convergence. The upper limit of the imperialistic competition occurs when there remains only one imperialist and several colonies that are the closest colonies to the

imperialist country with respect to position. The ideal situation is when colonies enjoy the same status and power as the imperialist.

Proposed Algorithm

Chaos is one of the most important research achievements of non-linear systems. A limited, unstable, dynamic behaviour that is dependent and sensitive to initial conditions covers the infinite irregular periodic movements in the non-linear systems. Although it appears to be random, it occurs in a definite non-linear system in definite conditions [158, 193, 261, 272].

Many of the chaotic maps in the literature possess confidence, ergodicity, and random features and qualities. Instead of random sequences, chaotic sequences have been recently used for various applications due to their relatively good and very interesting results [232, 270]. They have also been used in association with some stochastic instead of exploratory optimization algorithms for the expression of optimization variables [15, 16, 47]. The selection of chaotic sequences is justified theoretically mainly due to their unpredictability features, such as their wide range, irregularity, complex temporal behaviour, and ergodic features and qualities.

In random-based optimization algorithms, the technique of using chaotic variables instead of random variables is called *chaotic optimization algorithm*. Optimization algorithms based on chaos theory [69] are random search methods that are different from swarm intelligence or any other existing evolutionary calculation methods. As chaos is not repeated, these methods do the total search with higher speed compared to a random search that is dependent on probabilities.

When a random number is required in the ICA algorithm, it can be produced through the repetition of one stage of a selected chaotic map that is started from one of the random primary conditions in the first ICA repetition. Uni-dimensional irreversible maps are the simplest systems, possessing the ability to producing chaotic movements [186].

In the rest of this chapter, we evaluate some of the famous uni-dimensional maps for glaucoma detection. In this chapter, ICICA is proposed for the segmentation of the fundus images. The proposed technique uses the ability of the improved chaotic imperialistic competition algorithm, a meta-heuristic algorithm based on chaotic and imperialistic competition algorithms for clustering and detecting appropriate segmentation of the fundus images.

In 2012, Talatahari et al. [236] presented a chaotic imperialist competitive algorithm. This algorithm is formed by modifying the movement stage of the original algorithm. Considering the movement process of the ICA, by substituting the random numbers for the ICA parameters with sequences generated from

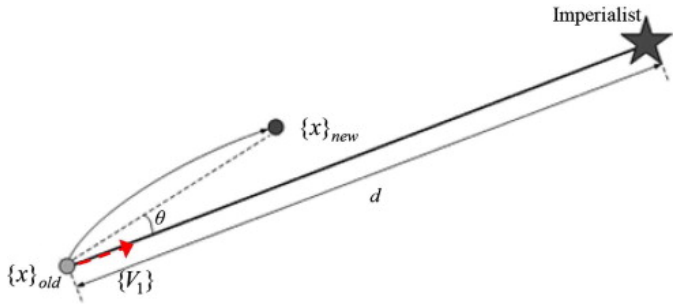


Figure 4.2: Movement of colonies toward their relevant imperialists [236].

chaotic systems. The algorithm not only uses different chaotic map values but also utilizes the orthogonal colony-imperialistic contacting line instead of θ for deviation of the colony as follows:

$$\begin{aligned} \{x\}_{new} = \{x\}_{old} + \beta \cdot d \cdot \{cm\} \otimes \{V_1\} \\ + cm \cdot \tan(\theta) \cdot d \cdot \{V_2\}, \\ \| \{V_2\} \| = 1 \end{aligned} \tag{4.2}$$

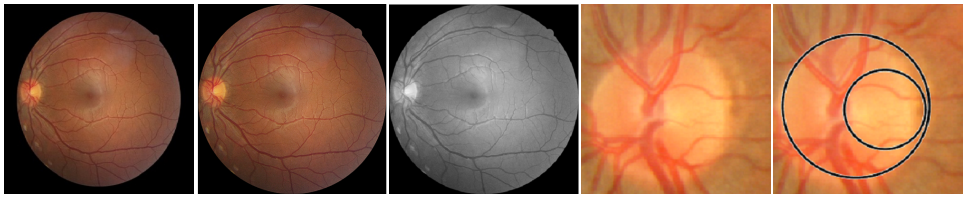
where β is a parameter whose value is greater than 1 and d is the distance between the colony and the imperialist (Fig. 4.2). $\{V_2\}$ is perpendicular to $\{V_1\}$, since this vector must cross the point obtained from the two first terms. $\{cm\}$ is a chaotic vector based on the selected map and the sign $\otimes$ denotes an element–by–element multiplication. However, in multi-dimensional problems like clustering, the calculation of $\{V_2\}$ is a very complicated task and is mathematically difficult.

Therefore, in this research, we change the movement step as follows:

- First, different chaotic map values are utilized for different components of the solution vector in place of only one value (4.2).
- Second, it is possible to obtain a random value for θ in third terms of the (4.2) that will be randomly changed in $(0, 1)$ at each iteration.

For a proper exploration and escape from local minima, the value of the selected chaotic map in (4.2) with an initial random value is calculated in each iteration. The equation is modified by

$$\begin{aligned} \{x\}_{new} = \{x\}_{old} + \beta \cdot d \cdot \{cm_k\} \otimes \{V_1\} \\ + \{cm_k\} \cdot \tan(\{\theta_k\}) \cdot d \end{aligned} \tag{4.3}$$



(a) Input (b) Remove background (c) PCA-grayscale (d) ROI (e) K-means output

Figure 4.3: Pre-processing and extraction of a region of interest with K-means algorithm [7].

where in (4.3) $\{cm_k\}$ is the k-th chaotic vector, $\{\theta_k\}$ is the k-th vector of θ values that will be randomly in the range of $(0, 1)$ at each iteration.

Equation (4.4) of the circle map in the chaotic map is as follows, and it is used to generate chaotic sequences that are required at each time of the algorithm. The random numbers are replaced by these chaotic sequences in the original imperialist competitive algorithm [162]:

$$X_{n+1} = X_n + b - \left(\frac{a}{2\pi}\right)\sin\left(2\pi X_n\right) \bmod (1) \tag{4.4}$$

Given $a = 0.5$ and $b = 0.2$, this chaotic map produces chaotic sequences in the interval $(0, 1)$.

4.3 Experiments and Results

We run the K-means clustering model to segment OC and OD regions, and the results demonstrated in Fig. 4.3. This model deployed on RIGA dataset (described in Chapter 3) and results were checked by an ophthalmologist [175].

The proposed algorithm was applied to the RIGA dataset (described in Chapter 3) while both colour conversion and ICICA clustering techniques were used (Fig. 4.4).

Moreover, the K-means algorithm was applied to the pre-processed images [183] of fundus for comparison by calculating the *root mean square error* (RMSE) [131].

For extracting the ROI from fundus images, the lowest RMSE was selected to find the background by drawing a box around the centre of the cluster Fig. 4.4d. Then, the ROI images were considered as input for the ICICA algorithm to extract the OD and OC.

Upon colour transformation of images from grey level to RGB, and then to LAB colour space, the ICICA was employed to cluster the input image and segment its

pixels. However, the clustering may lead to the production of separate regions with black holes or dots. Furthermore, Otsu's model was applied to obtain the final segmented regions. The following test shows how the proposed algorithm was used to detect OD and OC on fundus images (Fig. 4.5).

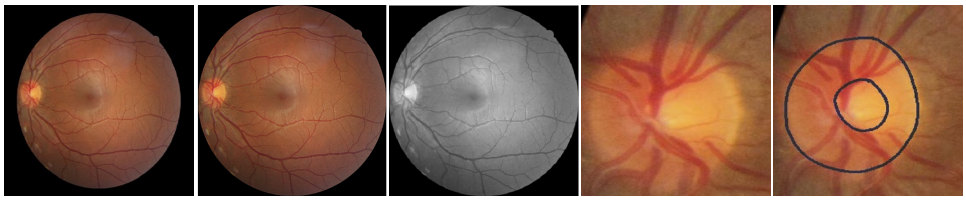
After transferring the original grey image to RGB and then to LAB colour space, the approach produced three components of "L" for light intensity, "A" related to the colour distribution along the green-to-red axis, and finally "B" for colour located along the blue-to-yellow axis. Therefore, all information related to colour is located in "A" and "B" layers.

The difference between two colours can be measured through the measurement of their Euclidean distance. The ICICA clustering algorithm was used to classify colour in the "A B" colour space. This algorithm, like the K-means clustering, requires specification of a number of clusters and distance measurement criteria to detect the degree of closeness of objects together.

Given the fact that the information related to the existing colour was located in the "A B" colour space, the objects were pixels with "A" and "B" value that were classified into three clusters by the ICICA clustering algorithm and Euclidean distance metrics.

Figure 4.5 shows the results of applying the ICICA algorithm to the processed images and extracting the final results are shown in Fig. 4.5i and Fig. 4.5j as comparative performance evaluations.

The OD and OC regions were marked by ophthalmologists manually as shown in Fig. 4.4e. It should be noted that the RIGA database was annotated by six different ophthalmologists, and in most cases, the decision was the same.



(a) Input (b) Remove background (c) PCA-grayscale (d) ROI (e) Ophthalmologist

Figure 4.4: Pre-processing and extraction of a region of interest manually by ophthalmologist [7].

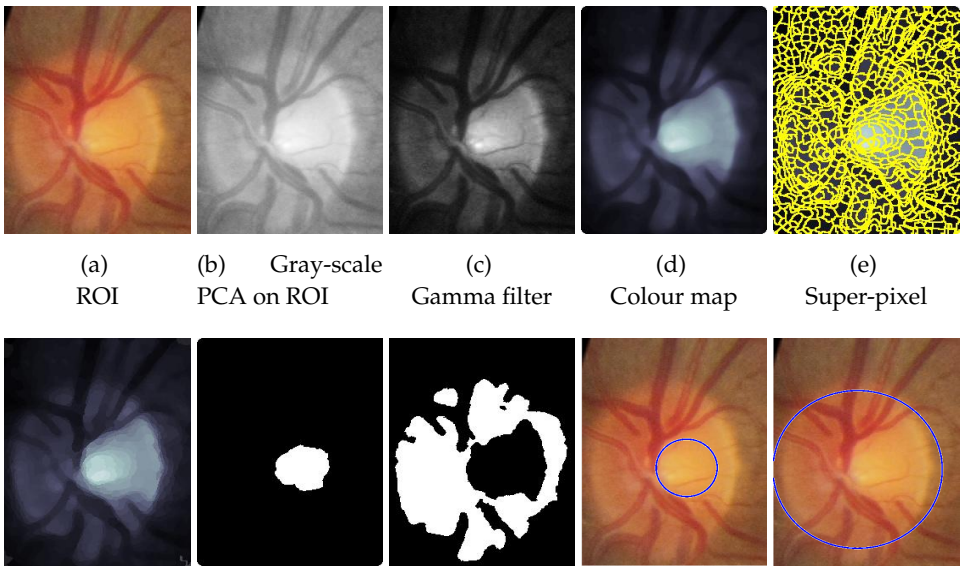
4.4 Summary

Image clustering is used to describe the high level of image contents which plays an important role in solving pattern recognition problems and medical image processing.

An image segmentation method based on ICICA and colour conversion techniques is proposed for segmentation of retinal images. The presented evaluations on fundus images show encouraging results.

In order to achieve a better performance, after de-noising [67] the image using the median filter and improving the quality of the image, the clean image was converted to RGB and then to the LAB colour space. Then, the ICICA clustering algorithm was employed to cluster the images and label the pixels. Finally, the Otsu method was used to completely segment the OC and OD.

The proposed method avoids the common problems of K-means clustering algorithms, such as dependence on initial values and early convergence that cause



(a) ROI (b) Gray-scale PCA on ROI (c) Gamma filter (d) Colour map (e) Super-pixel

(f) Output superpixel (g) OC cluster with ICICA (h) OD cluster with ICICA (i) Proposed method OC (j) Proposed method OD

Figure 4.5: Region of interest including OC and OD results with ICICA algorithm [7].

poor results.

In order to evaluate the performance of the proposed method, the final results were compared with the results of K-means clustering using the RMSE parameter. According to the results, the OC and OD regions can be precisely separated from the colour image and the proposed method enables ophthalmologists to measure OC and OD areas accurately.

Chapter 5

Superpixel Segmentation on Stereo Fundus Images and Disparity Map

Population-based studies report high rates of undiagnosed glaucoma with over 50% of the population with glaucoma living in developed countries remaining undiagnosed and unaware of their disease. Clinical diagnosis rests on the detection of the characteristic optic disc signs. Stereo optic-disc imaging improves intra and inter-observer agreement in the detection of optic disc abnormalities. We propose a robust method to help the specialist in detecting some abnormalities in stereo optic-disc images using stereo vision and superpixel segmentation concepts. A stereo vision system produces a disparity map for the input stereo images of the retina in which abnormalities are more distinguishable. In the region of interest in the disparity map, we can clearly visually recognize all abnormalities because of changes in distances for abnormality parts in glaucoma patients. The produced disparity map is segmented using two different superpixel segmentation algorithms (simple linear-iterative clustering and simple non-iterative clustering) to detect abnormalities. The original stereo images are also segmented using the same concept; results are compared with the segmented disparity map. Material discussed in this chapter has been published in my publication [177].

5.1 Stereo Vision and Glaucoma Abnormalities

In this research, there is a potential to detect one of the abnormalities that is RIM shrinking based on the disparity map technique. The general definition of abnormalities for glaucoma aspects is as follows:

- Increased cup-to-disc ratio (i.e. the ratio of the optic cup in the central optic disc to the margin of the optic disc),
- Notching of the neuroretinal rim (focal loss of the margin of the optic disc margin)
- The symmetry of optic disc cupping

- Loss of retinal nerve fibre layer

We assessed optic discs from patients with different subtypes of glaucoma defined by abnormal *stereo OD images* (SODIs). The RIM-ONE [292], release three, used in this chapter is a public SODIs dataset that was described in Chapter 3. Glaucoma manifests as ONH cupping; clinically this is most easily recognised at the superior and inferior poles of the optic disc as well as focal and diffuse retinal nerve fibre layer loss [81,219].

Stereo vision extracts 3D information from multiple 2D views of a scene. Stereo vision works by using a stereo-matching algorithm which finds corresponding points in a stereo image pair and produces a disparity map. This map encodes the difference in horizontal coordinates of corresponding image points. The values in the disparity map are inversely proportional to the scene depth at the corresponding pixel location [303]. Up-to-date stereo matching is robust (due to progress in matching algorithm design) and fast (as it only processes one stereo image pair). Many modern applications, such as advanced driving assistance system or robot navigation, work on principles of stereo vision to estimate the actual distance or the range of objects of interest relative to the camera [302]. In this chapter, we apply a stereo-matching algorithm to our input SODIs to assess the difference in depth values for abnormalities compared to OD. We use a commercial stereo-vision system SP1 [214] which produces disparity maps for input SODIs.

The process of partitioning a digital image into multiple segments is called *image segmentation*. These multiple segments are sets of pixels which are also known as *superpixels*. Pixels that share certain characteristics such as similar colours, intensities or grey-levels are grouped together to form superpixels [218]. By doing so, image segmentation simplifies the representation of an image and makes it easier to analyze [87]. The *simple linear-iterative clustering* (SLIC) algorithm is one of the widely used algorithms. It efficiently generates compact, nearly uniform-sized superpixels by clustering pixels in a combined five-dimensional colour and image coordinate space [299]. In this chapter, we apply the SLIC algorithm to divide input SODIs and disparity maps into distinct areas to estimate the abnormalities on OD. We also apply the *simple non-iterative clustering* (SNIC) algorithm on SODIs and disparity maps; SNIC is an improved version of SLIC. Results of SLIC and SNIC algorithms are compared at the end.

5.2 Stereo Vision Materials

The SP1 stereo vision system has been used in this research; it is a product of Nerian Vision Technologies (NVCom) (see [214]). It performs stereo matching in

real-time using a semi-global matching algorithm [109]. It is a stand-alone processing system with an integrated *field programmable gate array* (FPGA) which produces a dense disparity map. The system can be configured through a web interface using its IP address. A gigabit Ethernet connection is established between the SP1 and a client computer for the transmission of images. The NVCom software is required to send input stereo images to the SP1 and to display and write the received images from the SP1. The system can process the images from two industrial USB cameras in real-time, or it can also generate disparity maps for input stereo images transmitted from a client computer. The SP1 can process images with a resolution from 320×240 pixels up to $1,440 \times 1,440$ pixels; the number of calculated disparities ranges from 32 to 256 pixels. It can also reconstruct the 3D location of corresponding scene points from the disparity map [214].

5.3 Iterative Clustering in Stereo Images

The flow chart in Fig. 5.1 represents our approach for the detection of abnormalities on ODs in SODIs. At first, input stereo fundus images are separated in left and right images. Then, the left and right images are cropped in order to have a *region of interest* (ROI) accurately. These two images are given to the SP1 system to calculate the disparity map. Afterwards, in two stages, the disparity map and the left image are entered into the SNIC and SLIC algorithms to achieve the segmented masks. Finally, the SNIC and SLIC masks are overlaid on the input left image as outputs (see Fig. 5.5). Current segmentation studies, as reviewed in [6, 10], are not used in the disparity map.

The concepts of stereo vision and superpixel segmentation for glaucoma detection are discussed below in the rest of this section; processes applying in this flow chart are explained in the following sections.

Stereo Vision

We cite [112]: "Stereopsis is a term that is most commonly used to refer to the perception of depth and 3-dimensional (3D) structure obtained on the basis of the visual information deriving from two eyes by individuals with normally developed binocular vision"; the difference in the relative horizontal position of objects in the two images is referred to as binocular disparity. The visual cortex of the brain processes disparities to yield depth perception.

Stereo vision is a well-known ranging method because it resembles the basic mechanism of the human eye. Computer stereo-vision systems use the same principle by replacing eyes with two CCD cameras. They are displaced

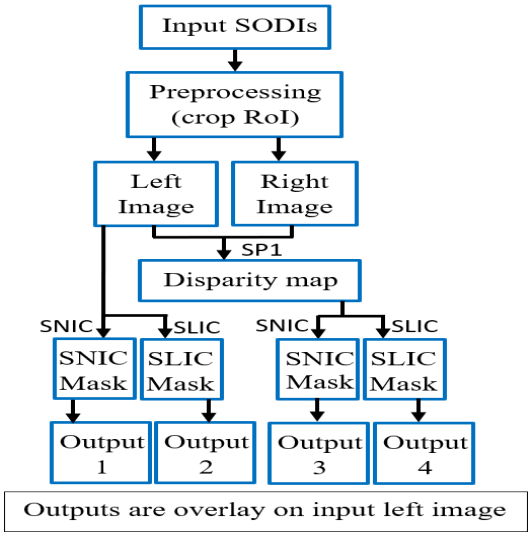


Figure 5.1: Overall approach for segmentation of abnormalities on optic discs.

horizontally to obtain two different views. A disparity map, which encodes the difference in horizontal coordinates of corresponding image points, can be obtained by comparing these two slightly different images. The values in the disparity map are inversely proportional to the scene depth at the corresponding pixel location [303].

Figure 5.2 illustrates matching in canonical stereo geometry; both the images have collinear rows y, defining the *epipolar lines*. Here, the left image is the base image indicated by B, and the right image is the matched image indicated by M. The projection of a 3D world point P in the base image is displayed by pixel $p = (x, y)$. Now, we must search for a corresponding pixel on the same epipolar line in the matched image M. The two pixels are corresponding if they are projections of

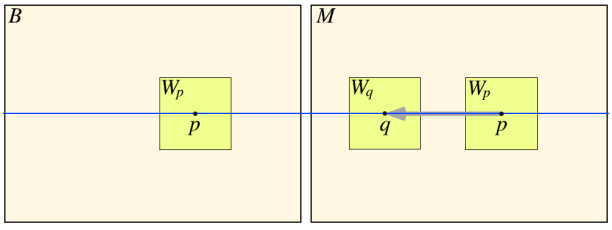


Figure 5.2: Left and right camera image for canonical stereo geometry [138].

the same point P in the scene.

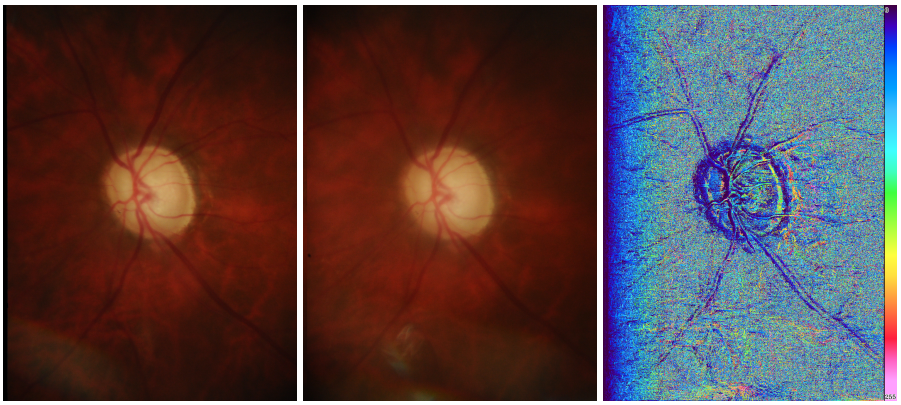
Let pixel $q = (x + d, y)$ be the corresponding pixel of pixel p. Here, d is the value of the disparity which is plotted in a disparity map [138]. Also, d is defined by equality (5.1) which is specified by the distance between pixels p and q in the same row to calculate the disparity map [283].

$$d = |\mathrm{x_p} - \mathrm{x_q}| \tag{5.1}$$

We start at pixel p in the base image (B), consider its neighbourhood defined by a square window (W), and compare with neighbourhoods around pixel q on the epipolar line (i.e. on the same row due to canonical stereo geometry) in the matched image (M) [138]. Therefore, W_q in match image M is the best window matching for corresponding W_p in B as the base image based on searching the window W_p in match image M, and q is the best point of p over on image M. Figure 5.3 shows an input stereo pair and a resulting disparity map; we use a colour key for visualising integer disparities.

The first step in our experiments is to perform stereo matching. We use the SP1 stereo-vision system to produce a disparity map for the input fundus images. The SP1 produces a disparity map from the perspective of the left camera image [301].

The input image data is transmitted from a client computer to the SP1 for stereo matching. The system performs stereo matching using a *semi-global matching* (SGM) algorithm and produces a disparity map with a bit depth of 8 bits per pixel. The NVCom application is used to receive and display disparity maps from the SP1



(a) Input left image (b) Input right image (c) Disparity map

Figure 5.3: Input stereo pair and resulting disparity map (used colour code is shown on the right).

[301]. Results for tested 3D scenes verified the accuracy of the SP1; the following SGM algorithm has been developed for accurate, pixel-wise matching at low run time [109]. It combines concepts of global and local stereo methods [213].

Simple Linear Iterative Clustering Algorithm

Superpixel algorithms group pixels into perceptually meaningful atomic regions. These pixels should adhere well to image boundaries [3]. The SLIC algorithm produces the desired number of regular, compact and nearly uniform superpixels with low computational overhead. The SLIC algorithm is very simple, fast and efficient, which makes it extremely easy to use. [299] documented that this algorithm achieved a segmentation quality better than four state-of-the-art methods (at that time).

SLIC adapts a *k*-means clustering approach to efficiently generate superpixels [3, 273]. There are two important distinctions compared to general *k*-means clustering:

1. The search space for the number of distance calculations in the optimization is limited to a region proportional to the superpixel size.

2. The colour and spatial proximity are combined by a weighted distance measure. It also provides control over the size and compactness of the superpixels [3].

SLIC has only one parameter k which is the desired number of approximately equally sized superpixels.

For colour images in the CIELAB colour space, the first step of the clustering procedure is to sample k initial cluster centres $\mathbf{c}_i=[\mathrm{l}_i, \mathrm{a}_i, \mathrm{b}_i, \mathrm{x}_i, \mathrm{y}_i]^\top$ on a regular

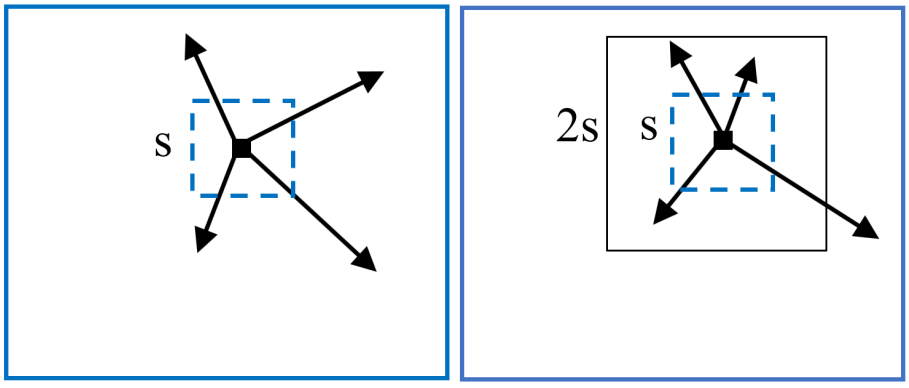


Figure 5.4: Search regions of *k*-means and SLIC algorithms [3].

grid, spaced s pixels apart. Here, $s = \sqrt{\frac{N}{k}}$ for roughly equally sized superpixels. To avoid centring a superpixel on an edge, the centres are transferred to seed locations corresponding to the lowest gradient position in a 3×3 neighbourhood. The next step is to assign each pixel at location p to the nearest cluster centre whose search region overlaps its location. As depicted in Fig. 5.4, the size of the search region is limited to reduce the number of distance calculations, in contrast to general *K*-means clustering. Since the expected spatial extent of a superpixel is a region of approximate size $s \times s$, the search for similar pixels is done in a region $2s \times 2s$ around the superpixel's centre. The last step is to adjust the cluster centres to be the mean vector $[l, a, b, x, y]^\top$ of all the pixels belonging to the cluster. The residual error E between the new and previous cluster centres is calculated. The last two steps of the algorithm are repeated iteratively until the error converges [3]. Finally, a post-processing step reassigns isolated pixels to nearby superpixels to enforce connectivity [3,76].

Simple Non-Iterative Clustering Algorithm

SNIC is an improved version of the SLIC algorithm [4]. It is simpler and faster as it is non-iterative. SNIC does not use a distance map and therefore requires less memory. It enforces connectivity from the start. The authors say: "SNIC algorithm retains the desirable properties of SLIC, namely computational efficiency, simplicity of implementation and use, and control over the number and compactness of superpixels." SNIC outperforms SLIC on quantitative benchmarks [4].

Superpixel partitioning provides accurate boundaries around different tissues which makes the extraction of image features easier [230]. In this chapter, superpixel segmentation is used to reduce the negative impacts of vessels on abnormality detection. Moreover, it groups the abnormalities on OD in one or more clusters which helps in glaucoma detection.

5.4 Experiments and Results

The first step of this experiment is to perform stereo matching on to the input dataset. The SODIs are cropped to ensure that the left and right camera images are of identical size. The size of the left and right camera images is set to $1,056 \times 1,424$ pixels. The stereo fundus images of glaucoma patients are then transferred from a client computer to the SP1 through gigabit Ethernet. The SP1 computes the disparity map and transmits it back to a client computer [301].

For computing the disparity map, the value of the maximum number of disparities is set to 256. For generating the disparity map, the SP1 uses intrinsic and extrinsic calibration parameters. The value of the re-projection error is inversely proportional to the accuracy of the disparity map. For our experiments here, the value of the re-projection error was 0.06. The SP1 produces a disparity map from the perspective of the left camera image. A rainbow colour-map is applied on the disparity map for better visualization. As we go from blue colour to pink colour, the disparity value increases gradually. See Fig. 5.3 for an example of an output disparity map for one SODI. As the abnormalities are present on or near the OD, we crop the input left camera image and the output disparity map into a ROI around the OD.

In the ROI of a disparity map, we can visually recognize all abnormalities because of changes in the distance (seen by changes in colour) for abnormality parts in glaucoma patients; we can easily distinguish abnormalities by comparing Figures. 5.5a and 5.5b.

The OD showed in Fig. 5.5a has advanced glaucoma as the inner line is close to the outer line indicating a loss of tissue between both lines; the tissue between the two lines is called the *neuroretinal rim*.

Figure 5.5b is a cropped disparity image. The size of the cropped images is uniformly 585×585 pixels. The original left image and the produced disparity map are now considered as inputs for the SLIC and SNIC algorithms. The SLIC and SNIC methods segment the images and produce boundary masks. Both algorithms are applied for the same number k (we decided $k = 200$) of superpixels for comparison purposes. The value of compactness in the SNIC algorithm is set to 40. If the number of superpixels increases (above 200), there are high chances that the clustering may lead to the production of regions having black holes or dots (showing non-simple region topologies).

The computed SLIC and SNIC boundary masks of the input fundus image and also of the disparity map are then overlaid on the original (cropped) input left image for better comparative visualization. As a result, four output images are generated for each SODI. Figures 5.5c and 5.5d show the segmented input fundus image using SLIC and SNIC algorithms, respectively. Resulting images, after overlaying the boundary masks of segmented disparity maps onto original input images, are shown in Figures. 5.5e and 5.5f. Figure 5.5e illustrates the result of applying the SLIC algorithm and Fig. 5.5f is the result of applying SNIC algorithm.

This chapter indicates a novel method which may help the specialist for glaucoma detection by using stereo matching. The output disparity map has different disparity values for abnormalities which helps to detect glaucoma.

Therefore, finding abnormalities in these clustered images is easier for specialists

compared to the fine structure of the pixel grid. The four types of segmented output images are suggested for being analysed by the specialist.

5.5 Summary

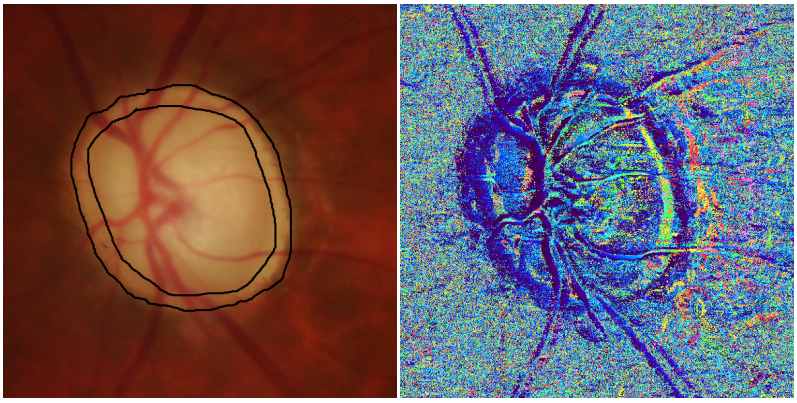
An image segmentation method was proposed for SODIs by combining two superpixel algorithms (SNIC and SLIC) and disparity map calculations, resulting in four kinds of segmentation of ROIs. The goal is to support the early detection of (small) abnormalities.

According to a consulted ophthalmologist, results as shown in Figs. 4.4c and 5.5d appear to be promising and useful due to the following factors:

1. Ability to identify the margin of the optic disc
2. Some outlining of blood vessels
3. Outlining of the optic cup (the centre depression of the optic disc)

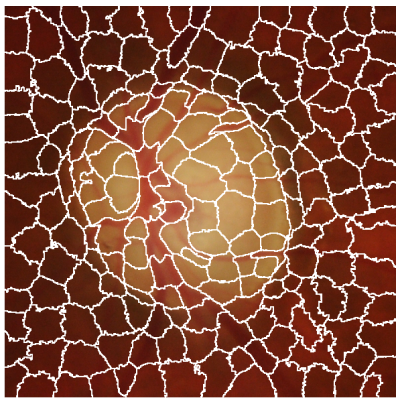
Also, results shown in Figs. 4.4c and 4.4d are more accurate in contrast to Figs. 5.5e and 5.5f. The results derived from segmented disparity maps use a recent approach which may be used in further studies for glaucoma detection. From our experiments, we can conclude that disparity values for abnormality regions tend to larger than disparity values for normal parts which assist with the specialist for glaucoma detection (Fig. 5.3c). This change in disparity values can also be further studied towards abnormality segmentation.

In future work, we will focus on the development of a system to eliminate the healthy segments of a disparity map image to achieve improved abnormality detection and have an accurate rim area segmentation.

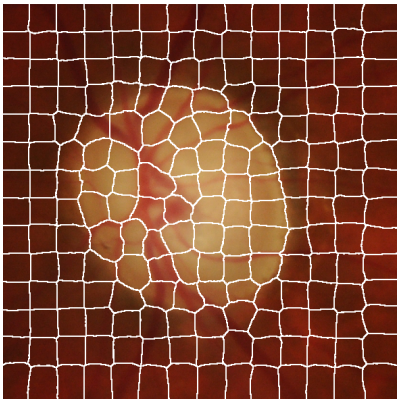


(a) ROI in input left fundus image

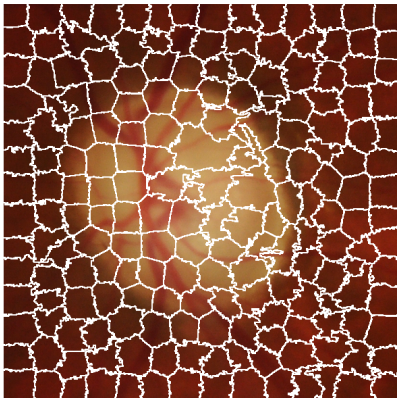
(b) ROI in disparity map



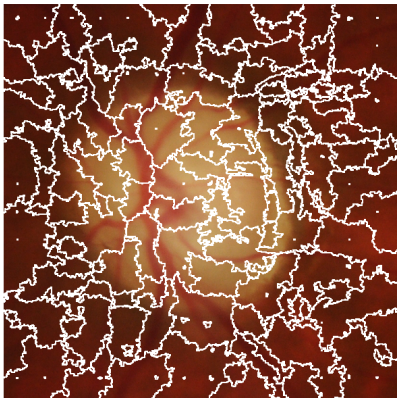
(c) Input image segmentation using SLIC



(d) Input image segmentation using SNIC



(e) Disparity map segmentation using SLIC



(f) Disparity map segmentation using SNIC

Figure 5.5: Results after applying SLIC and SNIC algorithms.

Chapter 6

Ensemble Traditional Classifiers on Retina Thickness

In this chapter, we present a proposed ensemble model on both SuperClinic and UCLA datasets. We also try to compare the results but the number of features on both datasets is not the same, so we cannot discuss and compare the performance of our model with the results of each.

We employed the UCLA dataset (discussed in Chapter 3) to run a hybrid machine learning model that can automatically identify glaucoma using peripapillary retinal nerve fibre layer thickness measurements [180]. Additionally, the proposed fused model on the SuperClinic dataset (also described in Chapter 3) is based on a stack of supervised classifiers combined via an ensemble learning method to achieve a robust and generalised model for glaucoma screening. Additionally, we implemented an unsupervised model based on K-means clustering with 80% accuracy for glaucoma screening. We have also followed two purposes: first, to assist the ophthalmologists in their daily patient examination to confirm their diagnosis, thereby increasing the accuracy of diagnosis. The second usage is glaucoma screening by optometrists in order to perform more eye tests and better glaucoma diagnosis. Therefore, thesis experimental tests illustrate that having only one SuperClinic dataset still allows us to obtain highly accurate results by applying both supervised and unsupervised models.

In this chapter, all ensemble, hybrid, fussed and stack model expressions have the same meaning and the reason why is to not be repetitive in my thesis. Material discussed in this chapter has been published in my publications [179,180].

6.1 Hybrid Machine Learning Model on UCLA Dataset

Glaucoma, suspected and healthy eye definitions, and eyes with glaucoma were defined as those with at least four abnormal visual field test points (at %5) confirmed once. Suspect eyes had either *glaucomatous optic neuropathy* (GON) based on stereo photograph examination or they had IOP greater than 21 and they had a normal

visual field at the baseline. Healthy eyes were defined as those with a normal disc, and IOP less than 21 mmHg [221], and normal visual field.

We wanted to examine whether machine learning classifiers can detect glaucoma onset from retinal nerve fibre layer thickness measurements prior to the development of visual field (VF) damage (VF conversion) in glaucoma suspect eyes. We used "RNFL thickness" to discriminate healthy eyes from eyes with glaucoma.

As we can see from instrument measurements in Fig 6.1, RNFL data belongs to the circle of the retina eye on the optic disc area. This area can separate into four main parts which are named Superior, Nasal, Inferior, and Temporal.

Participants had advanced glaucoma progression in 97 eyes and 154 healthy eyes on the UCLA dataset with RNFL data (discussed in Chapter 3). RNFL thickness measurements were collected from Spectralis (Heidelberg Engineering) with circumpapillary circle scans, 768 A-scans, and seven sectoral and global parameters at UCLA through Dr Kouros Nouri-Mahdavi. Also, the most discriminating RNFL thickness inputs to the classifier were average RNFL thickness, temporal-inferior RNFLthickness, and A-scans 60, 231, and 296. We need to somehow define the topography of these A-scans.

In this section we try to describe a solution of the research question: can machine learning classifiers detect glaucoma from the *retinal nerve fibre layer* (RNFL) thickness measurements? We discuss the solution of this question which informs combining several models to have reliable and robust outcomes. We aim to optimize the machine learning model to glaucoma classification based on RNFL data which needs to be analysed, and do data cleaning and pre-processing on it.

First designing a study map to classify the healthy and glaucoma patients (see Fig. 6.1). Before feeding the input to the proposed model we need to do data cleaning. Therefore, some of the features are eliminated which do not have effective value or without value on it. There is A-scan data as raw data with 768 features for each patient which is applied to data cleaning to have access to the significant features as an output of the data cleaning step for the remaining 726 attributes. Thus, we designed a framework to detect glaucoma as shown in this diagram then we fed the classifier with seven different instrument parameters and 768 A-scans and we trained the classifiers and assigned the test eyes to wither of the two classes.

Table 6.1 shows the mean and standard deviations calculated for healthy and glaucoma patients on different regions and also globally. The average age for glaucoma patients is 68.9 years and for healthy patients is 51.9 years. You can see demographic clinical information for both healthy and glaucomatous categories; this table shows the demographic information of the study eyes.

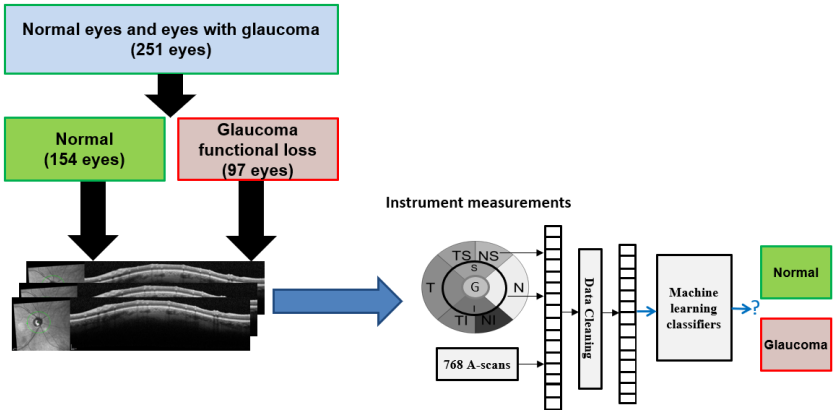


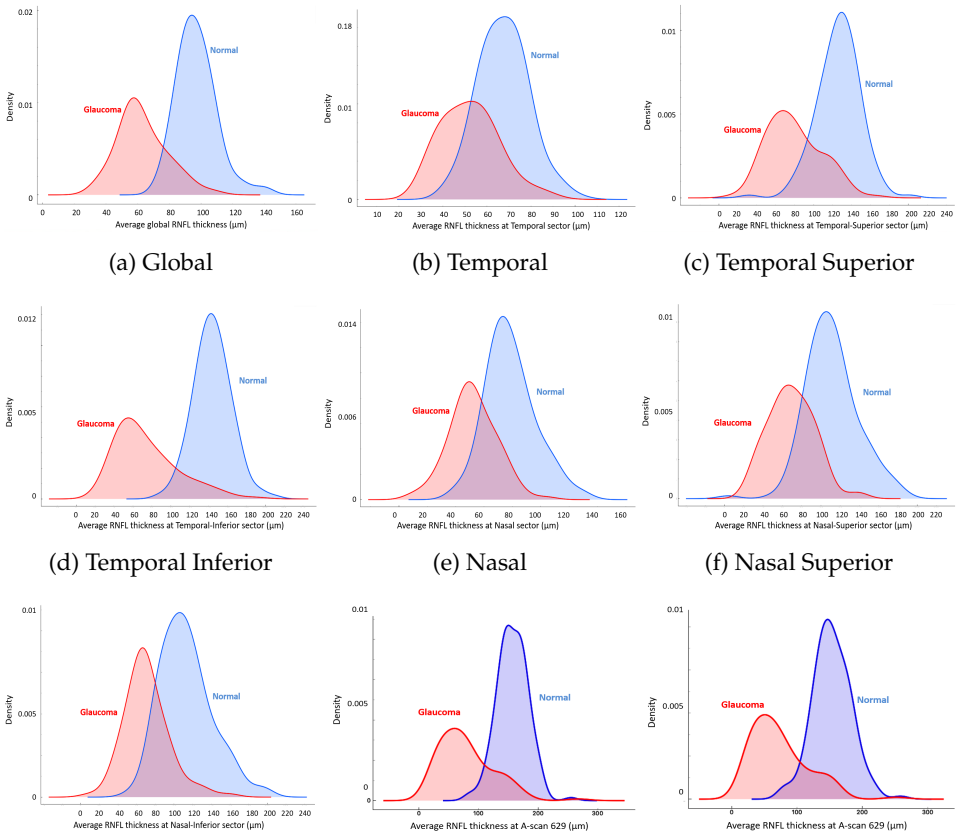
Figure 6.1: Design study of the hybrid machine learning model.

Table 6.1: Demographic and clinical information.

Parameter	Normal	Glaucoma	P value
Number of eyes	154	97	-
Number of subjects	146	64	-
Age (years)	51.9 (13.5)	68.9 (9.3)	<0.001
Mean deviation (MD; dB)	-0.6 (1.3)	-8.6 (6.7)	<0.001
RNFL thickness Global (µm)	97.2 (12.3)	63.1 (15.3)	<0.001
RNFL thickness at Temporal (µm)	97.5 (11.2)	52.4 (12.9)	<0.001
RNFL thickness at Temporal-Superior (µm)	126.5 (22.3)	79.7 (28.2)	<0.001
RNFL thickness at Temporal-Inferior (µm)	144.1 (19.7)	74.5 (32.5)	<0.001
RNFL thickness at Nasal (µm)	78.8 (17.3)	52.2 (16.4)	<0.001
RNFL thickness at Nasal-Superior (µm)	111.6 (26.8)	69.6 (23.6)	<0.001
RNFL thickness at Nasal-Inferior (µm)	113. (27.4)	70.8 (23.5)	<0.001

It should be noted the number of visits, follow-up length, and baseline *mean deviation* (MD) were significantly different in non-converted and VF-converted eyes. Remember, we excluded the visits after the VF conversation date from the VF-converted eyes.

Now, let us have a look at the slope of the RNFL thickness of seven instrument parameters. Analysing the data and illustrating the RNFL data flow on different diagrams can distinguish healthy from glaucoma patients (see Fig. 6.2). How do we distinguish eyes in two groups from global RNFL thickness? The challenging area is the overlap section.



(a) Global (b) Temporal (c) Temporal Superior

(d) Temporal Inferior (e) Nasal (f) Nasal Superior

(g) Nasal Inferior (h) A-scan629 (i) A-scan638

Figure 6.2: The histogram shows the distribution of the mean RNFL thickness in different sectors.

We first examined the "global RNFL loss" in non-converted versus converted eyes. Histogram 6.2a shows the distribution of the mean RNFL thickness slope in non-converted eyes (in light gray) versus VF-converted eyes (in dark gray). The mean values of the slopes of these two groups are significantly different (at the .05 level). Here, we have overlaid the distribution of the slopes of RNFL thickness of these two groups. As you can see, the distribution of the slopes of non-converted eyes (in blue) is pretty symmetric while the distribution of the slopes of the VF-converted eyes (in red) is not symmetric with a longer tail towards the left.

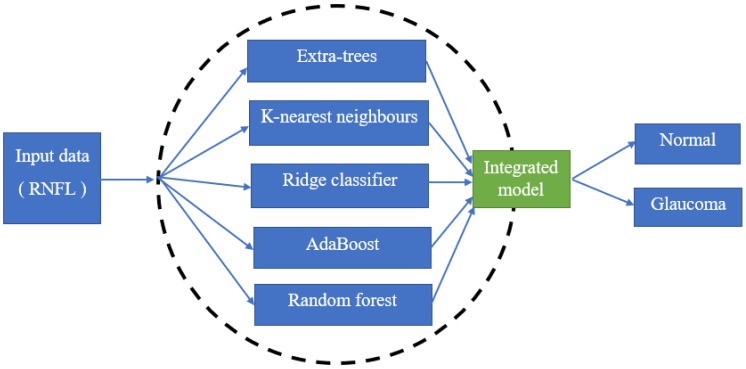


Figure 6.3: Diagram of the hybrid machine learning model for identifying glaucoma from retinal nerve fibre layers (RNFL) thickness measurements.

This, in part, indicates that the VF-converted eyes have faster RNFL loss compared to non-converted eyes. The important point here is that, although the mean of the RNFL slopes in these two groups are significantly different (p-value less than .05), there is a significant overlap between the RNFL slopes of the eyes in these two groups. As you can see, the same trend is available on the other histograms in Fig. 6.2.

The method of the proposed hybrid model is based on the processing of retinal nerve fibre layer (RNFL) thickness measurements from 154 healthy eyes and 97 eyes with glaucoma was acquired with Spectralis OCT. We developed five independent machine learning classifiers (discussed in Chapter 2) including extra-trees, K-nearest neighbours, ridge classifier, AdaBoost, and random forest and integrated them into a hybrid model to generate a more informed decision (see Fig. 6.3).

The reason why five classifiers were chosen was that 26 different classifiers were developed and based on 5-fold cross-validation and area under the receiver operating characteristics (ROC) curve metrics. We chose the top five of them to be integrated into each other based on the ensemble model to boost the result and have a robust model. We used a total of 733 RNFL measurements as input to the machine learning classifiers consisting of RNFL thickness in six sectors, average RNFL thickness, and 726 peripapillary A-scans. We used 5-fold cross-validation to develop and test the models. We examined the accuracy of the hybrid model and each independent classifier using annotated instances and ROC.

6.2 Ensemble Classifier on the SuperClinic Dataset

Aim and Materials Overview

Structural and functional methods are utilized to detect and monitor glaucomatous damage. The relationship between these detection measures is complex and differs between individuals, especially in early glaucoma [157]. Cross-sectional studies, comparing the visual field with optic disc appearance, established that there is a relationship between structural aspects of the optic nerve head and functional areas of the visual field [125, 130].

However, although some longitudinal studies have demonstrated structural change before functional change [231], other studies have shown the opposite [106]. These findings demonstrating objective detection of early structural change depends upon the variability of the tests involved, in addition to the criterion that is used to determine that structural change has occurred.

Due to the irreversible nature of the pathological changes, early diagnosis is imperative in order to preserve functional vision [157]. While glaucomatous structural damage can be assessed subjectively by clinically examining the *optic nerve head* (ONH) and *peripapillary retinal nerve fibre layer* (RNFL), the introduction of ocular imaging modalities into clinical management has allowed for supplemental objective and quantitative evaluation of ocular structure [46].

We used a fused pattern recognition model where multiple supervised models are combined to develop a hybrid model [180]; this was used to evaluate RNFL thickness in healthy eyes and those of patients with glaucoma. As a result, we classify towards diagnosing glaucomatous eyes and healthy eyes and develop a proposed model to detect glaucoma accurately with two properties, robustness and generalisation.

In this section, we aim at evaluating the relationship between retinal nerve fibre layer (RNFL) thickness and glaucoma patients. Thus, we develop a fused pattern recognition model to detect healthy vs. glaucoma patients. We also used the SuperClinic dataset that includes 175 eyes of 87 (42 female and 45 male) healthy and glaucoma patients. We propose an ensemble machine learning approach for healthy and glaucoma patients by applying and comparing supervised and unsupervised models.

Statistical Data Analysis

We present the complexity of data in two different classes based on the $Z - score$ [190] to compare across RNFL measures. There are 175 RNFL in the

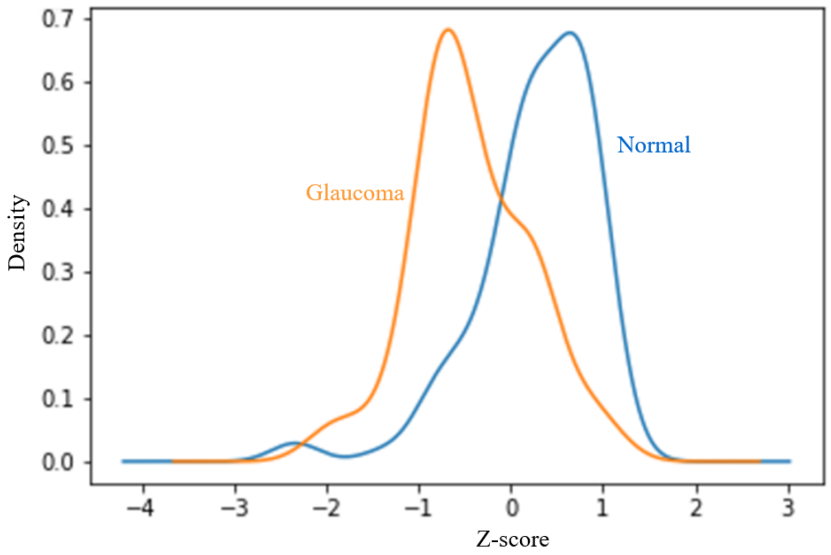


Figure 6.4: RNFL distribution of normal versus glaucoma patients.

SuperClinic database which has a non-homogeneous distribution of healthy class versus glaucoma class cases and this dataset was presented in Chapter 3.

There are 1,024 attributes for each eye, indicating the thickness of the retina at points in the circle to the centre of the optic nerve and within a radius of six micrometres. Also, data distribution with PCA (described in Chapter 2) to reduce the dimensionality of the data is shown in Fig. 6.4 and 6.5.

We also calculated the *Z-score* (see Eq. (6.1)) for all 1,024 features of each value in the normalized data set; then the mean of each record was determined. Also, we applied a *kernel density estimation* (KDE) function to generate the probability density function for density estimation to present the data distribution for each class. Afterwards, we used the Gaussian kernels for visualising the distribution of those high-dimensional data [222].

What is a Z-Score? A Z-score is a numerical measurement used in statistics of a value's relationship to the mean (average) of a group of values, measured in terms of standard deviations from the mean. If a Z-score is 0, it indicates that the data point's score is identical to the mean score. A Z-score of 1.0 would indicate a value that is one standard deviation from the mean. Z-scores may be positive or negative, with a positive value indicating the score is above the mean and a negative score indicating it is below the mean. Also, Z-scores are measures of an observation's variability.

Therefore, the Z-score is the distance of the standard deviation away from the mean, and it is calculated by

$$z = \frac{\mathbf{x} - \mu}{\sigma} \tag{6.1}$$

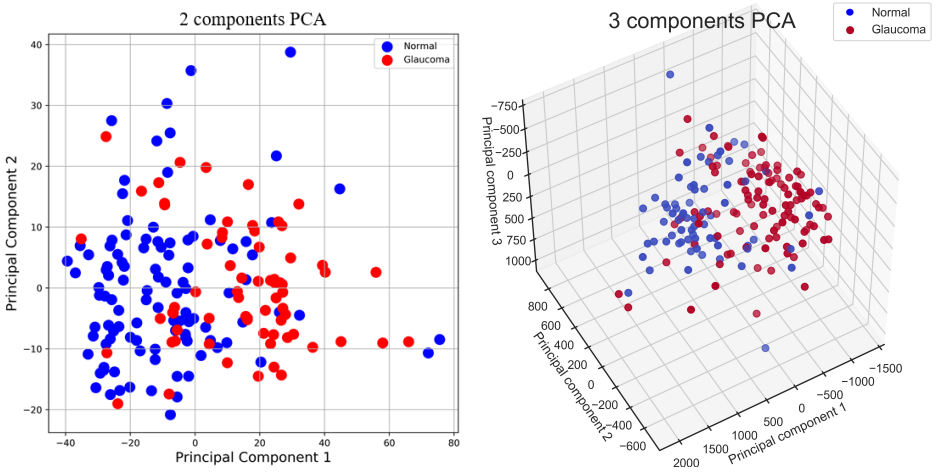
where **x** represents the data point (RNFL value for each feature) from the data set, μ is the mean, and σ is the standard deviation.

For a data point, a Z-score helps to point out how "unusual or usual" a data point is compared to other values. Also, Z-scores have a normal distribution curve where the mean is zero and the variance is equal to 1, so Z-scores do not follow the normal distribution unless the original data follows the normal distribution.

Figure 6.4 shows an RNFL standard normal probability distribution of healthy versus glaucoma patients. There is a significant overlap between healthy and glaucoma RNFL thickness data. Therefore, we proposed a hybrid classifier to detect glaucoma accurately.

Besides, the number of variables to analyse the data is 1,024. Thus, to reduce the dimensionality of each of them, we used *principal component analysis* (PCA) [101]. The overlapping section is the challenging area to classify eyes into two groups based on global RNFL thickness (see Fig. 6.5).

Also, the first three components of PCA (described in Chapter 2) show the complexity of the data. Figure 6.5 on the bottom illustrates that there is high complexity in our data set so we were not able to classify the data correctly by



(a) 2–PCA components (b) 3–PCA components

Figure 6.5: The complexity of data distribution based on 2– and 3–PCA components.

using a plane in 3-dimensional space.

Methods

In Fig. 6.6, we sketch our overall approach for automated glaucoma diagnosis based on a fused pattern recognition model with RNFL data.

Overview of Hybrid Model

The schematic diagram is a mind map of the proposed model that represents the main components of the process. RNFL data are entered as input numeric values. Then, data cleaning is performed on input data as a pre-processing step. Data cleaning has been done on missing values that includes dropping missing features and filling in not available values with any number that comes directly after it in the same row. Afterwards, data is fed into the fused pattern recognition model. There are two stages in the proposed model: unsupervised and hybrid supervised phases. The hybrid model was trained on 75% of the data. Meanwhile, all data are used for clustering by using a K-means model.

The proposed hybrid system works with an ensemble model that includes the top ten classifiers with high 5-fold cross-validation accuracy and 5-fold cross-validation F1 scores. All results of applied models in one particular ensemble are amalgamated to acquire hard votes for all models together. Even though we

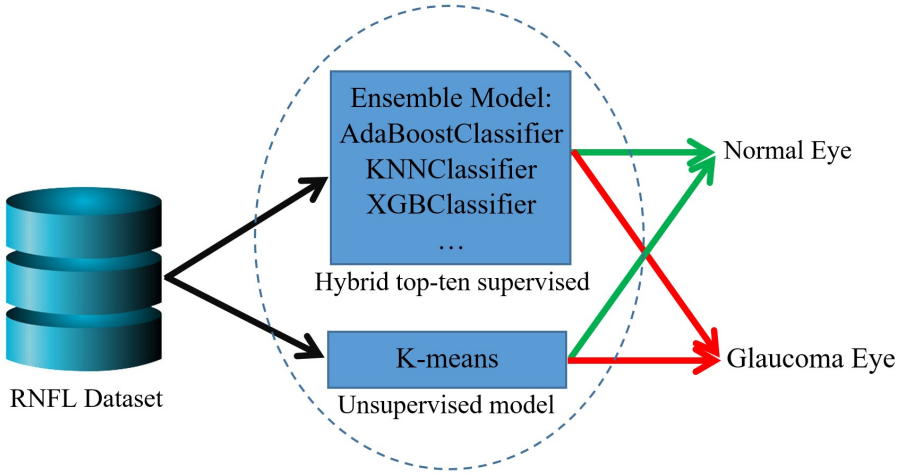


Figure 6.6: Overall approach based on the proposed fused pattern recognition model.

have high accuracy in an individual model, we apply the ensemble model to be more robust and generalised in case of our relatively small data set. Appropriately, the proposed ensemble model accomplishes the integrated output with high accuracy and F1 score by applying 5-fold cross-validation. The selected classifiers for our ensemble model are listed in Table 6.2 which we modified elements to get a better result. The top ten selected classifiers for our ensemble model are listed in Table 6.7 on Section 6.3 which we modified elements to get a better result.

We also determine the hyper-parameters of the classifiers with nested loops defined by the upper and lower bound of the elements to get the highest performance. We are combining results of classifiers with the mean of the probability of top ten classifiers to generate a hybrid model that could improve robustness and a generalisation of the final model as the main advantage of this research.

Table 6.2: Examples of tuned elements (hyper-parameters) include classifiers that are used on the tuned hybrid model.

Our supervised model	**Modified elements**
XGBClassifier_robust	random_state=0 , max_depth = 3 , learning_rate = 0.1, n_estimators= 50, silent =True , objective= 'binary:logistic', booster='gbtree', min_child_weight= 3, max_delta_step= 1
AdaBoostClassifier_robust	algorithm="SAMME", learning_rate=0.1 , random_state=0 , n_estimators=500
RandomForestClassifier_robust	criterion='entropy', max_depth=1, max_features='log2', random_state=0,warm_start=True, n_estimators=50, bootstrap=True
KNeighborsClassifier_robust	n_neighbors=13, weights='uniform' , algorithm='auto' , leaf_size=5
RidgeClassifier_robust	max_iter=100, fit_intercept=True, normalize=True
svm_svc_robust	C = 1 , kernel= 'rbf', gamma = 'scale', probability = False, max_iter= -1, decision_function_shape= 'ovo', random_state=0
ExtraTreesClassifier_robust	criterion='entropy', max_depth=1, max_features='auto',random_state=0, warm_start=True, n_estimators=50, bootstrap=True)
RidgeClassifiercv_robust	max_iter=100, fit_intercept=True, normalize=True
NearestCentroid__robust	metric='cityblock', shrink_threshold=None

The test data set was entered into the tuned hybrid classifiers to detect healthy and glaucoma eyes. Even though the size of the testing set is not very large, we used 5-fold cross-validation on hole data to detect the top ten classifiers that can effectively have high accuracy on our data set.

Also, we used all of our labelled data set to evaluate the performance of our tuned unsupervised model, and the results on the K-means model compare with labels that show us 80% accuracy with our tuned unsupervised model.

Proposed Hybrid Supervised Model

RNFL thickness measurements from 107 healthy eyes and 68 eyes with glaucoma were acquired with Spectralis OCT. We developed 26 independent machine learning classifiers including the top ten listed in Table 6.7.

We integrated the result of those well-tuned 26 classifiers into an ensemble learning model to generate a "more informed" decision. We used a total of 1,024 RNFL measurements as input for the machine learning classifiers. We used 5-fold cross-validation to find out the top ten models and test the models and examined the accuracy and the F1 score of the hybrid model and each independent classifier to have a robust and generalised model.

We selected the top ten most discriminating classifiers based on the best 5-fold cross-validation F1 score and accuracy measures. The hybrid model was generated by ensemble models of the top ten classifiers which are mentioned based on the accuracy of 5-fold cross-validation: AdaBoost, GridSearchcv (cv; cross-validation), XGBoost, RandomForest, NuSVC, KNN, RidgeClassifier, SVM_svc, ExtraTrees and SVC. This defines our stack of top ten classifiers to achieve an accurate model.

Proposed Unsupervised Model

We proposed an unsupervised model based on a tuned K-means classifier for screening. This model was trained based on two clusters and without any random state. Finally, the predicted labels were compared with the real labels annotated by an ophthalmologist, and we obtained 80% performance based on the unsupervised tuned model. This shows that the proposed model can be used for screening to aid ophthalmologists. Also, the unsupervised model can be used by optometrists for screening participants and refer suspected glaucoma cases to do more medical examinations and be checked by an ophthalmologist accurately.

Validation of Our Study in SuperClinic Dataset

We divided the data into 75% for training and 25% for testing for validation of our hybrid model, where the test set was not to be used during training. The evaluation

has been done based on the test set. In addition, K-fold cross-validation will be conducted to tune the parameters using the training data.

Multiple measures were used for performance evaluation (evaluation metrics are described more in Chapter 2) of the proposed model such as accuracy (AC), specificity (SP), sensitivity (SE), and F1 score (F1).

Hybrid and boosting Classifiers

In this section, we discuss the state-of-the-art boosting models with selected classifiers as mentioned in the Section 6.2 and demonstrate the results of our proposed model in the next Section 6.3. Moreover, the selected classifiers are discussed in Chapter 2.

To illustrate the concept of boosting, let us give an example. Suppose we have a series of sample test questions, and use three classifiers such as: KNN, decision tree and NN. From the beginning, we will read and learn with the answers given marking the ones that are most difficult for us to read for later review. The same concept applies to the boosting classifiers. In this first stage of the classifier method, we assume; KNN classifier observes a sample of data (using a method such as the bagging) and constructs its own classifier. Then, it is given the same unseen examples from the training set and this classifier is likely to misunderstand some of the examples. Now, to select the subset of data for the second classifier (decision tree) in the boosting method, those samples failed by the first classifier are, it is more likely to be selected for the second classifier. The harder samples are more likely to be selected than the simpler ones. Similarly, to create a subset of data for the third classifier, those that were more difficult in the first and second classifiers are more likely to be selected [280]. Hybrid classifiers generally prevent the algorithm from over-fitting, and in many cases produce better results than other algorithms.

6.3 Results and Discussion

Results on UCLA Dataset

We assessed the ROC curves of RNFL thickness slopes of all instrument parameters in these two groups (see Table 6.3). Indeed, the ROC curves confirmed all our previous findings. Moreover, the area under the ROC curve of all of the instrument parameters (RNFL sector) were not any better and there is no chance of improving by separating the non-converted from VF-converted eyes (see Table 6.4). Machine

learning outcomes can be seen in Table 6.5, and the result of the proposed machine learning model is significantly convincing.

The sensitivity of the hybrid model was 92% for detecting glaucoma eyes correctly and specificity was 98% for detecting healthy eyes correctly. The area under the ROC curve of the hybrid model was 0.97 (see Fig. 6.7). The most discriminating RNFL thickness inputs to the classifier were average RNFL thickness, temporal-inferior RNFL thickness, A-scans 60, 231, and 296, respectively. We need to somehow define the topography of these A-scans.

Also, as a quick discussion, when data is good (features are discriminative), the type of classifier is not a bottleneck. Whenever the features are too overlapping, the role of the classifier is more critical. A machine learning classifier can learn from RNFL data and provide diagnostic information accurately. Combining machine learning classifiers and generating hybrid models could improve the robustness and generalisability of the final model. It should be noted, when conventional classifiers are performing convincingly, there is no reason for using deep learning.

Results on SuperClinic Dataset

Moreover, in this study on the SuperClinic dataset, we achieved an F1 score of 0.82 and an accuracy of 82% using 5-fold cross-validation on a data set of 107 RNFL data from healthy eyes and 68 RNFL data from eyes with glaucoma; 25% of the data had been selected randomly for testing. We also used different evaluation measures [187] to show the probability that a patient has glaucoma or not. We

Table 6.3: The area under the ROC curve of RNFL thickness at different sectors; the best area under the ROC curve is 0.91.

RNFL sector	**Area under the ROC curve**
Average RNFL thickness Global	0.91
Average RNFL thickness at Temporal-Inferior	0.91
Average RNFL thickness at Temporal-Superior	0.88
Average RNFL thickness at Nasal-Inferior	0.88
Average RNFL thickness at Nasal	0.86
Average RNFL thickness at Nasal-Superior	0.86
Average RNFL thickness at Temporal	0.79

Table 6.4: The area under the ROC curve of RNFL thickness ad different A-scans; the best area under the ROC curve is 0.90.

RNFL sector	Area under the ROC curve
RNFL thickness at A-scan 629	0.90
RNFL thickness at A-scan 638	0.90
RNFL thickness at A-scan 639	0.90
RNFL thickness at A-scan 640	0.89

Table 6.5: Machine learning with the hybrid model and traditional classifiers outcomes.

Machine learning method	Area under the ROC curve
Hybrid model	0.97
Extra-trees	0.96
K-nearest neighbours	0.96
Ridge classifier	0.94
AdaBoost	0.94
Random forest	0.94

obtained the results (shown in Table 6.6) by our proposed robust model.

Table 6.6: Results of our hybrid model based on 1,024 features for 175 patients.

F1_5cv	Acc_5cv	F1	Acc	Recall	Precision
0.82	81.7	0.82	81.8	0.82	0.82

Sensitivity or recall was 82% for detecting glaucoma eyes correctly. Also, the F1 score for 5-fold cross-validation was 0.82 for detecting the top ten classifiers to obtain healthy versus glaucoma eyes successfully, and the F1 score is usually more indicative than accuracy.

All results of the top ten selected classifiers for our ensemble model are listed in Table 6.7.

We also used the top ten classifiers for testing of 25% of unseen data with different evaluation measures. The confusion matrix (see Fig. 6.8) shows the

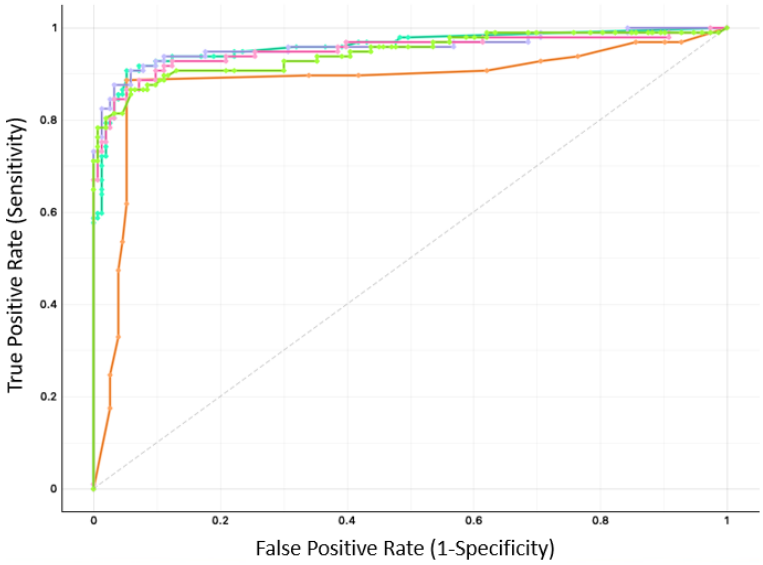


Figure 6.7: ROC curves of the hybrid machine learning framework for detecting glaucoma from RNFL thickness measurements; K-nearest neighbours (orange colour), extra-trees (light green colour), ridge classifier (pink colour), random forest (purple colour), and AdaBoost (dark green colour).

performance of the proposed model.

Based on the results of our proposed fused pattern recognition model, we demonstrate that our model can work on a variety of new data from the same device with 80% accuracy. We also obtained 80% accuracy with the unsupervised model which is quite promising because ophthalmologists usually can classify 80% of patients based on one modality data.

Detailed Comparison and Discussion

Although, there are the same types on both datasets (text data) the number of features and complexity of patients' data are different and, because of that, the performance of the results is not the same as each other. We also added one more XGBClassifier in to a proposed ensemble model to get better performance on the SuperClinic dataset. This classifier boosts the performance of our outcome. Additionally, we used on both datasets different evaluation metrics to tune and optimize the model. Finally, our proposed model shows that performance on it is over 80% which is valuable for glaucoma screening and aiding ophthalmologists.

Table 6.7: Results of the proposed hybrid model based on the top ten of 26 classifiers.

Proposed top ten models		All data		Test data set							
No	**Models**	**F1_5cv**	**Acc_5cv**	**F1**	**Acc**	**Recall**	**Precision**	**TN**	**FP**	**FN**	**TP**
1	**AdaBoostClassifier_robust**	**0.82**	**82.3**	**0.77**	**77.3**	**0.77**	**0.77**	**19**	**4**	**6**	**15**
2	**GridSearchCV**	**0.82**	**81.7**	**0.82**	**81.8**	**0.82**	**0.82**	**20**	**20**	**3**	**5**
3	**XGBClassifier_robust**	**0.82**	**81.8**	**0.77**	**77.3**	**0.77**	**0.77**	**19**	**4**	**6**	**15**
4	**RandomForestClassifier_robust**	**0.82**	**81.7**	**0.75**	**75.0**	**0.75**	**0.76**	**20**	**3**	**8**	**13**
5	**NuSVC**	**0.82**	**81.2**	**0.82**	**81.8**	**0.82**	**0.82**	**1**	**20**	**3**	**5**
6	**KNeighborsClassifier_robust**	**0.82**	**81.2**	**0.80**	**79.6**	**0.80**	**0.80**	**20**	**3**	**6**	**15**
7	**RidgeClassifier_robust**	**0.81**	**81.1**	**0.82**	**81.8**	**0.82**	**0.83**	**21**	**2**	**6**	**15**
8	**svm_svc_robust**	**0.81**	**81.1**	**0.82**	**81.8**	**0.82**	**0.82**	**20**	**3**	**5**	**16**
9	**ExtraTreesClassifier_robust**	**0.81**	**81.2**	**0.75**	**75.0**	**0.75**	**0.76**	**20**	**3**	**8**	**13**
10	**SVC**	**0.82**	**80.6**	**0.75**	**75.0**	**0.75**	**0.76**	**1**	**20**	**3**	**8**
11	RidgeClassifiercv_robust	0.80	80.6	0.77	77.3	0.77	0.79	21	2	8	13
12	LogisticRegression	0.78	78.3	0.72	72.7	0.73	0.74	1	20	3	9
13	GaussianNB_robust	0.77	77.2	0.73	72.7	0.73	0.73	18	5	7	14
14	NearestCentroid__robust	0.77	76.5	0.77	77.3	0.77	0.77	18	5	5	16
15	MLPClassifier	0.76	76.5	0.77	77.3	0.77	0.78	20	3	7	14
16	LinearSVC	0.75	74.9	0.68	68.2	0.68	0.68	17	6	8	13
17	OneVsRestClassifier	0.74	74.9	0.77	77.3	0.77	0.77	19	4	6	15
18	Perceptron	0.73	74.8	0.70	70.5	0.71	0.71	15	8	5	16
19	PassiveAggressiveClassifier	0.73	73.1	0.65	65.9	0.66	0.67	19	4	11	10
20	GradientBoostingClassifier	0.72	72.6	0.72	72.7	0.73	0.73	1	19	4	8
21	SGDClassifier	0.72	73.6	0.53	61.4	0.61	0.78	23	0	17	4
22	BaggingClassifier	0.71	70.9	0.65	65.9	0.66	0.66	18	5	10	11
23	DecisionTreeClassifier	0.68	68.2	0.63	63.6	0.64	0.64	1	18	5	11
24	LinearDiscriminantAnalysis	0.63	62.9	0.64	63.6	0.64	0.64	1	16	7	9
25	QuadraticDiscriminantAnalysis	0.49	50.3	0.38	45.5	0.46	0.39	1	18	5	19
26	BernoulliNB	0.43	53.2	0.40	52.3	0.52	0.51	22	1	20	1

6.4 Summary

We used UCLA dataset to train a hybrid machine learning classifier that detected glaucoma with high specificity and sensitivity based on RNFL thickness measurements. In other words, a stack of five models ensemble-voting on the outcome can be more robust and generalisable. This approach could be valuable in detecting glaucoma in clinical practice and for research purposes.

We also propose the tuned ensemble method on the SuperClinic dataset and we did data cleaning prior to analysis to obtain a better performance. Our results were approved by an ophthalmologist. The performance of our model is 82% and this documents two features, robustness and generalisation, which points to the usefulness for screening. The proposed model is ready to be retested on large and diverse data sets and can assist the ophthalmologists in their daily examination tasks to confirm their diagnoses, thereby increasing the accuracy of diagnosis. This study also shows that the proposed ensemble model and K-means performs well

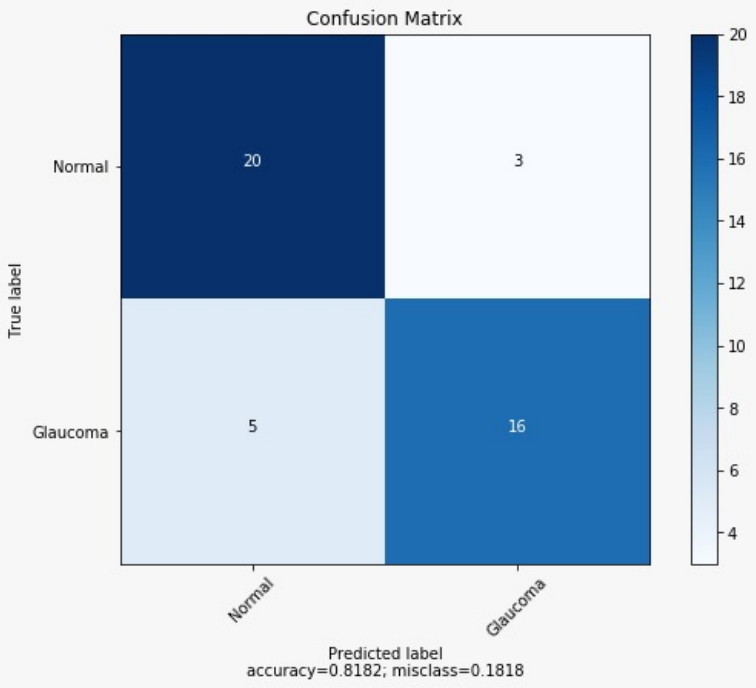


Figure 6.8: Confusion matrix based on the proposed hybrid model on testing data.

for real-time implementation of a classifier for healthy and glaucoma eyes. It is also suitable for the development of medical devices based on IoT (*internet-of-things*) and remote monitoring via a web interface. Finally, the contributions of this chapter can be summarised as follows:

- Classification of healthy versus glaucoma patients through RNFL data with 1,024 feature points with 0.82 of F1 score
- Visualisation of our complex data with Z-score metric in both glaucoma and healthy classes
- Robustness and generalisability improvement through results combination of traditional classifiers to generate the hybrid

There are not any public RNFL data sets. There is also some related research on private and not large data sets [60, 149, 180] and available high accuracy results on this research because of less complexity of data, but we got promising results based on small, complex and high dimensional data with the ensemble technique.

Moreover, the accuracy of our result is not quite high, but it can generally detect over 80% of ophthalmologist's diagnoses. Additionally, other studies [60, 149, 180] used different data features and complexity and that was why no comparison results with other state-of-the-art algorithms are given. For the future, we will use time series and multi-modal retina data for the prognosis of glaucoma disease and will do retests on our proposed model on the new data set with the same data structure.

Chapter 7

Transfer Learning in Glaucoma Detection

This chapter aims to develop a machine learning-based algorithm for glaucoma diagnosis in patients with lack of annotated fundus images, so we use the transfer learning model. In order to build a deep transfer learning model to diagnose glaucoma using a fundus image, we re-designed and tuned an optimized transfer learning model to detect glaucoma efficiently. First, we propose a system for clinical testing based on an automated glaucoma diagnosis using the transfer learning technique and validate it with test and retest data sets. Then, we make a comparison of transfer learning techniques, deep convolutional neural network, and multilayer neural network methods for the diagnosis of glaucomatous optic neuropathy. There is also a potential to have a tool/web-application glaucoma prediction via traditional and deep machine learning models based on the cloud.

Moreover, we used the developed deep learning framework that can automatically identify clinically relevant biomarkers on glaucoma fundus images. Material discussed in this chapter has been published in my publications [178,181,182].

7.1 Glaucoma Patients and Rules for Assessment of Fundus

Glaucoma, the second leading cause of blindness in the world, is a group of optic neuropathy disorders that lead to loss of vision if left untreated [7,136].

Early glaucoma diagnosis prevents permanent structural optic nerve damage and consequent irreversible vision impairment. Longitudinal studies have described both baselines structural and functional factors that predict the development of glaucomatous change in ocular hypertensive and glaucoma suspects. Although there is neither a gold standard for disease diagnosis nor progression, photographic assessment of the optic nerve head remains a mainstay in the diagnosis and management of glaucoma suspects and glaucoma patients.

Because of the rapid increase in aging populations, accurate diagnosis is critical for making treatment decisions to preserve vision and maintain quality of life [123,

126,240,279,287,314]. Stereoscopic disc photos provide an appropriate record of the optic nerve, independent of the specialized viewing instrument [171, 173, 239, 255, 279, 287].

Stereoscopic disc photos remain one of the most widely-used and accepted methods for documentation of the optic nerve head [127, 171]. However, due to its subjective nature, assessment of optic disc photographs for the presence of glaucoma is labour-intensive and prone to interpretation errors. From a clinical perspective, many eye care specialists prefer to have access to more objective analyses for glaucoma diagnosis. Five rules for assessment of fundus stereo-photographs to identify glaucoma and monitor its progression over time have been described by Fingret et al. [38, 239].

Identifying glaucomatous optic neuropathy (GON) based on ONH photographs is one of the standard methods used for glaucoma diagnosis [38]. This process is labour-intensive and biased by reader variations.

Optic disc change supersedes visual field testing as a reference standard in both disease diagnosis and progression [163, 219]. Structural change is detected earlier than visual field abnormalities in over half of patients progressing to an initial diagnosis of glaucoma [100]. Optic disc photography is the most commonly used technique to objectively document structural ONH damage due to its reproducibility and low cost [73]. In addition it has been used as an endpoint in randomized clinical trials: the *ocular hypertension treatment study* (OHTS), the *early manifest glaucoma trial* (EMGT), and the *European glaucoma prevention study* (EGPS) [219].

We present a decision support tool for differentiating normal optic nerves from those diagnosed with glaucoma/glaucoma suspects. The extracted parameters include cupping the area, and vertical and horizontal cup to disc ratio. This study was conducted in order to detect slight changes in these parameters automatically using *digital fundus image* DFIs and dividing images into two different categories of glaucoma/glaucoma suspects and normal optic nerves. The *optic disc critical region* (ODCR) is the small region of the DFIs, and processing ODCRs takes less time in comparison with the processing of the DFIs.

We applied different machine learning algorithms such as transfer learning models, deep *convolutional neural network* (CNN), and deep *multilayer neural network* (MNN) to the ODCRs and extracted clinically relevant parameters automatically. The ODCR parts are available in DFIs in the RIM-ONE data set as input to apply to the deep learning algorithms. Generally, deep CNN was developed for image segmentation and classification [53, 75].

Transfer learning of machine learning techniques restores the weights and labelled data which are trained on the ImageNet dataset [64]. It can be applied to

other tasks that are complicated such as glaucoma classification and robust optic disc segmentation [251]. Moreover, transfer learning models have been improved and developed to be applicable to small data sets. Recently, research studies indicated that many approaches that have attempted to segment the optic cup and optic disc have used unsupervised machine learning [175].

Recent advances in artificial intelligence and significant growth in available data have enhanced identification of ocular disorders including glaucoma diagnosis. In particular, deep learning techniques can identify highly complex patterns to detect various ocular pathologies [6, 143].

In this chapter, we propose an automated software application based on deep learning and transfer learning that can differentiate between healthy eyes and those with glaucoma using ONH photographs. We selected the regions of interest within the ONH photographs, namely regions which included the cup. In fact, the *cup-to-disc ratio* (CDR) is one major parameter for identifying glaucoma [239]. Besides, we compared the transfer learning models and deep convolutional neural network and deep neural network to convince that the proposed model based on transfer learning is accurate, robust and objective.

7.2 Materials

We used two independent data sets as described below, from universities in the USA and in Germany/Czech Republic.

The first data set was obtained from patients with healthy eyes and those with glaucoma who visited the glaucoma clinic at the *University of California Los Angeles* (UCLA) and it was presented in Chapter 3.

The second data set is the *high-resolution fundus* (HRF) data set that is publicly available. This data set was also presented in Chapter 3.

RIM-ONE is the third data set that we used in this chapter to compare the transfer learning models with deep CNN and NN; we employed a public retinal image database on optic nerve evaluation for two healthy and glaucoma categories that are exclusively focused on *optic nerve head* (ONH) segmentation. The second version (V2) of RIM-ONE data set [88] was presented in Chapter 3 as well.

In the following sections of this chapter, we propose an automated software application based on deep learning and transfer learning that can differentiate between healthy eyes and those with glaucoma using ONH photographs. We selected the regions of interest within the ONH photographs, namely regions which included the cup. In fact, the *cup-to-disc ratio* (CDR) is one major parameter for identifying glaucoma [239]. In addition, we compared the transfer learning models and deep convolutional neural network and deep neural network with

each other to convince that the proposed model based on transfer learning is accurate, robust and objective.

7.3 Very Deep Pre-trained Networks

The transfer learning (pre-trained) method is a state-of-the-art supervised machine learning technique. This model is developed and trained on a massive standard data set (ImageNet) and then used for other tasks. Pre-trained algorithms are first trained on that data set, and then weights and parameters are saved. Thus, it applies to deploy a pre-trained algorithm on glaucoma detection, but results are dependant on setting the hyper-parameters so it is tricky to get trustworthy results. Pre-trained methods do not require segmentation of DFI structures, so these algorithms can serve as prior knowledge for measuring different glaucomatous symptoms associated with fundus segmentation.

Convolutional neural networks (CNNs) have been widely used for image segmentation and classification [52, 53, 154, 200]. Transfer learning is implemented in developing deep learning frameworks to address restrictions due to the limited number of input samples as well as computational resources for running deep learning techniques. Transfer learning employs the weights and parameters that were learned from previous large labelled data sets and applies them to the new task [251].

We use very deep learning along with transfer learning to detect glaucoma from ONH photographs. Transfer learning (transferring the pre-trained parameters and weights to a new deep learning model) is a state-of-the-art machine learning technique that is used widely to train deep learning approaches. Thus, we have briefly explained in this section about pre-trained models that apply to the next section.

VGG19

VGG19 [223] has been widely used for different applications. As its name implies, VGG19 has 19 layers, with 16 convolutional layers and three fully connected layers [313].

VGG19 accepts a default input size of 244×244 for a colour image and it has been recently used for glaucoma detection on visual field tests [153, 223]

In this research, the input size is modified to 299×299 and the number of layers is extended up to 25 layers to address the over-fitting problem. Similar to InceptionResNet, most of the parameters are trainable and a small fraction is kept as default.

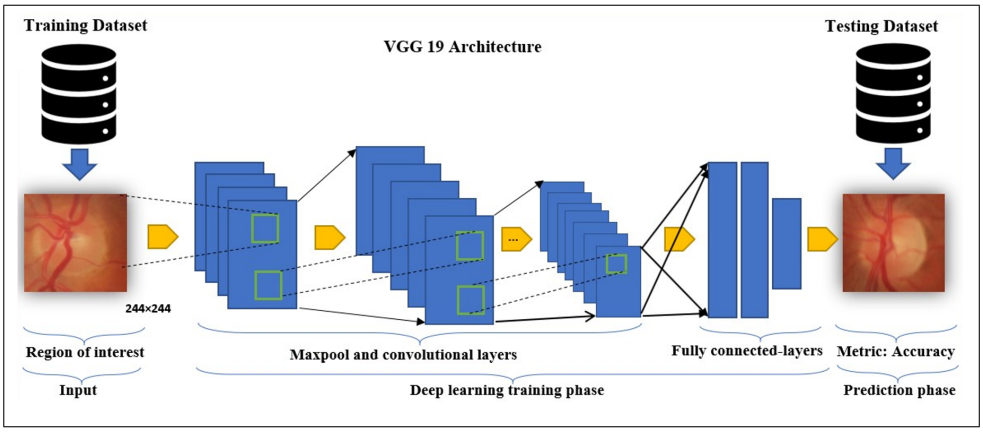


Figure 7.1: Schematic block diagram of VGG19 [54].

VGG19 has three fully connected layers at the end and all hidden layers use the *rectifier units* (ReLU) activation function. VGG19 provides a flexible architecture for different tasks. Similar to InceptionResNet, we used data augmentation in training. Figure 7.1 represents the architecture of the VGG19 model.

VGG19 is an appreciable architecture to evaluate a specific task, and it also works well on image classification for big image data sets. Scale jittering was used as one of data augmentation techniques during training. The ReLU activation function was used after each convolution layer and trained with batch gradient descent. Two training and validation scores have not converged into each other in different epochs, up to fifty. Therefore, in this case, VGG19 is an overfit model and not applicable.

InceptionResNet-V2

InceptionResNet-V2 is a very deep convolutional network (825 layers) [289] that has been employed in different image recognition tasks in recent years. InceptionResNet has multiple layers including input, output, convolutional, pooling, residual, concatenate, dropout, and fully connected layers. The default image input size is 299×299 in colour format [233]. In our study, we required most of the parameters to be trainable and only a small fraction was selected as default. In order to optimize the training computational complexity, we used a cloud-based *graphics processing unit* (GPU). Figure 7.2 shows the details of InceptionResNet layers.

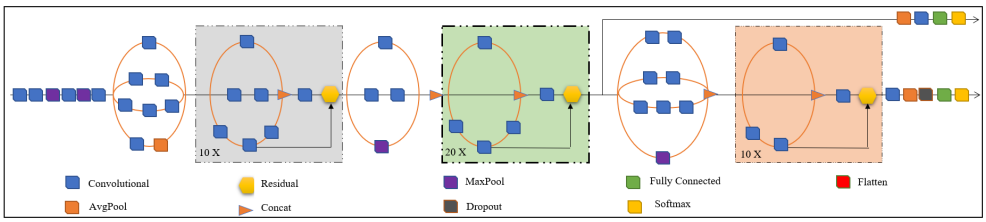


Figure 7.2: Schematic block diagram of InceptionResNet-V2 [308].

Multilayer Neural Networks

Our proposed MNN has three fully connected layers with, five, fifteen and five neurons in each layer. The ReLU activation function on the first three layers and the softmax activation function on the output layer were employed. All images are applied as an input vector to the first layer and then transferred to the hidden layers while each hidden layer is connected to preceding layers. The task of the activation function is to express an output which is given by some inputs.

Two training and validation scores have not converged into each other in one hundred epochs. Therefore, in this case, the MNN method is not applicable; it has an over-fitting problem.

Convolutional Neural Networks

ConvNets [147, 269] can learn weights and biased values from the processed images. The CNN had been employed to detect glaucoma by training seven convolutional layers with an average accuracy of 92.68 % [237] on the Drive database, which is a public data set. In our work, a deep CNN was employed with six convolution-layer blocks. In the first block, two convolution-layers are available with sigmoid activation functions, and there are two convolution-layers with ReLU activation functions in the other five blocks. At the end of this ConvNet architecture, there are three fully connected layers with a softmax activation function. It should be mentioned that the filter size is 3×3 and dropout is 5% in all blocks. Two training and validation scores have not converged into each other for the first one hundred epochs; thus, the proposed method fails to meet the objectives because of over-fitting.

Although the current CNN model works very well on big data sets, the transfer learning method can be deployed on the limited data set to get the best result.

Transfer Learning - InceptionResNet

The Inception-ResNet [233] is applicable to glaucoma detection, and both validation and training loss scores converged into each other in ten epochs. The accuracy of test data with the Inception-ResNet method on the RIM-ONE data set is 85%. In our study, we required most of the parameters to be trainable and only a small fraction was selected as a default. In order to optimize the training computational complexity, we used a cloud-based *graphics processing unit* (GPU).

Transfer Learning - Xception

Another pre-trained model to check the results of our application is Xception [59]. The Xception model is a linear stack of depth-wise separable convolutional layers with residual connections [92]. A depth-wise separable convolution is a kind of an Inception module [54] with a maximally large number of towers. This approach leads to the proposal of a novel deep CNN model inspired by Inception, where Inception modules have been replaced by depth-wise separable convolutions. It slightly improved results compared to Inception V3 on ImageNet and has the same number of parameters as Inception V3. Performance gains are not due to increased capacity but rather to a more efficient use of model parameters [59].

The default image input size is 299×299 in colour format and training and validation input images feed into the model. In our study, convs layers were frozen, so available weights of Xception were transferred from convs layers into fully connected layers. These fully connected layers were trained by input images. Accuracy and loss scores converged into each other. The best result is achieved on the 20th epoch: 90% of healthy images and glaucoma cases are correctly identified.

Proposed Method Based on Transfer Learning - NASNet

In this study, another pre-trained model was used to recheck the result of our application. NASNet [274] is a pre-trained model that can be used for small databases. Our proposed model based on NASNet is explained in detail as follows: The feeding image size is 299×299. The proposed model is modified in the training phase by an augmentation. There are four fill-mode augmentations: constant, nearest, reflect, and wrap. Due to the high performance, the best result was achieved using a proposed fill mode with "nearest" in augmentation. Furthermore, images have been horizontally filled. Other tuned factors are rotation range (equal to 40), width shift range (0.2), height shift range (0.2), shear range (0.2), zoom range (0.2), and channel shift range (10); the remaining factors are the same as NASNet defaults.

The proposed model has 85,723,220 parameters, of those, 85,523,960 are trainable and the rest of them are non-trainable parameters. Our model started by calling the NASNet-Large model with ImageNet weights. The model was flattened to build a long feature unit vector and use it by a fully connected layer to make a decision for the final classification with the softmax activation function; dropout is assumed to be 0.1. Freezing layers have been used, except in the fully connected layers. The proposed model uses Adam optimizer and a lower learning rate of $1e-5$ since our model is fine-tuning to achieve our goal which is a better performance and accuracy to detect glaucoma or non-glaucoma eyes. Three learning rates have been tested: $1e-3$, $1e-4$ and $1e-5$. Except for the $1e-5$ learning rate (that we have chosen), the other learning rates lead to over-fitting in the model.

We will compare the reported results with a number of transfer learning models and deep learning methods to identify the best performing approach.

7.4 Transfer Learning Generalisation

We developed a deep learning algorithm for identifying glaucoma on optic nerve head (ONH) photographs. Transfer learning was applied to overcome over-fitting on the small training sample size that we employed. The transfer learning framework that was previously trained on large data sets such as ImageNet, uses the initial parameters and makes the approach applicable to small sample sizes. We then classified the input ONH photographs as healthy ("normal") or "glaucoma".

Using this approach, deep learning frameworks are trained using a large standard, like ImageNet [64], then weights and parameters are saved to be used for another task. Thus, we transferred the pre-trained parameters as the initial setting of the deep learning framework and then tuned the parameters using the ONH photographs from the UCLA data set for glaucoma diagnosis.

Basically, it is challenging to train a robust model with a limited number of input images without transfer learning. In other words, transfer learning can serve as prior knowledge for measuring the difference between glaucomatous signs manifested in ONH photographs of eyes with glaucoma and those from healthy eyes.

The block diagram in Fig. 7.3 represents our proposed approach to automatically diagnose glaucoma using ONH photographs. Training and validation data of the UCLA data set were entered as input images. Then, all images were cropped as the pre-processing step. Afterwards, data were fed into VGG19 and InceptionResNet-V2 models, then hyper-parameters (features on deep learning models that should be initialized such as learning rate, batch size, and so forth) were tuned to have accurate and tuned classifiers. Finally, the test and retest

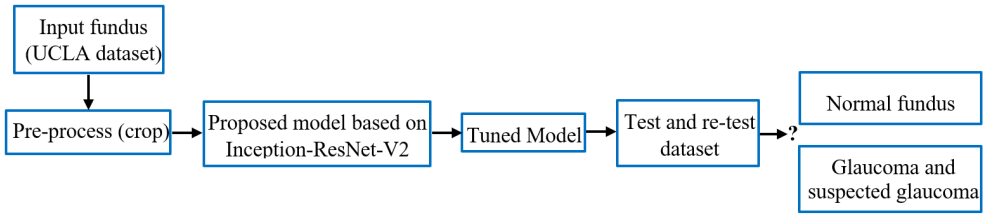


Figure 7.3: The overall approach for automated glaucoma diagnosis.

data sets were entered into the tuned classifiers to detect healthy and glaucoma eyes.

We used two deep learning models; VGG19 and InceptionResNet-V2. Our motivation is to identify which model performs better when we use a small number of input fundus images.

ONH photographs were entered into both VGG19 and InceptionResNet-V2 models. We kept the initial parameters of these two models and trained them using UCLA ONH photographs. Different layers of these two models extract features (related to glaucoma symptoms) at different resolutions.

We used 447 ONH photographs from the UCLA data set that included 277 images from healthy eyes and 170 images from eyes with glaucoma. We randomly selected 70% of the images for training, 25% for the validation and 5% for testing.

We also used the HRF data for retesting the model and assessing its generalisability. The region of interest of each image (optic disc) was cropped manually and fed to the models.

7.5 CNN and MNN in Contrast to Transfer Learning

We describe a method aimed at both detecting pathologic changes, characteristic of glaucomatous optic neuropathy in optic disc images, and classification of images into categories glaucomatous/suspect or normal optic discs. Three different deep learning algorithms used are transfer learning, deep convolutional neural network, and deep multilayer neural network that extract features automatically based on clinically relevant optic-disc features.

Deep learning demands extensive data for training. Optimized weights and image features are chosen during learning, and also for evaluation using accuracy metric. Then, the best model is selected. Finally, after training the selected model, the test data set is used to check the quality of the trained model.

The research is aimed at classifying medical images; a high-quality deep

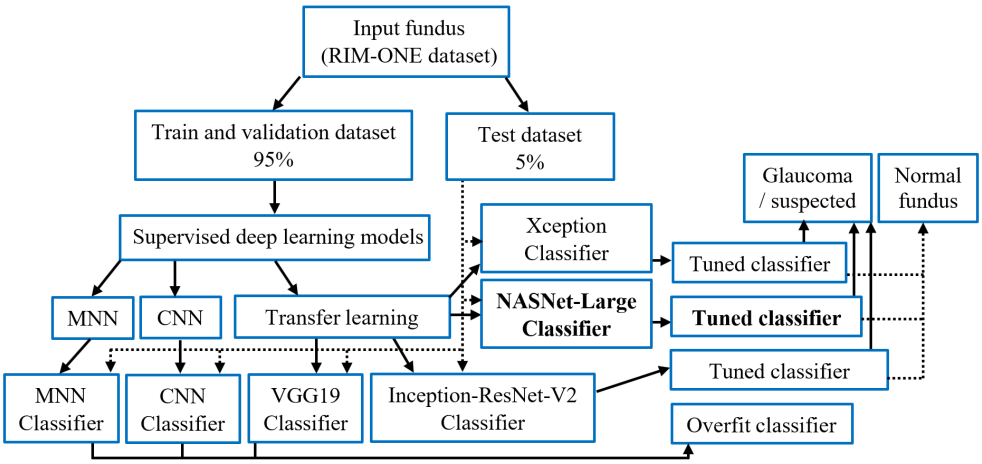


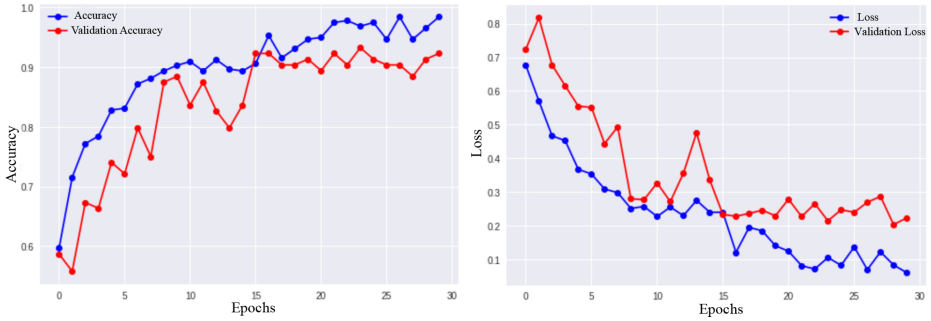
Figure 7.4: Block diagram for glaucoma classification.

learning model based on a limited number of low resolution (about 500×600) images from RIM-ONE V2 has been employed. Due to the small size of this image data set, it is divided into three parts, approximately 75% for the training set, 20% for evaluation, and 5% for testing. In order to overcome the requirement of having a large number of images, we used transfer learning models which are applicable to the weights trained on a large number of images in ImageNet. For the evaluation of the proposed approach, it has been compared with three popular types of deep learning: transfer learning, CNN and MNN. It was found that due to the lack of data, the aforementioned methods did not provide desirable results. By using transfer learning methods and adjusting the hyper-parameters, we was able to achieve improved results. Altogether, in this research we compared three methods for diagnosing glaucoma: CNN, MNN, and (four different) transfer learning methods which are represented as a block diagram in Fig. 7.4. This block diagram includes three supervised learning techniques and a deep learning approach. In the following, we explain briefly these three deep learning methods.

7.6 Evaluation Criteria and Results

Results of Transfer Learning Generalisation

The proposed approach based on InceptionResNet-V2 for generalisation achieved a validation accuracy of 92.3% on a data set of 277 ONH photographs from healthy



(a) Accuracy scores (b) Loss scores

Figure 7.5: Scores of accuracy and loss function for training and validation stages on InceptionResNet-V2 model.

eyes and 170 ONH photographs from eyes with glaucoma. In order to retest the accuracy and generalisability of the proposed approach, we retested the algorithm using an independent data set of 30 ONH photographs. The retest accuracy was 80.0% on average.

This section presents our results when applying the proposed model based on InceptionResNet-V2 model.

Figure 7.5 presents accuracy versus loss in the training and validation stages of the InceptionResNet-V2 model. Figure 7.5a shows how accuracy improves both training and validation with an increase in the number of epochs. Figure 7.5b illustrates how loss decreases for both training and validation with an increase in the number of epochs. Accuracy and loss scores are totally converged into each other. The test and retest results show that this model is working accurately.

Table 7.1: Results of InceptionResNet-V2 model.
"VAL" indicates validation, "ACC" is accuracy, "H" is healthy, and "G" is glaucoma or suspected. Bold data belongs in the best epoch for the proposed method.

	Results on UCLA data set						Results on HRF data set	
Epoch	Train_ loss	Train_ ACC (%)	VAL_ loss	VAL_ ACC (%)	Test_H ACC (%)	Test_G ACC (%)	Retest_H ACC (%)	Retest_G ACC (%)
5	0.40	81.88	0.61	71.15	70	100	86.67	60
10	0.26	90.00	0.31	86.54	80	90	86.67	40
15	0.16	93.13	0.26	92.31	80	100	66.67	40
20	0.12	96.88	0.17	92.37	80	90	93.33	20
25	0.11	96.25	0.28	91.35	70	100	73.33	40
30	**0.06**	**98.44**	**0.22**	**92.31**	**100**	**90**	**93.33**	**66.67**
35	0.05	97.81	0.23	89.42	70	100	53.33	53.33
40	0.03	99.06	0.19	92.31	100	90	93.33	40
45	0.02	99.37	0.16	92.31	90	100	66.67	66.67
50	0.05	98.75	0.07	96.15	90	100	86.67	53.33

We used the pre-trained deep learning model (InceptionResNet-V2 model; see Table 7.1) for generalisation.

Figure 7.5 indicates that scores of InceptionResNet-V2 model converged into each other, and it shows that the record of epochs is between 0 to 30.

Table 7.1 shows more detailed results on training, validation, testing, and retesting for all data sets. We recorded the outcome of the model on epochs five to 50, for every five epochs.

As can be seen, the loss of training decreases consistently with an increase in the number of epochs, except for the 15th epoch in the InceptionResNet-V2 model (Table 7.1).

The best result is achieved on the 30th epoch, in that all healthy images and 90% of glaucoma cases are identified correctly. Moreover, in this epoch, the system detected 93.3% of the healthy eyes and 66.7% of glaucoma eyes correctly.

The results of comparing Transfer Learning with CNN and MNN

The Inception-ResNet [233] is applicable to glaucoma detection, and both validation and training loss scores converged into each other in ten epochs. The accuracy of test data with the Inception-ResNet method on the RIM-ONE data set is 85%. In our study, we required most of the parameters to be trainable and only a small fraction was selected as a default. In order to optimize the training computational complexity, we used a cloud-based *graphics processing unit* (GPU).

The proposed model based on NASNet was trained for five to 50 epochs with elements as mentioned above. The best result was achieved for 50 epochs. The loss fiction result was 0.04, the accuracy value was 98.16%. The result for feeding 20 unseen healthy and glaucoma cases was 90%. The result indicates that our model learned to predict binary classification based on the data set. Some of the results for the test data set are shown in Fig. 7.6.

All results are shown in Table 7.2. "Test ACC" refers to the accuracy of twenty images randomly selected from RIM-ONE, which were unknown for the proposed system. We used a loop with checking thirty different seed random values to choose the highest accuracy with the lowest loss value. Then, we fixed the seed to have a reproducible result to test and retest our proposed model.

When checking the twenty test DFIs, an ophthalmologist was able to detect 70% of them correctly based on the given low-resolution fundus images provided on RIM-ONE as a test set. Expectedly, our proposed method can detect with 18 images correctly out of 20. This method has got 90% accuracy and you can see the results of the proposed method in the highlighted row in Table 7.2.

Of the total of 455 cases extracted from the RIM-ONE public data set (version

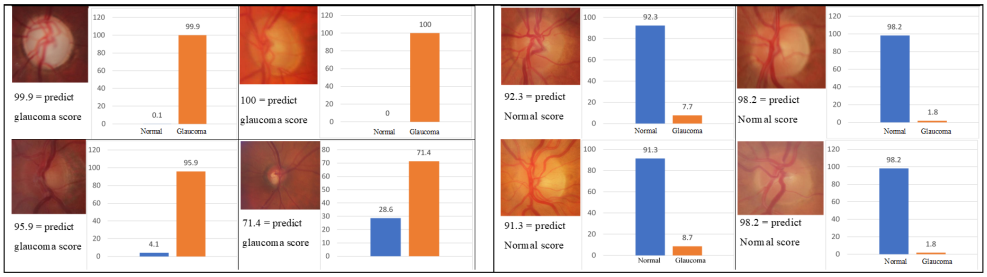


Figure 7.6: Detected glaucoma *(left)* and healthy cases *(right)* on test ODCRs in the RIM-ONE data set with the proposed model based on NASNet.

Table 7.2: Glaucoma detection results for different deep learning methods and "–" indicates that a result is not provided because of over-fitting.

Results in comparison with each other				
Classifier	Best epoch	Train ACC (%)	Train loss	Test ACC (%)
MNN	–	97.25	0.10	50.00
CNN	–	64.31	0.65	50.00
VGG	–	88.77	0.27	50.00
InceptionResNet	10	82.56	0.41	85.00
Xception	20	79.14	0.43	90.00
Proposed model	**50**	**98.16**	**0.04**	**90.00**

2), consisting of 348 training, 87 validation and 20 test cases, the proposed approach classified images with a training accuracy of 98.16%. We hypothesise that this approach can support the clinical decision algorithm in the diagnosis of glaucomatous optic neuropathy.

In this research, InceptionResNet-V2 model has got a much deeper architecture than the VGG model that could generate more accurate outcomes [178, 181].

7.7 Identification of Clinically Relevant Biomarkers

Several deep learning methods are proposed for glaucoma identification from retinal fundus images, however, most of these models provide a black box rather than an interpretable approach. We propose a deep learning approach for glaucoma screening using an interpretable approach to develop a deep learning framework that can automatically identify glaucoma and can identify and localize clinically relevant glaucoma biomarkers on fundus images rather than performing

merely as a black box.

Fundus photographs from 267 healthy eyes and 160 eyes with glaucoma were included. We developed a deep learning model based on the pre-trained NASNet and used a heat map technique to assess parts of the fundus image that were driving the classification, thus allowing localization of clinically relevant objects on retinal fundus images. After training the model, all 427 fundus images were used as preprocessed inputs [67] to the proposed model (based on a deep pre-trained classifier) consisting of the region of interest on fundus photographs. The clinical diagnosis labels of fundus images were validated by a glaucoma specialist and the outcome of deep learning was assessed by experts to assure clinical relevance. We used another independent data set including fundus photographs from 455 eyes for validating the model.

To develop the proposed model; firstly, fundus images were enhanced prior feeding to the deep learning model by a mixed noise removal algorithm based on the super-resolution algorithm and CNN [67]. Then, different augmentations were used such as shift, shear, zoom, and channel shift range, with horizontal flip and nearest fill mode. Deep pre-trained NasNet was used to distinguish fundus images of healthy eyes from fundus images of eyes with glaucoma.

Finally, the heat map technique identifies regions in fundus photographs that are major drivers of glaucoma identification and thence localizes glaucoma signatures. It consists of multiple maps (heat-maps) in different activation and convolution layers.

Note that the deep learning model, a supervised deep pre-trained learning model was applied to the data in the limited fundus images to identify distinct structures literally.

For validation of clinically relevant signatures, we validated the approach using sample fundus images from eyes with glaucoma from the UCLA data set [181]. The clinical diagnosis labels of fundus images were validated by a glaucoma specialist and the outcome of deep learning was assessed by experts to assure clinical relevance.

Results

Moreover, the accuracy of the deep learning model to identify biomarkers in discriminating healthy eyes from eyes with glaucoma was 92%. The validation accuracy on an independent data set of 455 images was 90%. Among fundus images that had been classified to glaucoma group, we observed that deep learning had identified significant features mostly in the superior/inferior peripapillary regions, within the optic nerve head, as well as in their pattern of the large blood

vessel structure.

The heat map model identified the following major signatures for identifying glaucoma:

- Selected boundaries located towards the right side of the optic disc nerve head on fundus images
- The overall optic nerve head structure (Fig. 7.7)
- The overall structure of large blood vessels (Fig. 7.8–Top)
- Neuroretinal rim structures (Fig. 7.8–Bottom)

7.8 Summary

We developed a deep pre-trained model for detection of glaucoma from retinal fundus images using InceptionResNetV2 glaucoma.

We used transfer learning to overcome the over-fitting problem caused by the limited number of input images. We used two independent data sets for training and retesting of the model to assure generalisability of the proposed model.

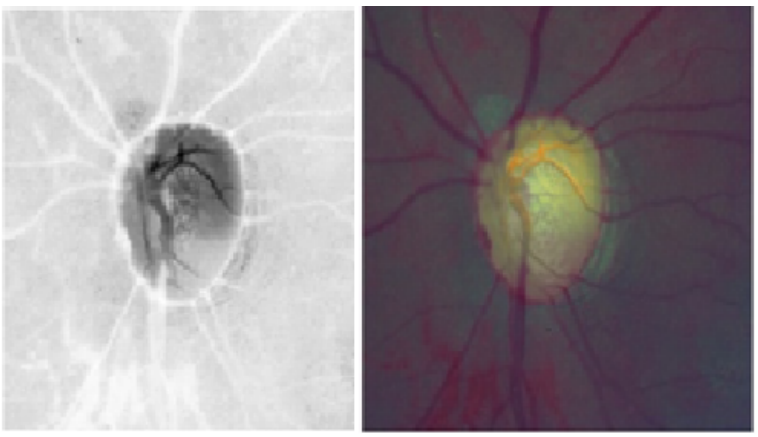


Figure 7.7: The utility of optic nerve head structure region in detecting glaucoma.
Left: Optic nerve head structures identified as a significant feature for the detection of glaucoma based on the proposed deep learning model,
Right: A sample input fundus image overlaid on the optic nerve head structure in the left panel.

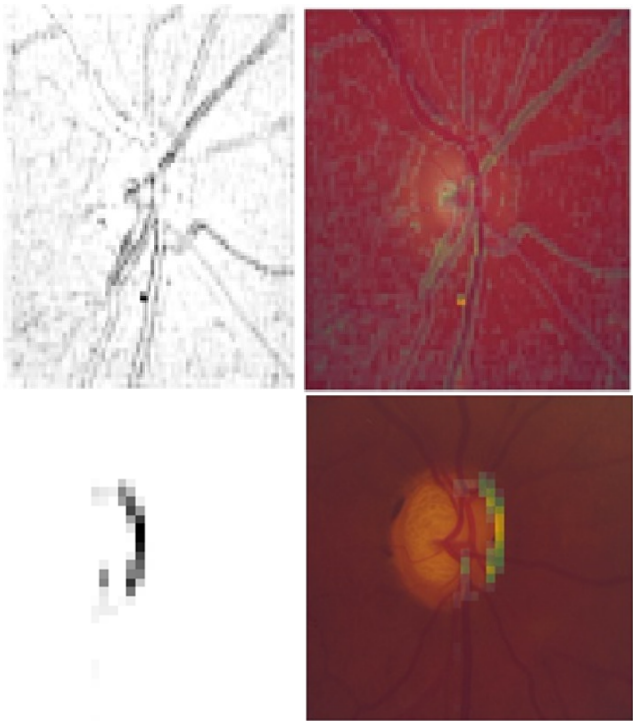


Figure 7.8: Relevant glaucoma biomarkers.
Top: The utility of blood vessel structure feature in detecting glaucoma.
Top left: Blood vessels structure identified as a significant feature for glaucoma detection.
Top right: The input fundus image overlaid on the blood vessel structure in the left panel.
Bottom: Neuroretinal rim was identified as a significant feature for detecting glaucoma.
Bottom left: Neuroretinal rim structures identified by deep learning as a significant feature for detection of glaucoma.
Bottom right: Input fundus image overlaid on the neuroretinal rim shown in the left panel.

InceptionResNet-V2 provides acceptable accuracy for validation, test, and retest data sets.

The average specificity and sensitivity of InceptionResNet-V2 on test and retest data sets were over 100%, 90.1%, 90.9% and 93.3%, respectively.

The proposed framework could be used in clinical and research settings for automated glaucoma diagnosis.

Moreover, in this chapter, we propose a method based on the NASNet model that was trained accurately on 75% of the RIM-ONE V2 data set and gained the least loss score for 50 epochs training. It was evaluated on a test data set which included 20 unseen images. It was successful in classifying glaucoma and suspected glaucoma versus healthy with 90% accuracy.

Totally, six models were trained and compared with each other where three of them (MNN, CNN, and VGG19) had an over-fitting problem. Also, three other models (InceptionResNet, Xception, and proposed method) were deployed to corroborate the performance of the transfer learning models; results were compared to each other. Contributions of this part are as follows:

1. We propose a strategy of using transfer learning techniques for detecting glaucoma. Transfer learning is an opportunity to use and test this method for the detection of healthy versus glaucoma fundus images.

2. Transfer learning models, CNN, and MNN were applied to investigate results in up to one hundred epochs. Moreover, the comparison of results indicates that the proposed method based on NASNet is an eminent method which is applicable to categorize data accurately (with 90% accuracy).

3. In this part, the study was done on a publicly available data set (RIM-ONE). This data set allowed us to evaluate images and extract different features locally or in the cloud. Reliable results are achieved despite the low quality and the small size of fundus images.

Overall, this chapter points toward an appropriate path for glaucoma detection using deep transfer learning.

In future work, we will compare the reported results with a number of classical machine learning and deep learning methods to identify the best performing approach through the multi-modal data. We plan to provide a refined statistical analysis for the performance of the proposed model.

Additionally, fundus photographs can be mined using transfer deep learning techniques to provide critical knowledge about glaucoma onset. Deep learning models based on pre-trained parameters can detect clinically relevant glaucoma signatures from fundus images with high accuracy. Clinically relevant glaucoma signatures can be visualized using the proposed deep learning and heat map framework. The proposed framework provides a step towards more interpretable deep learning approaches rather than black–box approaches. Moreover, we developed the proposed deep learning model based on pre-trained parameters that

were able to detect clinically relevant glaucoma features from fundus images with high accuracy. This approach could be useful in glaucoma clinics as well as in general practice settings as an assistive tool for screening glaucoma in the absence of glaucoma clinicians. Validation of our findings in an independent cohort with larger number of fundus images is required.

Chapter 8

Discussion and Conclusions

This chapter discusses the key contributions and achievements of this thesis and further articulates the research questions, posed in Chapter 1. We also addressed all research questions as stated in Chapter 1. Some main limitations of this work are then discussed along with an overview of future directions.

Discussion

We discuss the research questions and how we got the solution that is approved by ophthalmologists . The research questions of this study are listed as follows:

> How can we detect changes on retinas in the rim area of ONH with stereo and monocular images?
>
> Stereo optic-disc imaging improves intra and inter-observer agreement in the detection of optic disc abnormalities. We proposed a robust method to help the specialist in detecting some abnormalities in stereo optic-disc images using stereo vision and superpixel segmentation concepts. A stereo vision system produces a disparity map for the input stereo images of the retina in which abnormalities are more distinguishable. In the region of interest in the disparity map, we can clearly visually recognize all abnormalities because of changes in distances for abnormality parts in glaucoma patients. The produced disparity map with the SP1 system shows and magnifies the changes in the rim section. It is also segmented using two different superpixel segmentation algorithms (simple linear-iterative clustering and simple non-iterative clustering) to detect abnormalities. The original stereo images are also segmented using the same concept; results are compared with the segmented disparity map.
>
> Moreover, measuring the cup-to-disc ratio (CDR) is a common approach for glaucoma detection. Glaucoma can be specified by thinning of the rim area

that identifies the CDR value. Clustering and image segmentation can simply divide fundus images into distinct areas to estimate the optic disc (OD) and the optic cup (OC). This paper is based on a robust method, using the improved chaotic imperialistic competition algorithm (ICICA) for determining the position of the OD and OC on colour fundus images for glaucoma detection. The predicted OD and OC boundaries are then used to estimate the CDR for glaucoma diagnosis. The performance of the proposed method was evaluated by using the publicly available RIGA dataset. It was found that some of the common problems of the K-means clustering algorithm can be addressed by the proposed method for achieving better results. Moreover, the OC and OD regions can be precisely separated from the colour image so that ophthalmologists can measure OC and OD areas more accurately.

How can a deep learning model be optimized for glaucoma detection and cope with the lack of fundus data to have a high-performance?

We developed a deep learning algorithm for identifying glaucoma on optic nerve head (ONH) photographs. We applied transfer learning to overcome over-fitting on the small training sample size that we employed. The transfer learning framework that was previously trained on large datasets such as ImageNet, uses the initial parameters and makes the approach applicable to small sample sizes. We then classified the input ONH photographs as "healthy" or "glaucoma". The proposed approach achieved a validation accuracy of 92.3% on a dataset of 277 ONH photographs from normal eyes and 170 ONH photographs from eyes with glaucoma. In order to retest the accuracy and generalisability of the proposed approach, we retested the algorithm using an independent dataset of 30 ONH photographs. The retest accuracy was 80.0% on average.

Additionally, we described a method aimed at both detecting pathologic changes, characteristic of glaucomatous optic neuropathy in optic disc images, and classification of images into categories glaucomatous/suspect or normal optic discs. Three different deep-learning algorithms used are transfer learning, deep convolutional neural network, and deep multilayer neural network that extract features automatically based on clinically relevant optic-disc features. Of the total of 455 cases extracted from the RIM-ONE public dataset (version 2), consisting of 348 training, 87 validation and 20 test cases, the proposed approach classified images with a training accuracy of 98.16%. We hypothesise that this approach can support the clinical decision algorithm in the diagnosis of glaucomatous optic neuropathy.

How can we develop a deep learning model to identify the biomarkers for glaucoma detection and find a method for glaucoma screening?

To develop a deep learning framework that can automatically identify clinically relevant features on glaucoma fundus images. The accuracy of the method in discriminating normal eyes from eyes with glaucoma was 92%. The validation accuracy on an independent dataset of 455 images was 90%. Among fundus images that had been classified to the glaucoma group, we observed that deep learning had identified significant features mostly in the superior/inferior peripapillary regions, within the optic nerve head, as well as in their pattern of large blood vessel structure. We developed a deep learning model based on pre-trained parameters that was able to detect clinically relevant glaucoma features from fundus images with high accuracy. This approach could be useful in glaucoma clinics as well as in general practice settings as an assistive tool for screening glaucoma in the absence of glaucoma clinicians. Validation of our findings in an independent cohort with a larger number of fundus images is required.

How can reliability and performance rate be optimized for a model with a thickness of retina data and apply RNFL for glaucoma screening?

We aimed at evaluating the relationship between RNFL thickness and glaucoma patients. Thus, we developed a fused pattern recognition model with 26 classifiers to detect healthy vs. glaucoma patients. We also achieved an F1 score of 0.82 and an accuracy of 82% using 5-fold cross-validation on a data set of 107 RNFL data from healthy eyes and 68 RNFL data from eyes with glaucoma; 25% of data have been selected randomly for testing. The proposed fused the model based on a stack of top ten supervised classifiers combined by an ensemble learning method to achieve a robust and generalised model for glaucoma detection in the early stages. Additionally, we implemented an unsupervised model based on the K-means clustering with 80% accuracy for glaucoma screening. In this research, we have followed two purposes: first, to assist the ophthalmologists in their daily patient examination to confirm their diagnosis, thereby increasing the accuracy of diagnosis. The second usage is glaucoma screening by optometrists in order to perform more eye tests and better glaucoma diagnosis. Therefore, our experimental tests illustrate that having only one data set still allows us to obtain highly accurate results by applying both supervised and unsupervised models. In future, the developed model will be retested on more substantial and diverse data sets.

How can we use the eye-tracking to detect glaucoma?

We developed a model to detect a particular movement of the pupil (OKN) that may be helpful to detect glaucoma in the early stages. *Optokinetic nystagmus* (OKN) is an involuntary repeated "beating" of the eye, comprised of sequences of slow tracking (slow phase) and subsequent quick re-fixation events (quick phase) that occur in response to (typically horizontally) drifting stimuli. OKN has a characteristic saw-tooth pattern that we detected here using a state-machine approach applied to the eye-tracking signal. Our algorithm transitions through the slow/quick phases of nystagmus (and a final state) in order to register the start, peak and endpoints of individual sawtooth events. The method generates duration, amplitude, and velocity estimates for candidate events, as well as repetition estimates from the signal. We tested the method on a small group of participants. The results suggest that false positive detections occur as single isolated events in feature space. As a result of this observation we applied simple criteria based on the repetitious "beating" of the eye. The number of true positives is high (94%) and false OKN detections are low (2%). We will aim to optimize and rigorously validate the proof-of-concept framework we propose.

We summarized selected latest research on glaucoma detection with our results which were shown in Table 2.2 and 2.3 in chapter2. Also, our study focused on developing a glaucoma detection and biomarkers visualisation model. Moreover, pattern recognition based on OKN detection with signal processing as a hypothesis may assist researchers in the machine learning area for glaucoma detection that is discussed in Appendix A. Thus, we were performing a software application to detect OKN with a specific pattern on eye movement that can show us any changes in VF data and quality of vision.

Summary of Remarks

Our goals of the PhD thesis have been achieved with novel research work [67, 175, 177–182] utilising deep learning, transfer learning, hybrid model and stereo matching to detect glaucoma at its early stages. These models exist in the literature and targeting glaucoma detection with high performance and low cost [135, 225]. Thus, we applied many traditional and deep machine learning algorithms to achieve our goals.

The proposed computer-aid application which was implemented in Python, Matlab and C/C++ using available GPU hardware at AUT (CeRV lab and Colab) through the Anaconda and Spyder, Jupyter and Pycharm IDEs *integrated development environment* to simplify deep learning design and reduce development

effort and time. The implemented system is considered the first transfer learning that exists at present.

The primary research contributions of the thesis can be summarised as follows:

- Describing publicly available and labelled glaucoma fundus image databases, highlighting different performance evaluation measures commonly used within the field, and outlining current approaches in feature-based fundus image classification.
- Not only providing a better understanding of deep CNN architectures but also facilitating future research activities and application developments in the field of traditional classifiers and transfer learning for glaucoma detection.
- Proposing the ICICA clustering algorithm for segmentation to avoid the common problems of K-means clustering algorithms, such as dependence on initial values and early convergence that cause poor results.
- Offering rim area segmentation on fundus images with stereo matching and disparity map calculations to support the early detection of (small) abnormalities.
- Proposing the classification of healthy versus glaucoma patients through RNFL data in two private data sets with high accuracy.
- Data visualization of the complex RNFL data with statistical analysis such as PCA and Z-score.
- Robustness improvement through results combination of traditional classifiers to generate the hybrid.
- Using transfer learning with regularization to overcome the over-fitting problem caused by the limited number of input images.
- Proposing different deep learning models that were trained and compared with each other to get the best accuracy.

Limitations and Future Directions

Besides the remarks that have been made hitherto, there are a number of avenues that could be further improved and extended toward a number of directions in the future as follows:

- Lack of data in particular multi-modal data for glaucoma detection. Thus, the RNFL trained models can only be valid on the defined data scope and this is not yet generalised to be tested on any new person's data stream.
- Develop a data type conversion model to transform the thickness of retina data to the images and deploy the deep convocational neural network to detect glaucoma.
- Detection any changes on the rim area with deep learning on segmented fundus data sets annotated by ophthalmologists.
- Using the pupil tracking technique to detect the number of healthy ganglion cells.
- The developed hybrid model will be retested on more substantial and diverse data sets to be trustworthy for use in medical tests.
- We will check the proposed model based on time series or longitudinal of retina's thickness.
- Utilising TensorFlow Quantum [42] to optimize our proposed model while using the longitudinal and big data.

Longitudinal is a sort of time series data which can aid ophthalmologists to glaucoma prognosis at its early stages. Time-series data typically arises from the collection of many data points over time from a single source, such as from "glaucoma suspected patient". Longitudinal data typically arise from collecting a few observations over time from many sources. For example, collecting a few blood pressure measurements from "many people".

Finally, one of the keys to future research directions will be using real-time pupil tracking techniques to detect glaucoma based on the quality of vision through pattern recognition algorithms. The quality of vision including different patterns may cause identification of the number of healthy ganglion cells on the retina for glaucoma detection.

Appendix A

Real-time Pupil Tracking for Detecting Optokinetic Nystagmus

Optokinetic nystagmus (OKN) is an involuntary repeated "beating" of the eye, comprised of sequences of slow tracking (slow phase) and subsequent quick re-fixation events (quick phase) that occur in response to (typically horizontally) drifting stimuli. OKN has a characteristic saw-tooth pattern that we detect here using a state-machine approach applied to the eye-tracking signal. Our algorithm transitions through the slow/quick phases of nystagmus (and a final state) in order to register the start, peak and endpoints of individual sawtooth events. The method generates duration, amplitude, velocity estimates for candidate events, as well as repetition estimates from the signal.

We test the method on a small group of participants. The results suggest that false positive detections occur as single isolated events in feature space. As a result of this observation, we apply simple criteria based on the repetitious "beating" of the eye. The number of true positives is high (94%) and false OKN detections are low (2%). Future work will aim to optimize and rigorously validate the proof-of-concept framework we propose.

In the previous chapters, we applied different data modalities to detect glaucoma that is valuable and well-noun for glaucoma detection. In this chapter, we promote a new trend of glaucoma detection by using the pattern of pupil movements. Therefore, there is a potential to have a glaucoma detection in the early stages via pupil tracking based on machine learning techniques with specific patterns [23, 24, 118, 148, 189]. Material discussed in this chapter has been published in my publication [176].

A.1 Optokinetic Nystagmus

Optokinetic nystagmus (OKN) is an involuntary movement of the eye that occurs as a person views a drifting stimulus. It is characterised by a slow tracking (the slow phase) and a subsequent resetting motion in the opposite direction (the quick phase) that allows the eye to fixate and track a different stimulus feature. The overall visual appearance of OKN, typically elicited by drifting vertical bars, is a

repeated "beating" of the eye, that appears as sawtooth eye movements in the horizontal displacement signal (see Fig. A.1).

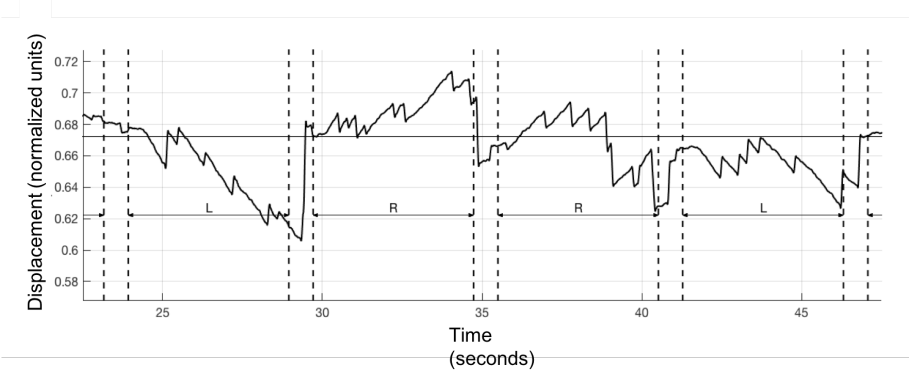


Figure A.1: An example of the horizontal displacement signal (coordinates normalised to the eye camera image) showing OKN. Also indicated on the figure is the direction of the stimulus (*i.e.*, in this case, L = leftward, R = rightward, synchronised to eye data.

OKN is an established means to detecting deficits along the visual pathway [68, 245,246]. The literature suggests [8,49,104,116,206] the presence/absence of OKN (in response to carefully designed stimulus) can be used to detect clinically significant deficits in *visual acuity* (VA), the (self-reported) ability of the eye to see fine detail. Visual acuity can be difficult to obtain in non-verbal patients, such as young children [13] and the involuntary nature of OKN presents a potential method for rapidly and accurately assessing VA in these patients.

The automated detection of the slow/quick phases of the optokinetic and related vestibulo-ocular reflexes have been studied by a number of authors. Velocity threshold was used to determine the saccadic portion of the signal [128, 188]. We found threshold of the velocity signal to be effective in an off-line situation in which a consumer grade camera was used to record video of the eye performing OKN in adult participants [243] as well as children [212]. Alternate approaches include a recursive digital filter that responds to the changes of phase of the signal obtained using *electronystagmography* (ENG) [128], and an system identification approach utilising ARX model for identifying the relationship between fast phase and slow phase velocity [207]. Recently, Ranjabaran et al. demonstrated a fully objective approach based on K-means clustering to provide an initial classification of data as belonging to fast/slow phase or non-slow phase followed by a system identification approach using ENG [202].

The major focus of this chapter is to provide proof-of-concept for a simple approach to OKN detection; highly suited for real-time application. Our method takes as input the horizontal eye displacement obtained from a head-mounted eye tracking system. The method generates time points corresponding to the onset/peak and end of triangular "sawtooth" features characteristic of OKN from the incoming signal. The resulting feature vectors are filtered using heuristic rules to determine whether they are legitimate candidates for OKN. From this process, we find that the repetitive nature of OKN appears to be a discriminative factor for the presence/absence of OKN. Our overall finding is a high true positive rate and low false detection rate using this as a discriminating criteria.

A.2 Background for Detecting OKN

We denote the input signal vector by $x(t) = (d(t), v(t), \gamma(t))$; the concatenation of the (horizontal) displacement of the eye $d(t)$, the velocity $v(t)$ and auxiliary information, $\gamma(t)$ (consisting of the direction of the stimulus and data quality). Our aim is to determine the start x_s, peak x_p and end points x_e of triangular features from $x(t)$, and to test the resulting feature vector (x_s, x_p, x_e) using reasonable decision criteria, to be described, to eliminate unlikely sawtooth candidates.

Consider a sample of $d(t)$ containing OKN as shown in Fig. A.2. The onset of a sawtooth is given by the point $x_s = (t_s, d_s)$, the point where a rising displacement in the eye signal is first detected, the peak of the sawtooth by $x_p = (t_p, d_p)$ which occurs as the eye transitions to the quick resetting eye motion, and the end of the sawtooth by $x_e = (t_e, d_e)$ where the descending edge now transitions to rising or stationary. These points yield the (average) slow/quick phase velocities, v_{SP} and v_{QP}:

$$v_{SP} = \frac{\triangle d_{SP}}{\triangle t_{SP}} = \frac{d_p - d_s}{t_p - t_s} \tag{A.1}$$

$$v_{QP} = \frac{\triangle d_{QP}}{\triangle t_{QP}} = \frac{d_e - d_p}{t_e - t_p} \tag{A.2}$$

where $\triangle t_{SP}$ and $\triangle t_{QP}$ are the slow/quick phase durations, and $\triangle d_{SP}$ and $\triangle d_{QP}$ are slow/quick phase amplitudes. Figure A.2 illustrates basic consistency constraints summarised by Table A.1(a). The durations ($\triangle t_{SP/QP}$) should be positive and non-zero: the quick phase duration must be shorter than the slow phase duration. The slow/quick phase velocities ($v_{SP/QP}$ must be non-zero and of opposite sign. The magnitude of the quick phase velocity should not exceed that of the slow phase. Furthermore, additional empirically based thresholds were applied

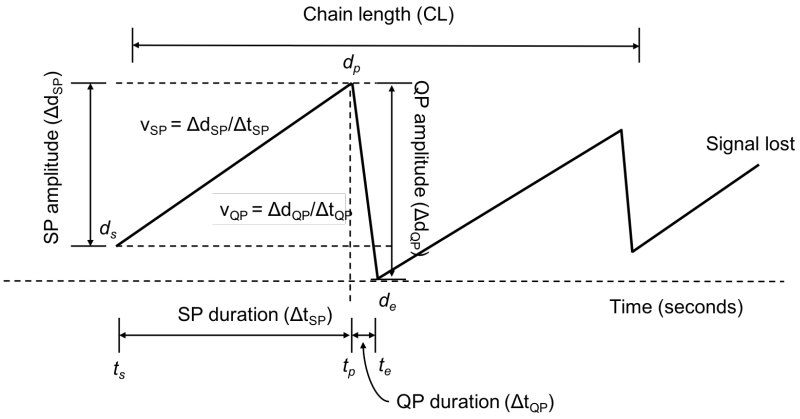


Figure A.2: Components of the OKN displacement signal

(a) Consistency constraints

Number	Constraint
1	$0 < \triangle t_{QP} \leq \triangle t_{SP})$
2	$v_{SP} \cdot v_{QP} < 0$
3	$\lvert v_{SP} \rvert < \lvert v_{QP} \rvert$

(b) Thresholded constraints

Number	Description	Variable	Lower limit	Upper limit
4	QP duration	$\triangle t_{QP}$	0.1	2.0
5	SP velocity[a]	$\lvert v_{SP} \rvert$	0.05	0.40
6	QP/SP amplitude	$\lvert \triangle d_{QP/SP} \rvert$	0.004	0.2

[a] Assumes the slow-phase velocity is the same as the stimulus direction

Table A.1: Constraints applied to the detected sawtooth features

as summarised by Table A.1(b). This thresholding was used to eliminate potential OKN candidates based on lower and upper estimates for quick/slow phase amplitude/duration and speed.

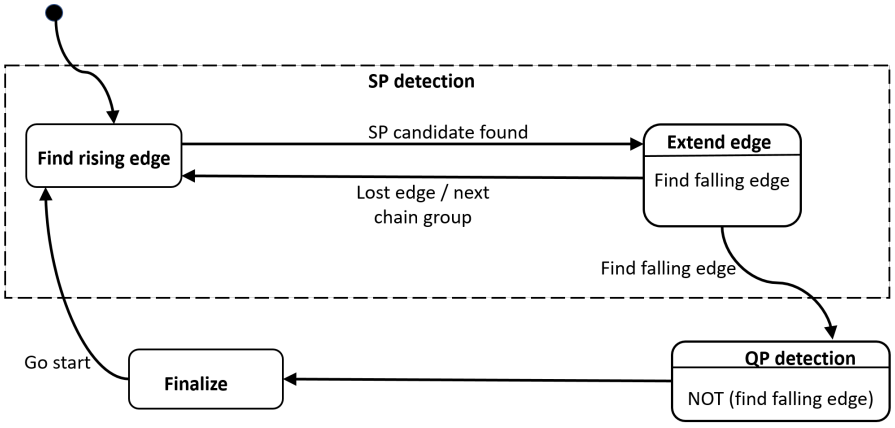


Figure A.3: State description of the algorithm.

State-machine description of algorithm

The optokinetic response is generated physiologically by independent slow phase and quick phase systems [250]. This behaviour naturally suggests a state-machine solution to detection; which is also highly suited for real-time implementation. Our algorithm is now explained. In broad terms, the input data stream $x(t)$ (dataset was presented in chapter 3) drives our machine through three states (either slow phase detection, quick phase detection or finalise as shown in Fig. A.3)

The purpose of the slow phase detection state is to find a rising edge of sufficient duration; if this is found then the start point of the rising edge is registered as the beginning of a sawtooth (t_s, d_s) and the machine transitions to the extend edge state. Whilst in the extension phase, the machine now looks for a falling edge to indicate the end of the slow phase, whereupon the end of SP/start of QP (t_p, d_p) will be registered. Any failure (for example, due to lost data caused by blinking) will discard the candidate and cause a reset to the find rising edge state (and also increment a chain group counter to be explained presently). In any event, given a successful detection for the start of a QP, the machine will transition to the QP phase. The machine now looks for a non-decreasing edge indicating the end of the QP (t_e, d_e).

Once the end is found, and given the points of the sawtooth, the machine transitions to the finalize state: the purpose of which is to retain OKN like sawtooth features and discard candidates unlikely to be due to OKN. We do this by applying the criteria described in Table A.1.

We measured also contiguous groupings of sawtooth features, using a chain

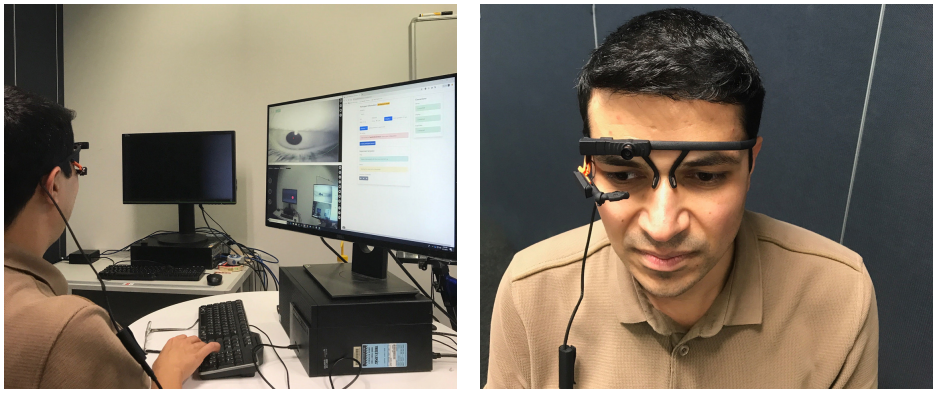


Figure A.4: *Left*: Experimental setup. *Right*: Used head-mounted eye tracking system [305].

group counter. This counter value was carried across to each new sawtooth candidate. In the case of a failure (*i.e.*, reset event) the group counter advanced thereby indicating a new group. This labelled each repeated sawtooth as belonging to a particular group, and the number of features belonging to a group gave the chain length as shown by variable CL in Fig. A.2. In this example, the chain length is $CL = 2$ the number of complete sawtooth features detected.

Our aim in this work was to see whether there were readily discernible patterns in the measured parameters that would allow us to determine whether stimulus was present or absent. We were particularly interested in whether slow phase properties (velocity and duration) or chain-length could be of value in differentiating between OKN present or absent cases.

A.3 Experimental Methods

The experimental setup is shown in Fig. A.4. Participants were asked to stare at the centre of the screen with one eye covered (the protocol was repeated for each eye). The viewing distance was 1.5m. The eye displacement was recorded using a 120 fps, head mounted eye-tracker (Pupil-Labs, Berlin, Germany). The raw horizontal displacement was smoothed using a Savitsky-Golay (SG) filter. The SG filter preserves high frequency content, and is conveniently specified in the time-domain by polynomial order N and frame-length M [196]. Moreover, the velocity $v(t)$ is readily computed from these filter coefficients. In this work, the order was $M = 2$, the region frame-length was chosen as $f = 13$.

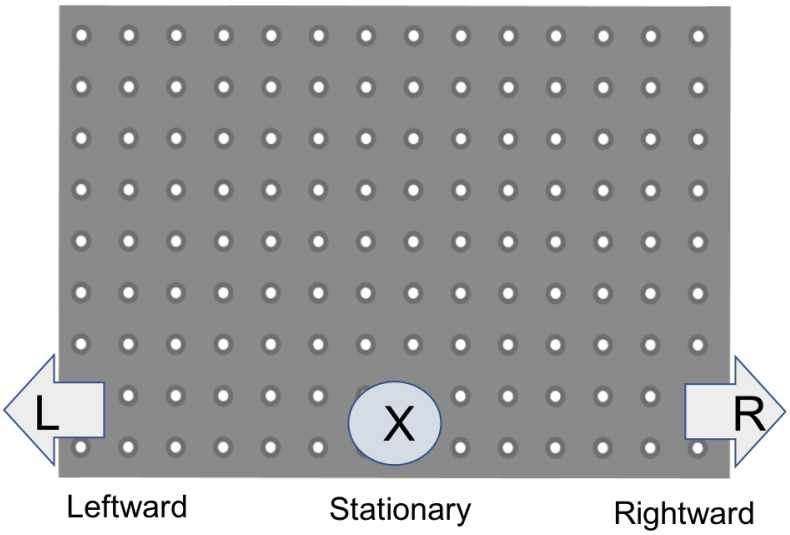


Figure A.5: The stimulus array shown to participants. The array was shown drifting leftward/rightward or stationary.

The algorithm was run with $(d(t), v(t), \gamma(t))$ data (across all participants and for each eye). The auxiliary function used in this work was $\gamma(t) = (\gamma_1(t), \gamma_2(t))$ comprising the known state of the stimulus ($\gamma_1(t) \in \{L, R, X, F\}$), and data quality measure ($0 \leq \gamma_2(t) \leq 1$) from the eye-tracker device. The latter measure was used to determine whether the data was of sufficient quality to be useful (maximum quality was 1 and minimum quality was 0). The former quantity passed the *a-priori* direction of the stimulus to the algorithm. Here, the potential values for $\gamma_1(t)$ were L = "leftward", R = "rightward", X = "stationary", or F = "fixation") (see Fig. A.5).

The slow phase velocity v_{SP}, duration ($\triangle t_{SP}$) and the chain length (CL) were extracted and categorized as depending on whether they were obtained whilst the stimulus was moving (*i.e.*, trials labelled "L" or "R") or whilst the stimulus was stationary (*i.e.*, trials labelled "X"). The fixation trials labelled "F" were ignored. For trials in which the stimulus array was stationary, the algorithm was run twice. This avoided choosing a particular direction for these trials; and allowed an unbiased estimate of the false positive (FP) detection rate of the method.

The apparent direction of travel of the stimulus as perceived by the observer was recorded, for comparison purposes, by asking the participant to press keys to indicate the direction of travel as they watched the stimulus on-screen. This was

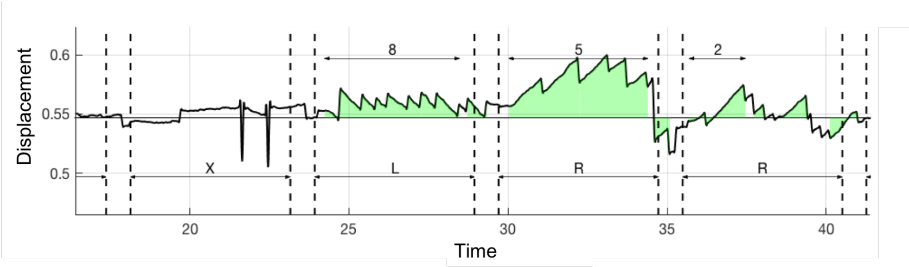


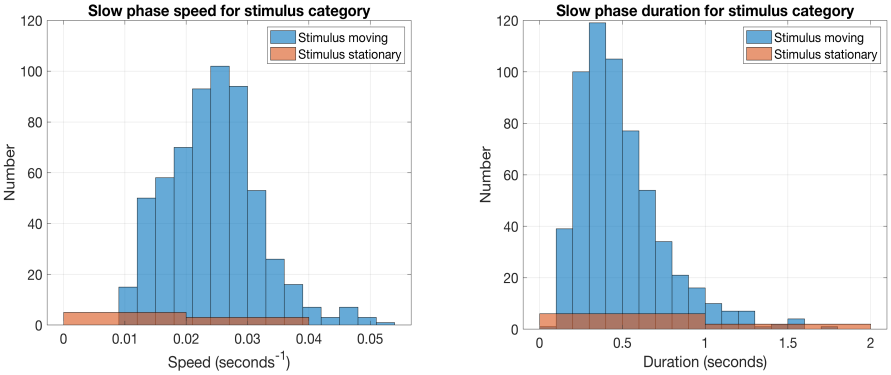
Figure A.6: The result of an example of a horizontal displacement signal.

used to confirm that the stimulus was perceived as expected by the observer. The eye tracking, experiment management and key-press collection where performed on a second computer running a custom web-server application (`node.js`) with an JavaScript/HTML5 interface. The display of the stimulus, ran on the primary display by accessing the server using the chrome web-browser. Figure A.4 shows the experimental setup used.

A.4 Results

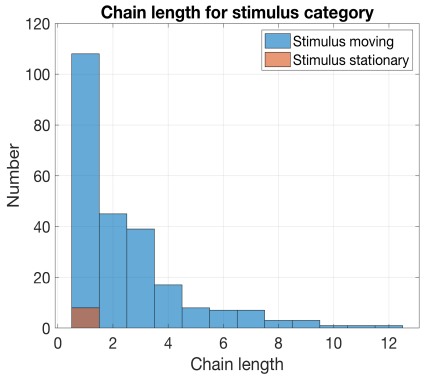
The stimulus was seen by all observers (n=6) for all trials ($N = 11$ eyes) as indicated by correct key-presses measured during the trial. Data for one eye was excluded because stimulus direction information $\gamma_1(t)$ was lost. The detection algorithm completed successfully for all other test runs/participants. The output from the system is given by Fig. A.6. The figure shows unshaded areas corresponding to intervals of time not identified as containing OKN, compared with shaded green areas where OKN was detected. The labels under the graph indicate the direction of travel as a function of time $\gamma_1(t)$, the numbered intervals above the graph show the chain length computed for the data (chain lengths of 1 are not indicated). Visual inspection (by the authors, who are experienced in identifying OKN) suggested the algorithm was effective in identifying regions containing OKN.

A total of $n = 661$ sawtooth features were detected for the moving stimulus category. A total of $n = 41$ sawtooth features were detected in the the stationary category. The FP rate at the level of features detected was 5.8% of all sawtooth detections. Figure A.7a shows the distribution of results obtained for the slow phase speed $|v_{SP}|$. The mean slow phase velocity (mean $\pm$ 2SD) were 0.024 ± 0.014 and 0.017 ± 0.016 (s^{-1}) for the moving ($n = 661$) and stationary categories ($n = 41$) respectively. Figure A.7b shows the distribution results obtained for the slow phase



(a) Slow phase velocities.

(b) Slow phase durations.



(c) Chain length numbers.

Figure A.7: Eye displacement signal for the first sample.

duration Δt_{SP}. The SP duration was 0.48 ± 0.52 s and 0.33 ± 0.50 s for the moving and stationary categories. The CL (*chain length*) is shown in Fig. A.7c. In this instance the mean and standard deviations were 3.10 ± 4.44 and 1.03 ± 0.30 for moving ($n = 215$ chains) and stationary categories ($n = 40$ chains). A two-sample t-test rejected the null hypothesis (no difference between moving and still distributions) at 5% significance level for the three parameters.

Figure A.8 summarizes all three features (SP velocity, SP duration and chain length CL) plotted on a single graph. The $n = 41$ false sawtooth detections appear as orange hued circles, compared to true detections ($n = 664$) shown in blue. Visual inspection of the data (*e.g.*, Fig. A.8) indicated that these false detections, were shifted toward lower durations and speeds. Most visually significant was the

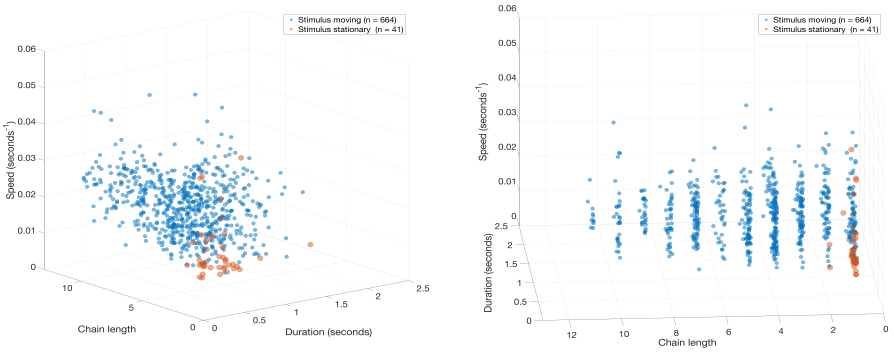


Figure A.8: Illustration of the distribution of three features from two views

CL threshold	TP rate	FP rate
1	109 (99%)	25 (46%)
2	102 (94%)	1 (2%)
3	89 (82%)	0 (0%)

Table A.2: Per trial TP and FP rates as a function of the CL parameter.

observation that FP detections were clustered along the $CL = 1$ and $CL = 2$ planes; indicating that CL could be a discriminating factor for OKN present/absent. The performance of CL as an indicator of OKN present/absent is shown in Table A.2. This table shows TP as a proportion of the total number of moving trials ($n = 110$) and FP as a proportion of total stationary trials ($n = 55$) shown to the observer. This table shows the reduction in FP rate for increasing CL as well as a drop in TP rate. The table suggests that a threshold of $CL = 2$ was the best balance between TP and FP for trial-by-trial detection of OKN. As a consequence of this result, the effects of SP duration and speed were not considered further, but would be the subject of future work.

A.5 Discussion

There is a clinical need for automated approaches able to identify the presence or absence of optokinetic nystagmus, particularly suited to real-time application. In this work we developed a method suitable for real-time detection of optokinetic

nystagmus based on a simple state-machine approach. The algorithm was developed and run on a small cohort of adult participants, who watched drifting or stationary patterns whilst having their eye movements recorded.

Sawtooth patterns were detected readily in moving trials ($n = 664$ detections) but also during some stationary trials ($n = 41$ detections) (a per feature FP detection rate of 5.8%). The effects of false detections were eliminated by considering the *per trial* criteria that OKN should be repetitive. For example, we found that a criteria of $CL >= 2$ would identify trials with moving stimulus ($TPratewas94\%$) whilst eliminating false detections ($FPratewas2\%$). We suggest that chain-length may be a key factor in assessing the presence or absence of OKN.

It is intended that the performance be evaluated more carefully in the future. We found that the mean slow-phase speed and durations for moving and stationary conditions were different, but we did not analyse this finding further. In this work, we presented data for the calibration set only, and we did not perform a robust validation analysis. We need to perform a full ROC analysis of the method which will the subject of further work. It is emphasised that the aim here was to present the basic concept, which was the use of a state-machine approach to determine and analyse OKN.

In this work we utilised a web-browser to display our stimulus. This was facilitated by the jsPsych package (`www.jsPsych.org`), a web browser based API for psychophysical trials. The jsPsych package utilises a plug-in architecture that allows the user to perform pre-programmed tasks (*e.g.*, show a movie, play audio, record a reaction) or custom tasks that execute in a sequence defined by the experimenter. We wrote a custom plugin was written that facilitated the display of the disk stimulus for the purpose of web browser display.

The web-browser display was controlled from a second computer (the controller) that managed the experiment. Crucially the system synchronised the start and end of each trial of the experiment with the pupil labs eye tracker. In this work the server code was written using `node.js` and the interface to the server was written JavaScript/HTML5 thereby maintaining a non-platform specific implementation with the possibility of distributing more widely in future work. Furthermore, we developed a batch extraction of pupil location which can extract the location of the pupil for many participants simultaneously. This model can extract the (x, y) of pupil based on 2-D frame on the video stream. In the future, we are going to have an autonomous pupil detection based on cloud computing platform.

There are a number of limitations of the present study. It would be desirable to increase the number of participants tested which would allow a more detailed examination of the behaviour of participants during eye testing and to allow

further generalisation (if possible) of the threshold we have used already. As mentioned, we require to perform a more in-depth sensitivity-specificity analysis. We looked at a limited set of parameters in this work (essentially chain-length and slow phase duration/speeds), and a more in-depth analysis would be required to indicate optimal features for quantification of OKN. Our aim is to provide these methods for clinical use, and therefore future studies will quantify performance on target groups such as children. Future work will now look to determine whether the present protocols and processing approaches can be improved, and work is under-way to examine whether machine learning approaches will benefit the technique we have developed.

A.6 Summary

We have presented a method for detecting optokinetic nystagmus designed for real-time applications. We have obtained encouraging results for a cohort of adult participants ($n = 11$ eyes). Further research is warranted, and we will continue to improve upon and further validate the methods presented here. In a forthcoming publication, we will use pupil tracking model to detect glaucoma through signal processing and machine learning techniques.

Bibliography

[1] S.J. Pan, and Q. Yang, "A survey on transfer learning", IEEE Transactions on knowledge and data engineering, 22(10):1345–1359, 2010.

[2] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems", J. CoRR, arXiv:1603.04467, 2016.

[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods", IEEE Trans. Pattern Analysis Machine Intelligence, 34(11):2274–2282, 2012.

[4] R. Achanta and S. Süsstrunk, "Superpixels and Polygons Using Simple Non-iterative Clustering", IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 4895–4904, 2017.

[5] R. Acharya, E. K. Ng, and J. S. Suri, "Image Modeling of the Human Eye", Artech House, Boston, 2008.

[6] B. Al-Bander, W. Al-Nuaimy, M.A. Al-Taee, and Y. Zheng, "Automated glaucoma diagnosis using deep learning approach", in: IEEE Int. Conf. Systems Signals Devices, pp. 207–210, 2017. https://doi.org/10.1109/SSD.2017.8166974

[7] B. Al-Bander, B. M. Williams, W. Al-Nuaimy, M. A. Al-Taee, H. Pratt, and Y. Zheng, "Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis", Symmetry, 10(4):87, 2018. https://doi.org/10.3390/sym10040087

[8] C. Aleci, M. Scaparrotti, S. Fulgori and L. Canavese, "A novel and cheap method to correlate subjective and objective visual acuity by using the optokinetic response", Int Ophthalmol, 1–15, 2017. https://doi.org/10.1007/s10792-017-0709-x

[9] A. Almazroa, "A novel Automatic Optic Disc and Cup Image Segmentation System for Diagnosing Glaucoma using RIGA dataset", Waterloo University, 2016.

[10] A. Almazroa, R. Burman, K. Raahemifar, and V. Lakshminarayanan, "Optic disc and optic cup segmentation methodologies for glaucoma image detection: A survey", J. Ophthalmology, 42: 162–189, 2015.

[11] R. Al-Rfou, G. Alain, A. Almahairi, C. Angermüller, D. Bahdanau, N. Ballas, F. Bastien, J. Bayer, A. Belikov, A. Belopolsky, Y. Bengio, A. Bergeron, J. Bergstra, V. Bisson, J. B. Snyder, N. Bouchard, N. Boulanger-Lewandowski, X. Bouthillier, A. de Brébisson, O. Breuleux, P. Luc Carrier, K. Cho, J. Chorowski, P. F. Christiano, T. Cooijmans, M. CÔté, M. CÔté, A. C. Courville, Y. N. Dauphin, O. Delalleau, J. Demouth, G. Desjardins, S. Dieleman, L. Dinh, M. Ducoffe, V. Dumoulin, S. E. Kahou, D. Erhan, Z. Fan, O. Firat, M. Germain, X. Glorot, I. J. Goodfellow, M. Graham, C. Gülçehre, P. Hamel, I. Harlouchet, J. Heng, B. Hidasi, S. Honari, A. Jain, S. Jean, K. Jia, M. Korobov, V. Kulkarni, A. Lamb, P. Lamblin, E. Larsen, C. Laurent, S. Lee, S. Lefrançois, S. Lemieux, N. Léonard, Z. Lin, J. A. Livezey, C. Lorenz, J. Lowin, Q. Ma, P. Manzagol, O. Mastropietro, R. McGibbon, R. Memisevic, B. van Merriënboer, V. Michalski, M. Mirza, A. Orlandi, C. Pal, R. Pascanu, M. Pezeshki, C. Raffel, D. Renshaw, M. Rocklin, A. Romero, M. Roth, P. Sadowski, J. Salvatier, F. Savard, J. Schlüter, J. Schulman, G. Schwartz, I. V. Serban, D. Serdyuk, S. Shabanian, É. Simon, S. Spieckermann, S. R. Subramanyam, J. Sygnowski, J. Tanguay, G. van Tulder, J. P. Turian, S. Urban, P. Vincent, F. Visin, H. de Vries, D. Warde-Farley, D. J. Webb, M. Willson, K. Xu, L. Xue, L. Yao, S. Zhang, and Y. Zhang. "Theano: A Python framework for fast computation of mathematical expressions", J. CoRR, arXiv:1605.02688, 472, p. 473, 2016.

[12] D. R. Anderson, V. M. Patella, "Automated Static Perimetry", London: Mosby, 1999.

[13] N.S. Anstice, and B. Thompson, "The measurement of visual acuity in children: an evidence-based update", Clin Exp Optom, 97(1):3–11, 2014. https://doi.org/10.1111/cxo.12086

[14] M. U. Akram, A. Tariq, S. Khalid, M. Y. Javed, S. Abbas, and U. U. Yasin, "Glaucoma detection using novel optic disc localization, hybrid feature set and classification techniques", Australasian Physical Engineering Sciences Medicine, 38(4):643–655, 2015. https://doi.org/10.1007/s13246-015-0377-y

[15] B. Alatas, "Chaotic bee colony algorithms for global numerical optimization", Expert Systems Applications, 37(8):5682–5687, 2010.

[16] B. Alatas, E. Akin, and A. B. Ozer, "Chaos embedded particle swarm optimization algorithms", Chaos Solitons Fractals, 40(4):1715–1734, 2009.

[17] A. Almazroa, R. Burman, K. Raahemifar, and V. Lakshminarayanan, "Optic disc and optic cup segmentation methodologies for glaucoma image detection: A survey", J. Ophthalmology, (2015), article ID 180972. https://doi.org/10.1155/2015/180972

[18] A. Almazroa, S. Alodhayb, E. Osman, E. Ramadan,M. Hummadi, M. Dlaim, M. Alkatee, K. Raahemifar, and V. Lakshminarayanan, "Retinal fundus images for glaucoma

analysis: The RIGA dataset", Proc. SPIE 10579:8 , Medical Imaging, 2018. https://doi.org/10.1117/12.2293584

[19] M. F. Armaly, "Optic cup in normal and glaucomatous eyes", Investigative Ophthalmology & Visual Science, 9(6):425–429, 1970.

[20] P. H. Artes, D. M. Hutchison, M. T. Nicolela, R. P. LeBlanc, B. C. Chauhan, "Threshold and variability properties of matrix frequency-doubling technology and standard automated perimetry in glaucoma", Investigative Ophthalmology Visual Science, (2005) 46(7):2451–2457. https://doi.org/10.1167/iovs.05-0135

[21] R. Asaoka, A. Iwase, K. Hirasawa, H. Murata, M. Araie, "Identifying "preperimetric" glaucoma in standard automated perimetry visual fields", Investigative Ophthalmology Visual Science, (2014) 55(12):7814-7820. https://doi.org/10.1167/iovs.14-15120

[22] R. Asaoka, H. Murata, A. Iwase, and M. Araie, "Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier", Ophthalmology, 123(9):1974–1980, 2016.

[23] D.S. Asfaw, P.R. Jones, V.M. Monter, N.D. Smith, and D.P. Crabb, "Does glaucoma alter eye movements when viewing images of natural scenes? A between-eye study", Investigative ophthalmology & visual science, 59(8):3189–3198, 2018.

[24] D.S. Asfaw, P.R. Jones, N.D. Smith, and D.P. Crabb, "Data on eye movements in people with glaucoma and peers with normal vision", Data in brief, 19:1266–1273, 2018.

[25] E. Atashpaz-Gargari and C. Lucas, "Designing an optimal PID controller using colonial competitive algorithm", Proc. Iranian Joint Congress Fuzzy Intelligent Systems, 2007.

[26] E. Atashpaz-Gargari and C. Lucas, "Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition", Proc. IEEE Congress Evolutionary Computation, pp. 4661–4667, 2007.

[27] E. Atashpaz-Gargari, F. Hashemzadeh, and C. Lucas, "Designing MIMO PIID controller using colonial competitive algorithm: Applied to distillation column process", Proc. IEEE Congress Evolutionary Computation, pp. 1929–1934, 2008.

[28] E. Atashpaz-Gargari, F. Hashemzadeh, R. Rajabioun, and C. Lucas, "Colonial competitive algorithm", Int. J. Intelligent Computing Cybernetics, 1(3):337–355, 2008.

[29] G.R. Barnes, "A procedure for the analysis of nystagmus and other eye movements", Aviat Space Environ Med, 53(7):676–682, 1982.

[30] J. Behnamian and M. Zandieh, "A discrete colonial competitive algorithm for hybrid flowshop scheduling to minimize earliness and quadratic tardiness penalties", Expert Systems Applications, 38(12):14490–14498, 2011.

[31] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, "Evaluation measures for models assessment over imbalanced datasets", J. Information Engineering Applications, 3(10):27–38, 2013.

[32] D. Bizios, A. Heijl, B. Bengtsson, "Integration and fusion of standard automated perimetry and optical coherence tomography data for improved automated glaucoma diagnostics", BMC Ophthalmology, 11(1):1–11, 2011. https://doi.org/10.1186/1471-2415-11-20

[33] D. Bizios, A. Heijl, and B. Bengtsson, "Trained artificial neural network for glaucoma diagnosis using visual field data: a comparison with conventional algorithms", J. Glaucoma, 16(1):20–28, 2007. https://doi.org/10.1097/IJG.0b013e31802b34e4

[34] R. Bock, J. Meier, G. Michelson, L. G. Nyúl, J. Hornegger, "Classifying glaucoma with image-based features from fundus photographs", Int. Joint Pattern Recognition Symp., pp. 355–364, 2007. doi: 10.1007/978-3-540-74936-3_36.

[35] R. Bock, J. Meier, L. G. Nyúl, J. Hornegger, and G. Michelson, "Glaucoma risk index: Automated glaucoma detection from color fundus images", Medical Image Analysis, 14(3):471–481, 2010. https://doi.org/10.1016/j.media.2009.12.006

[36] C. Bowd, J. Hao, I. M. Tavares, F. A. Medeiros, L. M. Zangwill, T. W. Lee, P. A. Sample, R. N. Weinreb, and M. H. Goldbaum, "Bayesian machine learning classifiers for combining structural and functional measurements to classify healthy and glaucomatous eyes", Investigative Ophthalmology Visual Science, 49(3):945–953, 2008. https://doi.org/10.1167/iovs.07-1083

[37] C. Bowd, I. Lee, M. H. Goldbaum, M. Balasubramanian, F. A. Medeiros, L. M. Zangwill, C. A. Girkin, J. M. Liebmann, and R. N. Weinreb, "Predicting glaucomatous progression in glaucoma suspect eyes using relevance vector machine classifiers for combined structural and functional measurements", Investigative Ophthalmology Visual Science, 53(4):2382–2389, 2012. https://doi.org/10.1167/iovs.11-7951

[38] C. Bowd, R. N. Weinreb, L. M. Zangwill, "Evaluating the optic disc and retinal nerve fiber layer in glaucoma. I: Clinical examination and photographic methods", Seminars Ophthalmology, 15(4):194–205, 2000.

[39] C. Bowd, R. N. Weinreb, M. Balasubramanian, I. Lee, G. Jang, S. Yousefi, L. M. Zangwill, F. A. Medeiros, C. A. Girkin, J. M. Liebmann, and M. H. Goldbaum, "Glaucomatous patterns in Frequency Doubling Technology (FDT) perimetry data identified by unsupervised machine learning classifiers", PLoS One, 9(1):e85941, 2014. https://doi.org/10.1371/journal.pone.0085941

[40] R. R. Bourne, "The optic nerve head in glaucoma", Community Eye Health, 19(59):44–45, 2006.

[41] C. F. Bosworth, P. A. Sample, C. A. Johnson , R. N. Weinreb, "Current practice with standard automated perimetry", Seminars Ophthalmology, 15(4):172–181, 2000.

[42] M. Broughton, G. Verdon, T. McCourt, A. J. Martinez, J. Hyeon Yoo, S. V. Isakov, P. Massey, M. Yuezhen Niu, R. Halavati, E. Peters, M. Leib, A. Skolik, M. Streif, D. Von Dollen, J. R. McClean, S. Boixo, D. Bacon, A. K. Ho, H. Neven, M. Mohseni, "TensorFlow quantum: a software framework for quantum machine learning", arXiv preprint arXiv:2003.02989, 2020.

[43] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images", Int. J. Biomedical Imaging, 20, 2013. https://doi.org/10.1155/2013/154860

[44] Z. Burgansky-Eliash, G. Wollstein, T. Chu, J. D. Ramsey, C. Glymour, R. J. Noecker, H. Ishikawa, and J. S. Schuman, "Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study", Investigative Ophthalmology Visual Science, 46(11):4147–4152, 2005. https://doi.org/10.1167/iovs.05-0366

[45] C. F. Burgoyne, "Optic nerve: The glaucomatous optic nerve", in Pearls of Glaucoma Management, pp: 1–15, Springer, Berlin, 2016.

[46] I. I. Bussel, G. Wollstein, and J. S. Schuman, "OCT for glaucoma diagnosis, screening and detection of glaucoma progression", British Journal of Ophthalmology, 98(Suppl 2):ii15–ii19, 2014.

[47] R. Caponetto, L. Fortuna, S. Fazzino, and M. G. Xibilia, "Chaotic sequences to improve the performance of evolutionary algorithms", IEEE Trans. Evolutionary Computation, 7(3):289–304, 2003.

[48] E. J. Carmona, M. Rincón, J. García-Feijoo, and J. M. Martínez-de-la-Casa, "Identification of the optic nerve head with genetic algorithms", J. Artificial Intelligence Medicine, 43(3):243–259, 2008. https://doi.org/10.1016/j.artmed.2008.04.005

[49] L.Y.L. Chang, P. Guo, B. Thompson, M. Sangi, and J. Turuwhenua, "Assessing visual acuity–test-retest repeatability and level of agreement between the electronic ETDRS chart (E-ETDRS), optokinetic nystagmus (OKN), and sweep VEP", Investigative Ophthalmology & Visual Science, 59(9):5789–5789, 2018.

[50] B. C. Chauhan, and C. F. Burgoyne, "From clinical examination of the optic disc to clinical assessment of the optic nerve head: a paradigm change", American J. Ophthalmology, 156(2):218–227, 2013. https://doi.org/10.1016/j.ajo.2013.04.016

[51] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system", In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785–794). ACM, 2016. https://doi.org/10.1145/2939672.2939785

[52] X. Chen, Y. Xu, S. Yan, D. W. K. Wong, T. Y. Wong, and J. Liu, "Automatic feature learning for glaucoma detection based on deep learning", Proc. Int. Conf. Medical Image Computing Computer-Assisted Intervention, pp. 669–677, 2015. https://doi.org/10.1007/978-3-319-24574-4_80

[53] X. Chen, Y. Xu, D.W.K. Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network", IEEE Annual Int. Conf. Engineering Medicine Biology Society, pp. 715–718, 2015. https://doi.org/10.1109/EMBC.2015.7318462

[54] J. Y. Choi, T. K. Yoo, J. G. Seo, J. Kwak, T. T. Um, and T. H. Rim, "Multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database," PloS one, 12(11): p.e0187336, 2017.

[55] M. Christopher, A. Belghith, C. Bowd, J. A. Proudfoot, M. H. Goldbaum, R. N. Weinreb, C. A. Girkin, J. M. Liebmann, and L. M. Zangwill, "Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs", Nature, Scientific reports, 8(1):16685, 2018. https://doi.org/10.1038/s41598-018-35044-9

[56] D. Ciresan, U. Meier, and J. Schmidhuber, Multi-column deep neural networks for image classification, in: IEEE Conf. Computer Vision Pattern Recognition, pp: 3642–3649, 2012. https://doi.org/10.1109/CVPR.2012.6248110

[57] C.J.W. Connell, B. Thompson, J. Turuwhenua, R.F. Hess, and N. Gant, "Caffeine increases the velocity of rapid eye movements in unfatigued humans", Psychopharmacology, 234(15):2311–2323, 2017.

[58] T. Coye, "A novel retinal blood vessel segmentation algorithm for fundus images," MATLAB Central File Exchange, 2016.

[59] F. Chollet, "Xception: Deep learning with depthwise separable convolutions", 2016. https://doi.org/10.1109/CVPR.2017.195

[60] M. Christopher, A. Belghith, R.N. Weinreb, C. Bowd, M.H. Goldbaum, L.J. Saunders, F.A. Medeiros, and L.M. Zangwill, "Retinal nerve fiber layer features identified by unsupervised machine learning on optical coherence tomography scans predict glaucoma progression", Investigative Ophthalmology & Visual Science, 59(7):2748–2756, 2018.

[61] R. R. Curtin, J. R. Cline, N. P. Slagle, W. B. March, P. Ram, N. A. Mehta, and A. G. Gray, "MLPACK: A scalable C++ machine learning library", J. Machine Learning Research, , 2013. 14:801–805.

[62] E. Decenciére, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, and A. Erginay, "Feedback on a publicly distributed image database: the Messidor database," Image Analysis Stereology, 33(3):231–234, 2014.

[63] S. Demirel, B. Fortune, J. Fan, R. A. Levine, R. Torres, H. Nguyen, S. L. Mansberger, S. K. Gardiner, G. A. Cioffi, and C. A. Johnson, "Predicting progressive glaucomatous optic neuropathy using baseline standard automated perimetry data", Investigative Ophthalmology Visual Science, 50(2):674–680, 2009. https://doi.org/10.1167/iovs.08-1767

[64] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and F.-F. Li, "Imagenet: A large-scale hierarchical image database", IEEE Conf. Computer Vision Pattern Recognition, pp. 248–255, 2009. https://doi.org/10.1109/CVPR.2009.5206848

[65] L. Deng, and D. Yu, "Deep learning: methods and applications", J. Foundations and Trends in Signal Processing, 7(3–4):197–387, 2014.

[66] A. Dey, and S.K. Bandyopadhyay, "Automated glaucoma detection using support vector machine classification method", J. Advances in Medicine and Medical Research, pp.1–12, 2016.

[67] L. Ding, H. Zhang, J. Xiao, B. Li, S. Lu, and M. Norouzifard, "An improved image mixed noise removal algorithm based on super-resolution algorithm and CNN", Neural Computing and Applications, pp. 1–12, 2018. https://doi.org/10.1007/s00521-018-3777-6

[68] M.R. Dix, "The Mechanism and Clinical Significance of Optokinetic Nystagmus", The J. Laryngology & Otology, 94(8):845–864, 1980. https://doi.org/10.1017/S0022215100089611

[69] L. dos Santos Coelho and V. C. Mariani, "Use of chaotic sequences in a biologically inspired algorithm for engineering design optimization", Expert Systems Applications, 34(3):1905–1913, 2008.

[70] S.M. Doustkouhi, P.R. Turnbull, and S.C. Dakin, "The effect of simulated visual field loss on optokinetic nystagmus", Translational Vision Science & Technology, 9(3):25–25, 2020.

[71] S. Dua, U. R. Acharya, P. Chowriappa, and S. V. Sree, "Wavelet-based energy features for glaucomatous image classification", IEEE Trans. Information Technology Biomedicine, 16(1):80–87, 2012. https://doi.org/10.1109/TITB.2011.2176540

[72] B. J. Erickson, P. Korfiatis, Z. Akkus, T. Kline, and K. Philbrick, "Toolkits and libraries for deep learning", J. Digital Imaging, 30(4):400–405, 2017. https://doi.org/10.1007/s10278-017-9965-6

[73] European Glaucoma Prevention Study Group and others, "Reproducibility of evaluation of optic disc change for glaucoma with stereo optic disc photographs", Ophthalmology, 110(2):340–344, 2003.

[74] E. Falkenauer, "Genetic Algorithms and Grouping Problems", Wiley, New York, 1998.

[75] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search", Biomedical Optics Express, 8(5):2732–2744, 2017.

[76] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation", Int. J. Computer Vision, 59(2):167–181, 2004.

[77] T.D. Fife, R.J. Tusa, J.M. Furman, D.S. Zee, E. Frohman, R.W. Baloh, T. Hain, J. Goebel, J. Demer, and L. Eviatar, "Assessment: Vestibular testing techniques in adults and children: Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology", Neurology, 55(10):1431–1441, 2000.

[78] M. Fingeret, "Cup-to-disc ratio not crucial to glaucoma documentation", Primary Care Optometry News, June 2005 (online).

[79] M. Fingeret, F. A. Medeiros, R. Susanna Jr, and R.N. Weinreb, "Five rules to evaluate the optic disc and retinal nerve fiber layer for glaucoma", Optometry J. American Optometric Association, 76(11):661–668, 2005. https://doi.org/10.1016/j.optm.2005.08.029

[80] J. Flammer, S. M. Drance, L. Augustiny, and A. Funkhouser, "Quantification of glaucomatous visual field defects with automated perimetry", Investigative Ophthalmology Visual Science, 26(2):176–181, 1985.

[81] P. J. Foster, R. Buhrmann, H. A. Quigley, and G. J. Johnson, "The definition and classification of glaucoma in prevalence surveys", British J. Ophthalmology, 86(2):238–242, 2002.

[82] Y. Freund, and R. Schapire, "A decision theoretic generalization of on-line learning and an application to boosting", J. Computer and System Sciences 55(1):119–139, 1997.

[83] D. S. Friedman, R. C. Wolfs, B. J. O'Colmain, B. E. Klein, H. R. Taylor, S. West, M. C. Leske, P. Mitchell, N. Congdon, J. Kempen, "Prevalence of open-angle glaucoma among adults in the United States", Archives Ophthalmology, 122(4):532–538, 2004. https://doi.org/10.1001/archopht.122.4.532

[84] H. Fu, D. W. K. Wong, J. Cheng, Y. Xu, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation", IEEE Trans. Medical Imaging, 37(7):1597–1605, 2018 (code on github.com/HzFu/DENet_GlaucomaScreen). https://doi.org/10.1109/TMI.2018.2791488

[85] J. G. Fujimoto, M. E. Brezinski, G. J. Tearney, S. A. Boppart, B. Bouma, M. R. Hee, J. F. Southern, and E. A. Swanson, "Optical biopsy and imaging using optical coherence tomography", Nature medicine, 1(9):970–972, 1995.

[86] R. L. Furlanetto, S. H. Teixeira, C.P.B. Gracitelli, C. L. Lottenberg, F. Emori, M. Michelan, E. Amaro Jr, and A. Paranhos Jr, "Structural and functional analyses of the optic nerve and lateral geniculate nucleus in glaucoma", PloS one, 13(3):e0194038, 2018. https://doi.org/10.1371/journal.pone.0194038

[87] B. Fulkerson, A. Vedaldi, and S. Soatto, "Class segmentation and object localization with superpixel neighborhoods", Proc. IEEE Int. Conf. Computer Vision, 2009.

[88] F. Fumero, S. Alayon, J. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "Rim-one: An open retinal image database for optic nerve evaluation", Proc. IEEE Int. Symp. Computer-based Medical Systems, (2011). https://doi.org/10.1109/CBMS.2011.5999143

[89] F. Fumero, J. Sigut, S. Alaón, M. González-Hernández, and M. González de la Rosa, "Interactive tool and database for optic disc and cup segmentation of stereo and monocular retinal fundus images", Proc. Int. Conf. Central Europe Computer Graphics Visualization Computer Vision, (2015) pp. 91–97.

[90] D. E. Gaasterland, B. Blackwell, L. G. Dally, J. Caprioli, L. J. Katz, F. Ederer, and *Advanced Glaucoma Intervention Study* investigators, "The advanced glaucoma intervention study (AGIS): 10. Variability among academic glaucoma subspecialists in assessing optic disc notching", Trans. American Ophthalmological Society, 99:177, 2001.

[91] M. L. Gabriele, H. Ishikawa, G. Wollstein, R. A. Bilonick, K. A. Townsend, L. Kagemann, M. Wojtkowski, V. J. Srinivasan, J. G. Fujimoto, J. S. Duker, and J. S. Schuman, "Optical coherence tomography scan circle location and mean retinal nerve fiber layer measurement variability", Investigative ophthalmology & visual science, 49(6):2315–2321, 2008.

[92] T. Gajarsky, H. Purwins, "An Xception residual recurrent neural network for audio event detection and tagging", in Proc. Sound Music Computing Conf. ,2018.

[93] G.O. Gajbhiye, A.N. Kamthane, "Automatic classification of glaucomatous images using wavelet and moment feature", In 2015 annual IEEE India conf. (INDICON), pp. 1–5, IEEE, 2015.

[94] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches", IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(4):463–484, 2011.

[95] S. Garbutt, M.R. Harwood, and C.M. Harris, "Comparison of the main sequence of reflexive saccades and the quick phases of optokinetic nystagmus", British Journal of Ophthalmology, 85(12):1477–1483, 2001.

[96] S. Garrido-Jurado, R. Munoz-Salinas, F.J. Madrid-Cuevas, and M.J. Marin-Jimenez, "Automatic generation and detection of highly reliable fiducial markers under occlusion", Pattern Recognition, 47(6):2280–2292, 2014. https://doi.org/10.1016/j.patcog.2014.01.005

[97] D. F. Garway-Heath, S. T. Ruben, A. Viswanathan, and R. A. Hitchings, "Vertical cup/disc ratio in relation to optic disc size: Its value in the assessment of the glaucoma suspect", British J. Ophthalmology, 82(10):1118–1124, 1998.

[98] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets", Proc. Advances Neural Information Processing Systems, pp. 2672–2680, 2014.

[99] H. A. N. S. Goldmann and T. H. Schmidt, "Applanation tonometry", Ophthalmologica, 134(4):221–242, 1957.

[100] M. O. Gordon, J. A. Beiser, J. D. Brandt, D. K. Heuer, E. J. Higginbotham, C. A. Johnson, J. L. Keltner, J. P. Miller, R. K. Parrish, M. R. Wilson, M. A. Kass, "The ocular hypertension treatment study: Baseline factors that predict the onset of primary open-angle glaucoma", Archives Ophthalmology, 120(6):714–720, 2002.

[101] N. Gour, and P. Khanna, "Automated glaucoma detection using GIST and pyramid histogram of oriented gradients (PHOG) descriptors", Pattern Recognition Letters, 2019.

[102] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks", Pattern Recognition, 77: 354–377, 2018. https://doi.org/10.1016/j.patcog.2017.10.013

[103] M. S. Haleem, L. Han, J. van Hemert, and B. Li, "Automatic extraction of retinal features from colour retinal images for glaucoma diagnosis: A review", Computerized Medical Imaging Graphics, 37(7):581–596, 2013. https://doi.org/10.1016/j.compmedimag.2013.09.005

[104] P.A. Harris, T. Garner, M. Sangi, P. Guo, J. Turuwhenua, and B. Thompson, "Visual acuity assessment in adults using optokinetic nystagmus", Investigative Ophthalmology & Visual Science, 60(9):5907–5907, 2019.

[105] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition", Proc. IEEE Conf. Computer Vision Pattern Recognition, pp. 770–778, 2016. https://doi.org/10.1109/CVPR.2016.90

[106] A. Heijl, MC.Leske, B. Bengtsson, L. Hyman, B. Bengtsson, M. Hussein, "Early Manifest Glaucoma Trial Group. Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial", Arch Ophthalmol. 120:1268–1279, 2002.

[107] G. E. Hinton, S. Osindero, and Y. -W. Teh, "A fast learning algorithm for deep belief nets", Neural Computation, 18(7):1527–1554, 2006. https://doi.org/10.1162/neco.2006.18.7.1527

[108] H. Hirasawa, H. Murata, C. Mayama, M. Araie, R. Asaoka, "Evaluation of various machine learning methods to predict vision-related quality of life from visual field data and visual acuity in patients with glaucoma", British J. Ophthalmology, 98(9):1230–1235, 2014. https://doi.org/10.1136/bjophthalmol-2013-304319

[109] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information", IEEE Trans. Pattern Analysis Machine Intelligence, 30(2): 328–341, 2008.

[110] J.M. Holmes, E.L. Lazar, B.M. Melia, W.F. Astle, L.R. Dagi, S.P. Donahue, M.G. Frazier, R.W. Hertle, M.X. Repka, G.E. Quinn, and K.K. Weise, "Effect of Age on Response to Amblyopia Treatment in Children", Arch Ophthalmol, 129(11):1451–1457, 2011.

[111] S. Hosseini, and A. Al Khaled, "A survey on the imperialist competitive algorithm metaheuristic: Implementation in engineering domain and directions for future research," Applied Soft Computing, 24:1078–1094, 2014.

[112] I. Howard and B. Rogers, "Binocular vision and stereopsis", New York: Oxford University Press, 1995.

[113] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications", arXiv preprint arXiv:1704. 04861, 2017.

[114] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks", Proc. IEEE Conf. Computer Vision Pattern Recognition, 1(2):3, pp. 4700–4708, 2017.

[115] K. H. Huang, M. C. Tai, L. C. Lee, T. H. Weng, Y. H. Chen, L. F. Lin, J. T. Chen, D. W. Lu, and C. L. Chen, "Positron emission tomography/computed tomography scan of Vogt–Koyanagi–Harada syndrome with associated autoimmune thyroid disease: A case report and literature review", Medicine, 97(9), 2018.

[116] J.Y. Hyon, H.E. Yeo, J.-M. Seo, I.B. Lee, J.H. Lee, and J.-M. Hwang, "Objective measurement of distance visual acuity determined by computerized optokinetic nystagmus test", Invest. Ophthalmol. Vis. Sci., 51(2):752–757, 2010. https://doi.org/10.1167/iovs.09-4362

[117] S. B. Imandoust, M. Bolandraftar, "Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background", Int. J. Engineering Research Applications, 3(5):605–610, 2013.

[118] Y. Ishiyama, H. Murata, and R. Asaoka, "The usefulness of gaze tracking as an index of visual field reliability in glaucoma patients", Investigative ophthalmology & visual science, 56(11):6233–6236, 2015.

[119] A. K. Jain, "Data clustering: 50 years beyond K-means", Pattern Recognition Letters, 31(8):651–666, 2010.

[120] A. Jasour, E. Atashpaz-Gargari, and C. Lucas, "Vehicle fuzzy controller design using imperialist competitive algorithm", Proc. Iranian Joint Congress Fuzzy Intelligent Systems, 2008.

[121] H. Jin, Q. Song, X. Hu, "Auto-keras: An efficient neural architecture search system", Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining, pp. 1946–1956, 2019. https://doi.org/10.1145/3292500.3330648

[122] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. "Caffe: Convolutional architecture for fast feature embedding", Proc. ACM Int. Conf. Multimedia, pp. 675–678, 2014. https://doi.org/10.1145/2647868.2654889

[123] C. A. Johnson, "Detecting functional changes in the patient's vision: Visual field analysis", in: Clinical Glaucoma Care, Springer, New York, pp. 117–159, 2014.

[124] C. A. Johnson, "Detecting functional changes in the patient's vision: visual field analysis", In Clinical Glaucoma Care pp. 229–263, Springer, New York, NY, 2010. https://doi.org/10.1007/978-0-387-76700-0_23

[125] C. A. Johnson, G. A. Cioffi, J. R. Liebmann, P.A. Sample, L. M. Zangwill, R. N. Weinreb, "The relationship between structural and functional alterations in glaucoma: a review", Semin Ophthalmol. 15:221–233, 2000.

[126] C. A. Johnson, P. A. Sample, G. A. Cioffi, J. R. Liebmann, and R. N. Weinreb, "Structure and function evaluation (SAFE): I. criteria for glaucomatous visual field loss using standard automated perimetry (SAP) and short wavelength automated perimetry (SWAP)", American J. Ophthalmology, 134(2):177–185, 2002. https://doi.org/10.1016/S0002-9394(02)01577-5

[127] J. B. Jonas and A. Dichtl, "Evaluation of the retinal nerve fiber layer", Survey Ophthalmology, 40:369–78, 1996.

[128] M. Juhola, "Detection of nystagmus eye movements using a recursive digital filter", IEEE Transactions on Biomedical Engineering, 35(5):389–395, 1988.

[129] L.P. Kaelbling, M.L. Littman, and A.W. Moore, "Reinforcement learning: A survey", J. artificial intelligence research, 4:237–285, 1996.

[130] M. A. Kass, D. K. Heuer, E. J. Higginbotham, C. A. Johnson, J. L. Keltner, J. P. Miller, R. K. Parrish, M. R. Wilson, and M. O. Gordon, "The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma", Arch Ophthalmol. 120:701–13, 2002.

[131] J. Kaur, S. Agarwal, and R. Vig, "A methodology for the performance analysis of cluster based image", Int. J. Engineering Research Application, 2(2):664–667, 2012.

[132] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to Platt's SMO algorithm for SVM classifier design", Neural Computation, 13(3):637–649, 2001. https://doi.org/10.1162/089976601300014493

[133] T. Khalil, M. U. Akram, S. Khalid, and A. Jameel, "Improved automated detection of glaucoma from fundus image using hybrid structural and textural features", IET Image Processing, 11(9):693–700, 2017. https://doi.org/10.1049/iet-ipr.2016.0812

[134] S.J. Kim, K.J. Cho, and S. Oh, "Development of machine learning models for diagnosis of glaucoma", PLoS One, 12(5):p.e0177726, 2017.

[135] M. Kim, J.C. Han, S.H. Hyun, O. Janssens, S. Van Hoecke, C. Kee, and W. De Neve, "Medinoid: Computer-Aided Diagnosis and Localization of Glaucoma Using Deep Learning", J. Applied Sciences, 9(15):3064, 2019. https://doi.org/10.3390/app9153064

[136] S. Kingman, "Glaucoma is second leading cause of blindness globally," Bulletin World Health Organization, 82:887–888, 2004.

[137] M. Kipp, "ANVIL - A Generic Annotation Tool for Multimodal Dialogue", 2001.

[138] R. Klette, "Concise computer vision," Stereo Matching, 8:287–330, Springer, London, 2014.

[139] R. Klette and A. Rosenfeld, "Digital Geometry", Morgan Kaufmann, San Francisco, 2004.

[140] D. E. Knuth, T. L. Larrabee, P. M. Roberts: *Mathematical Writing*, Mathematical Association of America, ISBN: 088385063X, 1996.

[141] M.J.G. Kooiker, J.J.M. Pel, H.J.M. Verbunt, G.C. de Wit, M.M. van Genderen, and J. van der Steen, "Quantification of visual function assessment using remote eye tracking in children: validity and applicability", Acta Ophthalmologica, 94(6):599–608, 2016

[142] J. Kotowski, G. Wollstein, L. S. Folio, H. Ishikawa, and J. S. Schuman, "Clinical use of OCT in assessing glaucoma progression", Ophthalmic Surgery Lasers Imaging Retina, 42(4):S6–S14, 2011. https://doi.org/10.3928/15428877-20110627-01

[143] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", Proc. Advances Neural Information Processing Systems, pp. 1097–1105, 2012. https://doi.org/10.1145/3065386

[144] B. R. LaHood, J. Erceg, T. H. Bevin, and G. Sanderson, "High rate of incidental glaucoma detection in New Zealand", Ethnicity, 25, p. 7, 2016.

[145] F. N. Landola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer. "SqueezeNet: AlexNet-level accuracy with 50$\times$ fewer parameters and < 0.5 MB model size", 2016. arXiv preprint arXiv:1602.07360.

[146] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", Proc. IEEE, 86(11):2278–2324, 1998. https://doi.org/10.1109/5.726791

[147] Y. LeCun, L. Jackel, L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger, P. Simard, U. A. Muller, E. Sackinger, P. Simard, V. Vapnik, "Learning algorithms for classification: A comparison on handwritten digit recognition", Neural Networks: The Statistical Mechanics Perspective, pp. 261–276, 1995.

[148] S.S.Y. Lee, A.A. Black, and J.M. Wood, "Effect of glaucoma on eye movement patterns and laboratory-based hazard detection ability", PloS one, 12(6):e0178876, 2017.

[149] J. Lee, Y. Kim, J.H. Kim, and K.H. Park, "Screening glaucoma with red-free fundus photography using deep learning classifier and polar transformation", J. of Glaucoma, 28(3):258–264, 2019.

[150] S.L. Levens, "Electronystagmography in normal children", British J. Audiology 22(1):51–56, 1988.

[151] F. Li and R. Klette, "A variant of adaptive mean shift-based clustering", Springer, LNCS 5506, pp. 1002–1009, 2008.

[152] F. Li, Z. Wang, G. Qu, Y. Qiao, and X. Zhang, "Visual field based automatic diagnosis of glaucoma using deep convolutional neural network", In: Computational Pathology Ophthalmic Medical Image Analysis, pp. 285–293, 2018. https://doi.org/10.1007/978-3-030-00949-6_34

[153] F. Li, Z. Wang, G. Qu, D. Song, Y. Yuan, Y. Xu, K. Gao, G. Luo, Z. Xiao, D. S. Lam, H. Zhong, Y. Qiao and X. Zhang, "Automatic differentiation of Glaucoma visual field from non-glaucoma visual filed using deep convolutional neural network", BMC Medical Imaging, 18(1):35, 2018. https://doi.org/10.1186/s12880-018-0273-5

[154] G. Lim, Y. Cheng, W. Hsu, and M. L. Lee, "Integrated optic disc and cup segmentation with deep learning", in IEEE Int. Conf. Tools Artificial Intelligence, pp. 162–169, 2015.

[155] X. Lin, G. Klette, R. Klette, J. Craig, and S. Dean, "Accurately measuring the size of the pupil of the eye", Proc. Image Vision Computing New Zealand, sprg.massey.ac.nz/ivcnz/Proceedings/IVCNZ_40.pdf, (2003).

[156] G. Litjens, T. Kooi, B. Ehteshami Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. Van Ginneken, C. I. Sánchez, "A survey on deep learning in medical image analysis", Medical Image Analysis, 42:60–88, 2017. https://doi.org/10.1016/j.media.2017.07.005

[157] K. A. Lucy, and G. Wollstein, "Structural and functional evaluations for the early detection of glaucoma. Expert review of ophthalmology", 11(5):367–376, 2016.

[158] J. MacQueen, "Some methods for classification and analysis of multivariate observations", Proc. Berkeley Symp. Mathematical Statistics Probability, 1(14):281–297, 1967.

[159] T. S. Madhulatha, "An overview on clustering methods", arXiv preprint, arXiv:1205.1117, 2012.

[160] S. Maetschke, B. Antony, H. Ishikawa, G. Wollstein, J. Schuman, and R. Garnavi, "A feature agnostic approach for glaucoma detection in OCT volumes", PloS one, 14(7):p.e0219126, 2019.

[161] P. Massin, E. Vicaut, B. Haouchine, A. Erginay, M. Paques, and A. Gaudric, "Reproducibility of retinal mapping using optical coherence tomography", Archives Ophthalmology, 119(8):1135–1142, 2001. https://doi.org/10.1001/archopht.119.8.1135

[162] R. M. May, "Simple mathematical models with very complicated dynamics," Nature, 261(5560):459–67, 1976.

[163] F. A. Medeiros, L. M. Zangwill, C. Bowd, P. A. Sample, R. N. Weinreb, "Use of progressive glaucomatous optic disk change as the reference standard for evaluation of diagnostic tests in glaucoma", American J. Ophthalmology, 139(6):1010–1018, 2005.

[164] J. Meier, R. Bock, G. Michelson, L. G. Nyúl, and J. Hornegger, "Effects of preprocessing eye fundus images on appearance based glaucoma classification", Proc. Int. Conf. Computer Analysis Images Patterns, pp. 165–172, 2007. https://doi.org/10.1007/978-3-540-74272-2_21

[165] X. Meng, J. Bradley, B. Yavuz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman,D. B. Tsai, M. Amde, S. Owen, and D. Xin, "Mllib: Machine learning in apache spark", J. Machine Learning Research, 17(1):1235–1241, 2016.

[166] S. Meng, M. Lazarescu, J. Ivins, A. Turpin, "Monitoring glaucomatous progression: classification of visual field measurements using stable reference data", In: Advances Machine Learning Cybernetics, Springer, Berlin, pp. 750–759, 2006. https://doi.org/10.1007/11739685_78

[167] F. S. Mikelberg, C. M. Parfitt, N. V. Swindale, S. L. Graham, S. M. Drance, R. Gosine, "Ability of the Heidelberg Retina Tomograph to Detect Early Glaucomatous Visual Field Loss", J. Glaucoma, 4(4):242–247, 1995.

[168] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement", Annals of internal medicine, 151(4):264–269, 2009.

[169] J. E. Morgan, I. Bourtsoukli, K. N. Rajkumar, E. Ansari, I. A. Cunliffe, R. V. North, and J. M. Wild, "The accuracy of the inferior $>$ superior $>$ nasal $>$ temporal neuroretinal rim area rule for diagnosing glaucomatous optic disc damage", Ophthalmology, 119(4):723–730, 2012. https://doi.org/10.1016/j.ophtha.2011.10.004

[170] S. J. Nanda and G. Panda, "A survey on nature inspired metaheuristic algorithms for partitional clustering", Swarm Evolutionary Computation, 16: 1–18, 2014.

[171] J. Nayak, R. Acharya, P. S. Bhat, N. Shetty, and T. -C. Lim. "Automated diagnosis of glaucoma using digital fundus images", J. Medical Systems, 33(5):337–346, 2009.

[172] T. Niknama, E. T. Fard, N. Pourjafarian, and A. Rousta, "An efficient hybrid algorithm based on modified imperialist competitive algorithm and K-means for data clustering," Engineering Applications Artificial Intelligence, 24(2):306–317, 2011.

[173] H. Nomoto, C. Matsumoto, S. Takada, S. Hashimoto, E. Arimura, S. Okuyama, and Y. Shimomura, "Detectability of glaucomatous changes using SAP, FDT, flicker perimetry, and OCT", J. Glaucoma, 18(2):165–171, 2009. https://doi.org/10.1097/IJG.0b013e318179f7ca

[174] A. Y. Ng, "Feature selection, L 1 vs. L 2 regularization, and rotational invariance", In Proc. the twenty-first inter. conf. Machine learning, pp. 78–86, 2004.

[175] M. Norouzifard, A. Abdollahi Dehkordi, M. Naderi Dehkordi, H.GholamHosseini, and R. Klette, "Unsupervised optic cup and optic disc segmentation for glaucoma detection by ICICA", Proc. Pervasive Systems Algorithms Networks (I-SPAN), 15th International Symposium on pp. 209–214, 2018. https://doi.org/10.1109/I-SPAN.2018.00041

[176] M. Norouzifard, J. Black, B. Thompson, R. Klette, and J. Turuwhenua, "A state machine for real-time eye tracking to detect optokinetic nystagmus", InAsian Conf. Pattern Recognition, Springer, pp. 143–155, 2019. https://doi.org/10.1007/978-3-030-41299-9_12

[177] M. Norouzifard, A. Dawda, A. Abdul-Rahman, H. GholamHosseini, and R. Klette, "Superpixel segmentation methods on stereo fundus images and disparity map for glaucoma detection", in: Int. Conf. Image and Vision Computing New Zealand, IEEE Explore, pp. 1–6, 2018. https://doi.org/10.1109/IVCNZ.2018.8634732

[178] M. Norouzifard, A. Nemati, A. Abdul-Rahman, H. GholamHosseini, and R. Klette, "A comparison of transfer learning techniques, deep convolutional neural network and multilayer neural network methods for the diagnosis of glaucomatous optic neuropathy", in: Int. Computer Symp.,Taiwan, Springer, pp. 627–635, 2019. https://doi.org/10.1007/978-981-13-9190-3_69

[179] M. Norouzifard, A. Nemati, A. Abdul-Rahman, H. GholamHosseini, and R. Klette, "A Fused Pattern Recognition Model to Detect Glaucoma Using Retinal Nerve Fiber Layer Thickness Measurements" In Pacific-Rim Symposium on Image and Video Technology, Springer, pp. 1–12, 2019. https://doi.org/10.1007/978-3-030-34879-3_1

[180] M. Norouzifard, A. Nemati, R. Klette, H. GholamHosseini, K. Nouri-Mahdavi, and S. Yousefi, "A hybrid machine learning model to detect glaucoma using retinal nerve fiber layer thickness measurements", Investigative Ophthalmology & Visual Science, 60(9):3924–3924, 2019.

[181] M. Norouzifard, A. Nemati, H. GholamHosseini, R. Klette, K. Nouri-Mahdavi, and S. Yousefi, "Automated glaucoma diagnosis using deep and transfer learning: Proposal of a system for clinical testing", in: Int. Conf. Image and Vision Computing New Zealand, IEEE Explore, pp. 1–6, 2018. https://doi.org/10.1109/IVCNZ.2018.8634671

[182] M. Norouzifard, A. Nemati, R. Klette, H. GholamHosseini, K. Nouri-Mahdavi, and S. Yousefi, "Identification of clinically relevant glaucoma biomarkers on fundus images using deep learning", Investigative Ophthalmology & Visual Science, 60(11):PB090–PB090, 2019.

[183] M. Norouzifard, A. Soleymani, and J. Shanbehzadeh, "A new method for measuring auto-fluorescence changes in neovascular-AMD after intravitreal injection of Bavecizumab", Proc. IEEE Int. Conf. Current Trends Information Technology, 2009. https://doi.org/10.1109/CTIT.2009.5423139

[184] T. Oberwahrenbrock, M. Weinhold, J. Mikolajczak, H. Zimmermann, F. Paul, I. Beckers, and A.U. Brandt, "Reliability of intra-retinal layer thickness estimates", J. PloS one, 10(9):e0137316, 2015.

[185] J.I. Orlando, E. Prokofyeva, M. del Fresno, and M.B. Blaschko, "Convolutional neural network transfer for automated glaucoma identification", Int. Symp. Medical Information Processing Analysis, SPIE, 10160: 10, 2017. https://doi.org/10.1117/12.2255740

[186] E. Ott, "Chaos in Dynamical Systems," second edition, p. 478, 2002.

[187] J.O. Palacio-Niño, "Evaluation metrics for unsupervised learning algorithms", arXiv preprint, arXiv:1905.05667, 2019.

[188] T. Pander, R. Czabanski, T. Przybyla, and D. Pojda-Wilczek, "An automatic saccadic eye movement detection in an optokinetic nystagmus signal", Biomedical Engineering / Biomedizinische Technik, 59(6):529–543, 2014. https://doi.org/10.1515/bmt-2013-0137

[189] J.W. Park, B.H. Kang, J.W. Kwon, and K.J. Cho, "Analysis of various factors affecting pupil size in patients with glaucoma", BMC ophthalmology, 17(1):168–175, 2017.

[190] N.B. Patel, M. Sullivan-Mee, and R.S. Harwerth, "The relationship between retinal nerve fiber layer thickness and optic nerve head neuroretinal rim tissue in glaucoma", Investigative ophthalmology & visual science, 55(10):6802–6816, 2014.

[191] J. Patterson, and A. Gibson, "Deep learning: a practitioner's approach", OŔeilly Media, Inc. 2017.

[192] I. Peate and N. Jones, "Glaucoma – early diagnosis could prevent vision loss", British J. Healthcare Assistants, 8(10):482–488, 2014. https://doi.org/10.12968/bjha.2014.8.10.482

[193] L. M. Pecora and T. L. Carroll, "Synchronization in chaotic systems", Physical Review Letters, 64(8):821–824, 1990.

[194] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas, "Scikit-learn: machine learning in Python", J. Machine Learning Research, 12:2825–2830, 2011.

[195] C. Pena-Betancor, M. Gonzalez-Hernandez, F. Fumero-Batista, J. Sigut, E. Medina-Mesa, S. Alayon, and M. González de la Rosa, "Estimation of the relative

amount of hemoglobin in the cup and neuro-retinal rim using stereoscopic color fundus images", J. Investigative Ophthalmology Visual Science, 56(3):1562–1568, 2015. https://doi.org/10.1167/iovs.14-15592

[196] W.H. Press, and S.A. Teukolsky, "Savitzky-Golay smoothing filters", Computers in Physics, 4(6):669, 1990.

[197] H. A. Quigley and A. T. Broman, "The number of people with glaucoma worldwide in 2010 and 2020", British J. Ophthalmology, 90(3):262–267, 2006. https://doi.org/10.1136/bjo.2005.081224

[198] H. A. Quigley, G. R. Dunkelberger, W. R. Green, "Retinal ganglion cell atrophy correlated with automated perimetry in human eyes with glaucoma", American J. Ophthalmology, 107(5):453–464, 1989. https://doi.org/10.1016/0002-9394(89)90488-1

[199] L. Racette, C. Y. Chiou, J. Hao, C. Bowd, M. H. Goldbaum, L. M. Zangwill, T. W. Lee, R. N. Weinreb, and P. A. Sample, "Combining functional and structural tests improves the diagnostic accuracy of relevance vector machine classifiers", J. Glaucoma. 19(3):167–175, 2010. https://doi.org/10.1097/IJG.0b013e3181a98b85

[200] U. Raghavendra, H. Fujita, S. V. Bhandary, A. Gudigar, J. H. Tan, and U. R. Acharya, "Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images", Information Sciences, 441:41–49, 2018. https://doi.org/10.1016/j.ins.2018.01.051

[201] B. Ramsundar, and R. B. Zadeh, "TensorFlow for deep learning: from linear regression to reinforcement learning", OŔeilly Media, Inc. 2018.

[202] M. Ranjbaran, H.L.H. Smith, and H.L. Galiana, "Automatic classification of the vestibulo-ocular reflex nystagmus: integration of data clustering and system identification", IEEE Transactions on Biomedical Engineering, 63(4):850–858, 2016.

[203] A. S. Raza, X. Zhang, C. G. De Moraes, C. A. Reisman, J. M. Liebmann, R. Ritch, and D. C. Hood, "Improving glaucoma detection using spatially correspondent clusters of damage and by combining standard automated perimetry and optical coherence tomography", Investigative Ophthalmology Visual Science, 55(1):612–624, 2014. https://doi.org/10.1167/iovs.13-12351

[204] N. Razmjooy, B. S. Mousavi, and F. Soleymani, "A hybrid neural network imperialist competitive algorithm for skin color segmentation," Mathematical Computer Modelling, 57(3-4):848–856, 2013.

[205] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection", Proc. IEEE Conf. Computer Vision Pattern Recognition, pp. 779–788, 2016.

[206] R.D. Reinecke, D.G. Cogan, "Standardization of objective visual acuity measurements: opticokinetic nystagmus us. snellen acuity", AMA Arch Ophthalmol, 60(3):418–421, 1958. https://doi.org/10.1001/archopht.1958.00940080436010.

[207] C.G. Rey, and H.L. Galiana, "Parametric classification of segments in ocular nystagmus", IEEE J. Transactions on Biomedical Engineering, 38(2):142–148, 1991.

[208] R. Rojas, "AdaBoost and the super bowl of classifiers a tutorial introduction to adaptive boosting", Freie University, Berlin, Tech. Rep. 2009.

[209] S. J. Ryan, and S. R. Sadda, "Ryanś retinal imaging and diagnostics ebook", Elsevier Health Sciences, 2013.

[210] A. A. Salam, T. Khalil, M. U. Akram, A. Jameel, and I. Basit, "Automated detection of glaucoma using structural and non structural features", J. Springerplus, 5(1): 1519, 2016. https://doi.org/10.1186/s40064-016-3175-4

[211] P. A. Sample, "Glaucoma is present prior to its detection with standard automated perimetry: is it time to change our concepts?" Graefe's Archive Clinical Experimental Ophthalmology, 241(3):168–169, 2003.

[212] M. Sangi, B. Thompson, and J. Turuwhenua, "An Optokinetic Nystagmus Detection Method for Use With Young Children', IEEE J. Translational Engineering in Health and Medicine, 3: 1–10, 2015.

[213] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", Int. J. Computer Vision, 47(1-3):7–42, 2002.

[214] K. Schauwecker, "SP1 stereo vision system," Nerian Vision Technologies, Germany, 2017.

[215] A. M. Schulz, E. C. Graham, Y. You, A. Klistorner, and S. L. Graham, "Performance of iPad-based threshold perimetry in glaucoma and controls", Clinical Experimental Ophthalmology, 46(4):346–355, 2018. https://doi.org/10.1111/ceo.13082

[216] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks", 2013. arXiv preprint arXiv:1312.6229.

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

[218] L. Shapiro and G. Stockman, "Computer Vision," pp. 279–325, Pearson, 2001.

[219] P. Sharma, P. A. Sample, L. M. Zangwill, and J. S. Schuman, "Diagnostic tools for glaucoma detection and management," Survey Ophthalmology, 53(6):S17–S32, 2008.

[220] J. Shlens, "A tutorial on principal component analysis", 2014. arXiv preprint arXiv:1404.1100.

[221] R. Sihota, D. Angmo, D. Ramaswamy, and T. Dada, "Simplifying "target" intraocular pressure for different stages of primary open-angle glaucoma and primary angle-closure glaucoma", Indian J. of ophthalmology, 66(4):495, 2018. https://doi.org/10.4103/ijo.IJO_1130_17.

[222] B. W. Silverman, "Density estimation for statistics and data analysis", Routledge, 2018.

[223] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409. 1556, 2014.

[224] S. Simonthomas, N. Thulasi, and P. Asharaf, "Automated diagnosis of glaucoma using Haralick texture features", Proc. Int. IEEE Conf. Information Communication Embedded Systems, page 1–6, 2014. https://doi.org/10.1109/ICICES.2014.7033743

[225] A. Singh, S. Sengupta, and V. Lakshminarayanan, "Glaucoma diagnosis using transfer learning methods", In Applications of Machine Learning Vol. 11139, p. 111390U, Int. Society for Optics and Photonics, Sep 2019. https://doi.org/10.1117/12.2529429

[226] J. Sivaswamy, S. Krishnadas, A. Chakravarty, G. D. Joshi, and A. S. Tabish, "A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis", JSM Biomedical Imaging Data Papers, (2015) 2(1):1004.

[227] C. Sitaraman, K. Sharma, A. Sahai, and W. Gangwal, "Analysis of RNFL thickness in normal, ocular hypertensives and glaucomatous eyes with oct", Proc. All India Ophthalmological Society, p. 246, 2009.

[228] J. Sivaswamy, S. R. Krishnadas, G. D. Joshi, M. Jain, A. U. S. Tabish, and A. Ujjwaft, "Drishti-GS: Retinal image dataset for optic nerve head (ONH) segmentation", Proc. IEEE Int. Symp. Biomedical Imaging, pp. 53–56, 2014. https://doi.org/10.1109/ISBI.2014.6867807

[229] M. Sokolova, J. Nathalie and S. Stan, "Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation", In Australasian Joint Conf. on Artificial Intelligence, pp. 1015–1021. Springer, Berlin, Heidelberg, 2006.

[230] M. Soltaninejad, G. Yang, T. Lambrou, N. Allinson, T. L. Jones, T. R. Barrick, F. A. Howe, and X. Ye, "Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI," Int. J. Computer Assisted Radiology Surgery, 12(2):183–203, 2017.

[231] A. Sommer, J. Katz, H. A. Quigley, N. R. Miller, A. L. Robin, R. C. Richter, and K. A. Witt, "Clinically detectable nerve fiber atrophy precedes the onset of glaucomatous field loss", Arch Ophthalmol. 109:77–83, 1991.

[232] M. Suneel, "Chaotic sequences for secure CDMA," arXiv Preprint nlin/0602018, pp. 6–9, 2006.

[233] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. "Inception-v4, inception-resnet and the impact of residual connections on learning," in AAAI Conf. Artificial Intelligence, pp. 4278–4284, 2017.

[234] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions", Proc. IEEE Conf. Computer Vision Pattern Recognition, 2015. https://doi.org/10.1109/CVPR.2015.7298594

[235] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision", Proc. IEEE Conf. Computer Vision Pattern Recognition, pp. 2818–2826, 2016. https://doi.org/10.1109/CVPR.2016.308

[236] S. Talatahari, B. Farahmand Azar, R. Sheikholeslami, and A. H. Gandomi, "Imperialist competitive algorithm combined with chaos for global optimization," Communications Nonlinear Science Numerical Simulation, 17(3):1312–1319, 2012.

[237] J. H. Tan, U. R. Acharya, S. V. Bhandary, K. C. Chua, S. Sivaprasad, "Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network", J. Computational Science, 20:70–79, 2017.

[238] M. Tayefeh Mahmoudi, N. Forouzideh, C. Lucas, and F. Taghiyareh, "Artificial neural network weights optimization based on imperialist competitive algorithm," Proc. Int. Conf. Computer Science Information Technology, pp. 244–247, 2009.

[239] N. Thakur and M. Juneja, "Survey on segmentation and classification approaches of optic cup and optic disc for diagnosis of glaucoma", Biomedical Signal Processing Control, 42:162–189, 2018. https://doi.org/10.1016/j.bspc.2018.01.014

[240] 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. https://doi.org/10.1016/j.ophtha.2014.05.013

[241] D.S.W. Ting, L.R. Pasquale, L. Peng, J.P. Campbell, A.Y. Lee, R. Raman, G.S.W. Tan, L. Schmetterer, P.A. Keane, and T.Y. Wong, "Artificial intelligence and deep learning in ophthalmology", British Journal of Ophthalmology, 103(2):167–175, 2019.

[242] S. Tong, D. Koller, "Support vector machine active learning with applications to text classification", J. Machine Learning Research, 2:45–66, 2001. https://doi.org/10.1162/153244302760185243

[243] J. Turuwhenua, T.Y. Yu, Z. Mazharullah, and B. Thompson, "A method for detecting optokinetic nystagmus based on the optic flow of the limbus", Vision Res, 103:75–82, 2014.

[244] M.E. Tyler, "Stereo fundus photography: Principles and technique," J Ophthalmic Photogr, 18(2):6–81, 1996.

[245] C. Valmaggia, J. Charlier, and I. Gottlob, "Optokinetic nystagmus in patients with central scotomas in age related macular degeneration", Br. J. Ophthalmol, 85(2):169–172, 2001. https://doi.org/10.1136/bjo.85.2.169

[246] C. Valmaggia, and I. Gottlob, "Optokinetic nystagmus elicited by filling-in in adults with central scotoma", Invest Ophthalmol Vis Sci., 43(6):1804–1808, 2002.

[247] C. Valmaggia, A. Rütsche, A. Baumann, C. Pieh, Y.B. Shavit, F. Proudlock, and I. Gottlob, "Age related change of optokinetic nystagmus in healthy subjects: a study from infancy to senescence", British Journal of Ophthalmology, 88(12):1577–1581, 2004.

[248] R. Varma, W. C. Steinmann, and I. U. Scott, "Expert agreement in evaluating the optic disc for glaucoma," Ophthalmology, 99(2):215–221, 1992.

[249] A. Voulodimos, N. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review", Computational Intelligence Neuroscience, article ID 7068349, 2018. https://doi.org/10.1155/2018/7068349

[250] J. Waddington, and C.M. Harris, "Human optokinetic nystagmus: A stochastic analysis", Journal of Vision, 12(12):5–5, 2012. https://doi.org/10.1167/12.12.5

[251] H. Wahab, S. R. Haider, S. Khitran, N. ul Huda, and M. U. Akram, "Bright region and vessel density based robust optic disc segmentation," in IEEE Int. Conf. Image Processing Theory Tools Applications, pp. 1–6, 2014.

[252] S. Wass, "The use of eye tracking with infants and Children", in Practical Research with Children, Routledge, 2016.

[253] C.J. Watkins, and P. Dayan, "Q-learning", Machine learning, 8(3-4):279–292, 1992.

[254] S. West, and C. Williams, "Amblyopia in children (aged 7 years or less)", BMJ Clin Evid, vol. 2016, 2016.

[255] M. C. Westcott, A. I. McNaught, D. P. Crabb, F. W. Fitzke, and R. A. Hitchings, "High spatial resolution automated perimetry in glaucoma", British J. Ophthalmology, 81(6):452–459, 1997.

[256] World health organization. "Global data on visual impairments 2010," Geneva: World Health Organization Organization, 2012.

[257] J. Wu, Q. Hu, C. Leng, and J. Cheng, "Shoot to know what: An application of deep networks on mobile devices", Proc. AAAI Conf. Artificial Intelligence, pp. 4399–4400, 2016.

[258] A. M. Wilson and A. Di Polo, "Gene therapy for retinal ganglion cell neuroprotection in glaucoma", Gene Therapy, 19(2):127–136, 2012. https://doi.org/10.1038/gt.2011.142

[259] S. Xie, R. Girshick, P. Dollàr, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks", Proc. IEEE Conf. Computer Vision Pattern Recognition, pp. 5987–5995, 2017.

[260] Z. Xu, C. Qian, S. Xu, and R. Klette, "Accurate ellipse extraction in low-quality images", Proc. Int. Conf. Machine Vision Applications, Tokyo, 2019.

[261] D. Yang, G. Li, and G. Cheng, "On the efficiency of chaos optimization algorithms for global optimization," Chaos Solitons Fractals, 34(4):1366–1375, 2007.

[262] S. Yousefi, M. H. Goldbaum, M. Balasubramanian, T. P. Jung, R. N. Weinreb, F. A. Medeiros, L. M. Zangwill, J. M. Liebmann, C. A. Girkin, and C. Bowd, "Glaucoma progression detection using structural retinal nerve fiber layer measurements and functional visual field points", IEEE Trans. Biomedical Engineering, 61(4):1143–1154, 2014. https://doi.org/10.1109/TBME.2013.2295605

[263] S. Yousefi, M. H. Goldbaum, M. Balasubramanian, F. A. Medeiros, "Learning from data: recognizing glaucomatous defect patterns and detecting progression from visual field measurements", IEEE Trans. Biomedical Engineering, 61(7):2112–2124, 2014. https://doi.org/10.1109/TBME.2014.2314714

[264] S. Yousefi, M. H. Goldbaum, E. S. Varnousfaderani, A. Belghith, T. P. Jung, F. A. Medeiros, L. M. Zangwill, R. N. Weinreb, J. M. Liebmann, C. A. Girkin, and C. Bowd, "Detecting glaucomatous change in visual fields: Analysis with an optimization framework", J. Biomedical Informatics, 58:96–103, 2015. https://doi.org/10.1016/j.jbi.2015.09.019

[265] S. Yousefi, M. H. Goldbaum, L. M. Zangwill, F. A. Medeiros, C. Bowd, "Recognizing patterns of visual field loss using unsupervised machine learning", Paper presented at: SPIE Conf. Medical Imaging, San Diego, 2014. https://doi.org/10.1117/12.2043145

[266] S. Yousefi, T. Kiwaki, Y. Zheng, H. Suigara, R. Asaoka, H. Murata, H. Lemij, and K. Yamanishi, "Detection of longitudinal visual field progression in glaucoma using machine learning", American J. Ophthalmology, 193:71–79, 2018. https://doi.org/10.1016/j.ajo.2018.06.007

[267] M. Yousefikhoshbakht and M. Sedighpour, "New imperialist competitive algorithm to solve the travelling salesman problem," Int. J. Computer Mathematics, 90(7):1495–1505, 2013.

[268] T.-Y. Yu, R.J. Jacobs, N.S. Anstice, N. Paudel, J.E. Harding, and B. Thompson, "Global Motion Perception in 2-Year-Old Children: A Method for Psychophysical Assessment and Relationships With Clinical Measures of Visual Function", Invest Ophthalmol Vis Sci, 54(13):8408–8419, 2013.

[269] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks", Proc. European Conf. Computer Vision, pp. 818–833, 2014. https://doi.org/10.1007/978-3-319-10590-1_53

[270] X. Zhang and W. Chen, "A new chaotic algorithm for image encryption," Proc. Int. Conf. Audio Language Image Processing, 29:889–892, 2008.

[271] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey", ACM Computing Surveys, 35(4):399–458, 2003. https://doi.org/10.1145/954339.954342

[272] W. M. Zheng, "Kneading plane of the circle map", Chaos Solitons Fractals, 4(7):1221–1233, 1994.

[273] C. Zitnick and S. Kang, "Stereo for image-based rendering using image over-segmentation," Int. J. Computer Vision, 75:49–65, 2007.

[274] B. Zoph, V. Vasudevan, J. Shlens, Q. V. Le, "Learning transferable architectures for scalable image recognition", arXiv preprint arXiv:1707.07012, 2(6), pp. 8697–8710, 2018. https://doi.org/10.1109/CVPR.2018.00907

Websites

[275] AdaBoost classifier in Python, and machine learning in datacamp , see https://www.datacamp.com/community/tutorials/, access date: 27 July, 2019.

[276] A foundational machine learning framework used across Apple products. Core ML2, developer.apple.com/machine-learning/, access date: 27 August, 2018.

[277] A Comprehensive Guide To Glaucoma, normal versus glaucoma schematic figure www.lasereyesurgeryhub.co.uk/a-comprehensive-guide-to-glaucoma/ access date: 22 August, 2019.

[278] J. Bender and T. Wang, Tilted and anti-tilted portfolios: A coherent framework for advanced beta portfolio construction, papers.ssrn.com/sol3/papers.cfm?abstract_id=2757802, access date: 27 August, 2018.

[279] Blind Foundation of New Zealand, Latest stats at a glance, Annual Report 2015, www.blindfoundation.org.nz/eye-info/latest-statistics/, access date: 20 August, 2018.

[280] Boosting method, https://chistio.ir/ access date: 7 October, 2019.

[281] F. Chollet, Keras: Deep Learning for humans, GitHub, github.com/fchollet/keras, access date: 27 August, 2018.

[282] R. Collobert, K. Kavukcuoglu, and C. Farabet, Torch7: A matlab-like environment for machine learning, Proc. BigLearn, NIPS workshop, no. EPFL-CONF-192376, http://torch.ch/, 2011.

[283] CVFX Lecture 15: Stereo correspondence in 2014 https://www.youtube.com/watch?v=kxsvG4sSuvA access date: 27 September, 2019.

[284] Data on eye movements in people with glaucoma and peers with normal vision, https://doi.org/10.5281/zenodo.1156863, access date: 12 August, 2019.

[285] A. Dubey, "Feature Selection Using Regularisation", https://towardsdatascience.com/, access date: 10 April, 2020.

[286] Glaucoma in 2015, see www.iapb.org/knowledge/what-is-avoidable-blindness/glaucoma/, access date: 20 July, 2019.

[287] Glaucoma New Zealand, see www.glaucoma.org.nz, access date: 22 August, 2018.

[288] J. Heaton, "Applications of deep neural networks", see https://sites.wustl.edu/jeffheaton/t81-558/, access date: 10 April, 2020.

[289] InceptionResNet-v2 Network, 2017, see www.mathworks.com, access date: 17 September, 2018.

[290] M. Lin, "AI Part 2- Classifier Showdown", see https://medium.com/fragile-by-design/first-dive-into-machine-learning-and-ai-compe\-titive-data-science-part-2-c2d6c124f0ee, access date: 4 October, 2019.

[291] M. Mayank, Ten machine learning algorithms you should know to become a data scientist, see kdnuggets.com/2018/04/10-machine-learning-algorithms-data-scientist.html, access date: 24 October, 2018.

[292] Medical image analysis group, University of La Laguna, Spain. medimrg.webs.ull.es/research/retinal-imaging/rim-one/, access date: 20 August, 2018.

[293] Microsoft Cognitive Toolkit, microsoft.com/en-us/cognitive-toolkit/, access date: 27 August, 2018.

[294] A. Ng, https://www.coursera.org/lecture/machine-learning/, access date: 25 March, 2020.

[295] OCT imaging, see www.aao.org/eye-health, access date: 02 September, 2019.

[296] Pytorch. Tensors and dynamic neural networks in Python with strong GPU acceleration, pytorch.org/, access date: 27 August, 2018.

[297] A. Ray. 7 limitations of deep learning algorithms of AI, amitray.com/7-limitations-of-deep-learning-algorithms-of-ai/, access date: 6 November, 2018.

[298] A. Shaleynikov, 10 best frameworks and libraries for AI, dzone.com/articles/progressive-tools10-best-frameworks-and-libraries, access date: 27 August, 2018.

[299] SLIC superpixels 2017, see ivrl.epfl.ch/research/superpixels access date: 27 August, 2019.

[300] C. R. Souza, A. Kirillov, M. Diego Catalano and Accord. NET contributors. The Accord.NET Framework, Open Source Machine Learning Data Mining Framework, accord-framework.net, 4(8), 2014.

[301] "SP1: Real-time 3D stereo vision through FPGA technology, 2018," Nerian Vision Technologies, see nerian.com/products/sp1-stereo-vision/

[302] Stereo matching 2018, see www.devpy.me/what-is-stereo-matching/

[303] "Stereo vision for depth estimation 2018", The Mathworks, Inc., see www.mathworks.com/discovery/stereo-vision.html

[304] TensorFlow Lite is TensorFlow's lightweight solution for mobile and embedded devices, tensorflow.org/mobile/tflite/, access date: 27 August, 2018.

[305] The future of eye tracking. see pupil-labs.com/, access date: 01 August, 2019.

[306] S. Xu, "Deep Reinforcement Learning: Build a Deep Q-network(DQN) to Play CartPole with TensorFlow 2 and Gym," see https://towardsdatascience.com/ access date: 28 March, 2020.

[307] The Glaucoma Foundation. see glaucomafoundation.org, access date: 21 August, 2018.

[308] The latest news from Google AI in 2016, see ai.googleblog.com/2016/08/improving-inception-and-image.html, access date: 16 September, 2018.

[309] These notes accompany the Stanford CS class CS231n: Convolutional Neural Networks for Visual Recognition. see cs231n.github.io/convolutional-networks/#conv, access date: 15 December, 2018.

[310] The random forest classifier, see https://towardsdatascience.com/understanding-random-forest-58381e0602d2, access date: 29 September, 2019.

[311] Trainable classifier: Support Vector Machine, nu-algorithme, http://www.37steps.com/prhtml/prtools/nusvc.html, access date: 04 October, 2019.

[312] Transfer Learning by andrew ng; see https://www.youtube.com/watch?v=yofjFQddwHE access date: 12 October, 2019.

[313] VGG19 Network, 2017, see www.mathworks.com, access date: 17 September, 2018.

[314] Vision problems in the U.S., 2010, see www.visionproblemsus.org

[315] World Health Organization, see www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment, 2018. access date: 22 July, 2019.

[316] XGBoost developers snapshot, see https://buildmedia.readthedocs.org/media/pdf/xgboost/latest/xgboost.pdf access date: 10 October, 2019.