Automated Cataract Grading using Smartphone Images

by

Mona Nasirzonouzi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Vision Science and Systems Design Engineering

Waterloo, Ontario, Canada, 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

Supervisors:	Dr. Vasudevan Lakshminarayanan Professor, School of Optometry & Vision Science, University of Waterloo
	Dr. John Zelek Associate Professor, Systems Design Engineering Department, University of Waterloo
Committee Members:	Dr. Lyndon Jones Professor, School of Optometry & Vision Science, University of Waterloo
	Dr. Alexander Wong Associate Professor, Systems Design Engineering Department, University of Waterloo
	Dr. Kaamran Raahemifar Adjunct Professor, School of Optometry & Vision Science, University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

A cataract is an age-related eye disease and is one of the leading ophthalmological public health problems in developed and developing countries. Early detection of cataracts is necessary to preserve sight and prevent the increase in blindness due to cataracts worldwide. Lacking eye clinicians and slit lamp cameras in poor and rural areas are the main causes of the cataract's late diagnoses. The recent research in this field indicates that it is possible to screen cataracts using image processing. As smartphones become universal in most urban areas, cataract self-screening with smartphones removes the limitations like cataract screening cost and travel/time burdens for patients. Accordingly, a novel computer-aided automatic cataract grading method is presented in the current dissertation to detect various cataract stages, including normal, early, pre-mature, and mature cataracts, from the digital camera images. The IIITD Cataract Mobile Periocular (CMP) dataset was used as the cataractous and normal data images in the current study. This dataset contains periocular images, including ocular regions such as the eyebrow, pupil, sclera vasculature, iris, and pupil. These images are captured in the unconstrained condition such as uncontrolled illumination, complex background, and geometric distortions and mostly have non-frontal view poses. The current dissertation addresses smartphone-based cataract grading by proposing a method to classify the periocular eye regions into four classes of normal, early, pre-mature and mature cataracts on deep features using Convolutional Neural Networks (CNNs). We designed and proposed a four-layer CNN for cataract grading of the IIITD detected eye regions in the first procedure. In the second procedure, three pre-trained ConvNets, including VGG-16, Inception V3, and ResNet-101, were fine-tuned on the target dataset. In the last procedure, to evaluate the classification technique with the standard supervised classifiers, the extracted features by the ResNet-101 pre-trained network were fed into the Support Vector Machine (SVM) classifier for cataract grading. The experimental results show that end-to-end Residual Network (ResNet)-101 with the accuracy rate of 89.62 % outperforms the four-layer CNN, VGG-16, Inception V3, and ResNet-101+SVM with the mean accuracy of 84.67%, 87.64%, 84.67%, and 87.14% respectively. Moreover, according to all the calculated evaluation metrics such as precision, recall, sensitivity, specificity, and also F-measure, which is the trade-off between recall and precision, the results show that for each class, ResNet-101 outperforms the other models and has a better grading result for IIITD with the imbalanced number of images for each class.

Acknowledgements

I would like to express my sincere gratitudes to my supervisors Dr. Vasudevan Lakshminarayanan and Dr. John Zelek for their invaluable guidance, continues support and excellent supervision. Thank you for spending all those hours to guide and mentor me throughout this process. This thesis changed directions many times. Each time I would give up hope that anything good can come out of this. But you continuously helped put things into perspective and were really patient with me, my mistakes, and the numerous emails I would send you all the time. I want to thank you from the bottom of my heart for giving me this opportunity where I learned so much in every step. This would not have been easy if it was not for amazing supervisors like you.

I would like to express my special thank to my committee members Dr. Lyndon Jones, Dr. Alexander Wong, and Dr. Kaamran Raahemifar for their valued knowledge, advice and time invested in guiding me throughout my degree.

I would also like to thank Abdul Rasheed Mohammed for spending the time to label my dataset as a well-trained optometrist and a supportive friend.

Last but not least, I want to genuinely thank my parents Ali and Minou, my brother Pouya and my friends Mohammad, Sevil, Ehsan and Parvaneh who were my family here in Waterloo, Canada. I would like to thank for their constant words of encouragement, wisdom and endless love. Most importantly thank you for being there for me. You have each supported me in so many ways that I cannot be appreciative enough for.

Dedication

To my beloved parents and brother Ali, Minou, Pouya.

Table of Contents

List of Tables

List of Figures

Abbreviations

AI artificial intelligence 5

ANN Artificial Neural Network 14, 17

AREDS Age-Related Eye Disease Study 5

AUC Area Under the ROC Curve 16

CC Cortical Cataract 2, 3, 15

CCD Charged Coupled Device 11

CMP Cataract Mobile Periocular 21

CNN Convolutional Neural Network xii, 8, 9, 18, 20, 22, 24, 26, 31, 37, 38, 42, 43, 45–47, 58, 59

EC Early Cataract xii, 27, 43

FC Fully Connected 40

FPR False Positive Rate 13, 38

GLCM Gray level Co-occurrence Matrix 14

HOG Histogram of Oriented Gradients 24

ILSVRC Large Scale Visual Recognition Challenge 58, 59

IOL Intra-ocular lens 3, 27

IOP Intraocular Pressure 3

K-NN K-Nearest Neighbor 14, 15

LOCS Lens Opacities Classification System 5

lr learning rate 39

MC Mature Cataract xii, 27, 43

MTCNN Multi-Task Cascaded Convolutional Neural Networks 34

NC No View of Cataract xii, 15, 27, 43

NO nuclear opalescence 5

NUC Nuclear Cataract 2, 3, 10

OXCGS Oxford Clinical Cataract Grading System 5

PMC Pre-Mature Cataract xii, 27, 43

PSC Posterior Subcapsular Cataract 2, 3

R-CNN Region-based Convolutional Neural Networks 8, 11, 12

RBF Radial Basis Function 32

ResNet Residual Network iv, xii, 8, 12, 18, 20, 30–32, 38–40, 42, 43, 45, 46, 58, 59

ROC Receiver Operating Characteristic curve xii, 16, 37–40, 43

ROI Region of Interest 11, 12, 14

RPN Region Proposal Network 12

SEC Smart Eye Camera 10

SGD Stochastic Gradient Descent 39

SVM Support Vector Machine xii, 8, 15, 17, 20, 32, 40, 42, 43, 45, 46

SVR Support Vector Regression 8

TIFF Tag Image File Format 13–15

TPR True Positive Rate 13, 38

VGG Visual Geometry Group xii, 18, 20, 30, 31, 38, 39, 42, 43, 45, 58, 59

WGS Wisconsin Grading System 5

WHO World Health Organization 5, 10

Chapter 1

Introduction

1.1 Background

The human eye is a complicated system consisting of interconnected organs, including the lens, pupil, iris, retina, cornea, and optic nerve [6]. There are several ocular diseases related to different components of the eye; age-related diseases such as cataracts are among the most common ones [7, 6]. If the ocular diseases are diagnosed late, it is challenging to repair vision effectively and lead to vision loss [6]. Although cataract can be cured [7], it remains one of the main problems in ophthalmological public health in developed and developing countries [8, 9, 10, 11, 12, 13, 14], and it is known as the leading cause of blindness in most countries [14, 6, 2, 15]. Studies show that 36 million people worldwide have blindness, and more than 12 million cases are diagnosed with cataract [10, 13]. It is estimated that this number will increase to 13.5 million people in 2020 [13, 14]. In 2015, about 3 million cataract surgeries were performed in the United States alone, with an estimated 6.8 billion USD in direct costs [7, 9]. Given the magnitude of both the number of individuals affected and the associated healthcare costs, assessing both the presence and severity of cataracts are imperative for diagnosing and monitoring the disorder's progression [9, 10, 11, 12, 13].

1.1.1 Cataract

The human eye's crystalline lens is an optically clear organ with ectodermal tissue, located between the iris and vitreous body and retina [16]. Due to the refractive index of the crystalline lens, the shape, and its clarity, the crystalline lens can focus the incident light on the

retina [17]. Besides the superficial strips of new cells, the lens's constant growth produces a series of laminae that are concentrically arranged and gradually increase lens fibers in life [16]. As a result, the existing crystalline proteins in the lens misfold and aggregate into insoluble clumps in aged people [18]. When the lens loses its optical clarity, the result would be a complication known as cataract [19]. Cataracts occlude the transmission of light to the retina, thereby impairing the vision, even causing blindness [20]. Figure 1.1 indicates the anatomy and structure of the adult human lens.

The lens has three layers, including the nucleus, cortex, and capsule [21]; the nucleus is the core in the lens surrounded by the cortex and capsule, respectively [2]. According to the locations of the grown opacity, cataracts are classified into three types namely Posterior Subcapsular Cataract (PSC), Cortical Cataract (CC), and Nuclear Cataract (NUC) [22] which might occur either alone or in combination with each other (Figure 1.2). Resultant changes are typically bilateral but commonly asymmetrical [19]. Cortical cataract CC is a radial, and white wedged-shaped opacity in the cortex [6]. It begins from the lens' outside edge and moves towards the center in a spoke-like manner [22]. Posterior subcapsular cataract (PSC) appears in the form of small sand-like particles sprinkled near the back of the lens and is more common in diabetic patients [22, 6]. Figure 1.2 shows three types of cataracts: nuclear, cortical, and posterior subcapsular cataracts.

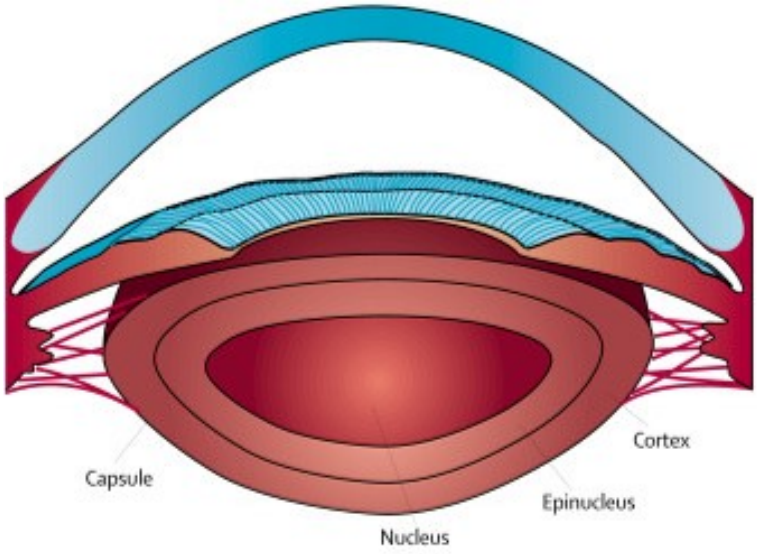

Figure 1.1: The human lens subsections[2].

Although cataracts can be cured [7], still, it is considered one of the most prevalent reasons for visual impairment worldwide [23, 8, 9, 14]. According to a systematic study conducted in February 2020, the overall prevalence of cataracts in women and men were

33.67% (95% CI: 25.90–41.44) and 32.57% (95% CI: 26.29–38.85) respectively; in addition, prevalence of different types of cataracts in women and men was as followed: cortical cataract CC: 15.22% (95% CI: 9.79–20.65) and 13.64% (95% CI: 9.17–18.11); nuclear cataract NUC: 14.09% (95% CI: 9.67–18.51) and 15.63% (95% CI: 11.44–20.33); and PSC cataract: 3.66% (95% CI: 3.34–4.98) and 3.70% (95% CI: 2.35–5.05) [14]. According to the reports, more than 90% of vision lost because of cataracts are observed in developing countries[10]. Therefore, it is significantly important to take measurement towards the early diagnosis of cataracts and early prevention of blindness due to cataracts.

Aging is one of the major causes of cataracts [8]. In general, the prevalence of cataracts varies in different age groups, and most diagnosed cases are over 60 [14]. However, genetic and environmental factors like exposure to ultraviolet light, diseases (e.g., diabetes and uveitis), smoking, specific jobs, Intraocular Pressure (IOP)-lowering medications/surgery, trauma, and steroids, may increase the risk of cataracts [24, 25, 26, 27, 14]. So far, no effective method has been developed against the formation of a cataractous lens. However, cataracts' removal through small-incision surgery, viscoelastic use, and development of Intra-ocular lens (IOL) have all affected the quality and time of treatment and visual recovery. Despite that, cataracts are still considered a significant public health problem that will worsen along with population increase worldwide.

In nuclear cataract, which is the most prevalent age-related cataract [28], new fiber layers are gradually added to the lens, resulting in compression, and hardening nucleus with a yellow lens [29]. Therefore, the changes caused by the age-related nuclear cataract consists of two processes: 1- opacification (clouding) and 2- coloration (browning) [15]. Nuclear cataract advances gradually over the years and does not significantly affect the vision in some cases. However, only a change in refraction (myopic shift) or second sight may cause patients not to use glasses for reading anymore [6]. Nevertheless, the further advance of nuclear cataract might cause color differentiation and vision loss, particularly distance vision [30].

Surgery is currently the primary method for cataract treatment [31]. In cataracts surgery, the ophthalmologist takes out the cataractous lens and replaces it with a clear artificial lens. However, sometimes, cataracts may be corrected without implanting an artificial lens called an intraocular lens (IOL). Surgical methods for correcting cataracts include using an ultrasound probe to break up the lens and removing out. The phacoemulsification ultrasound probe emits energy into the eye lens, breaks up cataracts, and facilitates emulsification and aspiration. The annual number of cataracts surgeries in the USA and Africa are currently 5000 and 200 cased per million [32]. In the late 90s, 1.35 million cataracts operations were conducted annually in the USA with a cost of 3.4 billion USD [33]. Expenses associated with non-treatment cataracts, which cause non-functional vision, are remarkably higher than the costs of cataracts surgery and treatment [34]. Therefore, addressing global access to high-quality cataracts surgery is of primary importance.

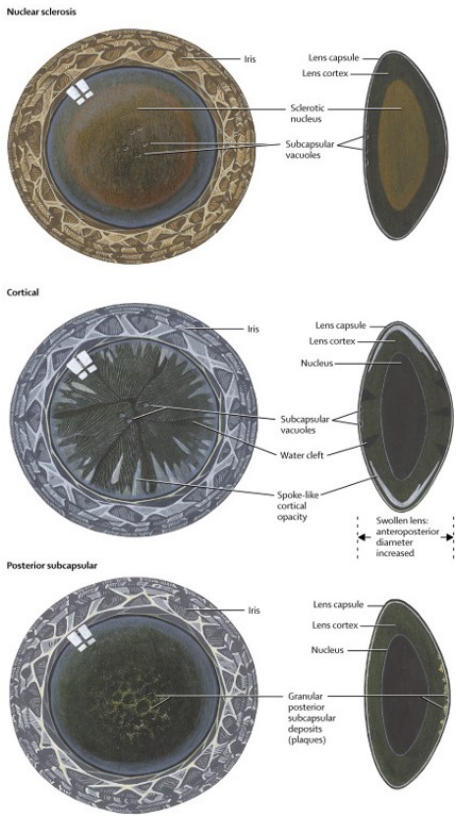

Figure 1.2: The Three Types of Cataract: Nuclear, Cortical and Posterior Subcapsular cataract[2]

Mass screening of cataracts among elderlies is essential from both a social and economic point of view [34].

1.2 Motivation

Cataracts are generally diagnosed by an eye care clinician who observes anatomical changes occurring in the eye lens by a slit-lamp [35]. The clinician then detects and grades cataracts to determine their type and severity level in order to be able to plan the necessary treatments [6]. Classification and grading of cataracts, especially in their progressive age-related forms, has been frequently addressed in clinical and epidemiological research [2, 28]. Cataracts are clinically diagnosed by well-trained eye care clinicians using a slit-lamp biomicroscopy and manually graded by comparing the opacity severity of each slit-lamp image

with a set of standard images [6]. Standard images show increasing severity of cataract indicated by increasing integer-valued grades [36]. Ungraded slit-lamp image is then matched up with standard images, and a/an decimal/integer-valued grade is specified that shows opacity's severity [37, 36]. Common standard grading protocols include Lens Opacities Classification System (LOCS) III, Age-Related Eye Disease Study (AREDS) cataract grading protocol, Wisconsin Grading System (WGS), Oxford Clinical Cataract Grading System (OXCGS) [38] and World Health Organization (WHO) Grading System. LOCS-III and Wisconsin grading systems are broadly employed by eye care clinicians [6]. The LOCS III system assesses four features, including nuclear color (NC), nuclear opalescence (NO), posterior subcapsular cataract (P), cortical cataract (C) [37].

This diagnostic procedure requires expensive medical equipment; besides, manual procedures of diagnosis are time-consuming and subjective according to clinician's experiences [38]. It is reported that when different eye care clinicians are asked to grade the same slit-lamp images according to the same grading system, only about 65% inter-observer agreement is reached [37]. Furthermore, eye clinicians are prone to unconscious and imprecise grading after inspecting numerous images [29]. It is worth mentioning that the correct screening of cataract hardness is crucial to reduce the surgical complications [15]. Selecting an incorrect phacoemulsification energy level in cataract surgery may disturb the posterior capsule [39]. Thus, it is essential to ascertain the optimal phacoemulsification energy level, which relies on the cataract's density and grading [39].

Although it takes a well-trained eye clinician to grade a cataract only a few minutes, recent developments in imaging technologies have prompted researchers to propose automatic systems with the ability to locate the lens landmarks and offer cataract grading scores. These landmarks are correctly correlated with LOCS III clinical grades and decrease testing time [40]. The increasing prevalence of artificial intelligence (AI) in ophthalmology is powered by ever-growing clinical big data [41]. However, AI 's experience and development are still limited in cataracts, compared to other eye diseases like diabetic retinopathy, age-related macular degeneration, and glaucoma [41]. Previous works used algorithms for automated cataract determination using various modalities such as slit lamp or color fundus, and, more recently, they are towards cataract grading using smartphone-based images [42, 43, 44, 45].

A device with a featured camera and remarkably bright and high-resolution screens can be potentially used for ophthalmology, and eye care [46]. There are billions of smartphone users worldwide, which causes revolutionary potential in healthcare applications. Smartphones are providing low-cost and automated alternatives for expensive medical diagnostics, especially in regions where it is difficult to access medical professionals and equipment. Although the slit-lamp is considered a standard protocol for cataract detection and grading, using digital camera images for screening cataract is more desirable due to simple and easy use. Considering all these reasons, the current research study seeks to investigate automated cataract grading using mobile phone images.

1.3 Objective

A novel computer-aided automatic cataract grading method is presented in the current study to detect various cataract stages, including normal, early, pre-mature, and mature cataracts stages using digital camera images. The IIITD dataset [1] with 2380 images is utilized. This dataset is captured in two pre-surgery (145 subjects) and post-surgery (99 subjects) sessions. The main purpose of the proposed method is to implement an automated end-to-end cataract grading system. According to the proposed method, the eye region will be detected automatically, and the extracted eye regions will feed into a CNN model as inputs. The CNN model categorizes the ungraded lenses into normal, early, pre-mature and mature groups.

Chapter 2

Related works

2.1 Overview

This chapter provides an overview of the existing literature on cataract grading methods and techniques and presents recent developments and state-of-the-art smartphone-based decision support systems in computer-aided cataracts screening and grading. The current chapter is then organized into several sections; first, cataract detection and grading methodologies using smartphone images are reviewed. Cataract grading consists of several steps in which each possible method and technique that have been so far applied are analyzed. Then, the shortcomings and advantages of these techniques and methods are discussed and some suggestions are introduced to improve the existing methods. Finally, proposed approach of the current dissertation is presented.

2.2 Cataract grading systems

Imaging modalities currently used for developing automated techniques for detection and grading of cataracts include slit-lamp, retro-illumination, retinal, digital/optical eye, and ultrasonic Nakagami images [6].

Slit-lamp images are mostly employed to detect nuclear cataracts, which normally affect the eye lens's nucleus; for that reason, it is automatically detected and graded by extracting features of the nucleus region [6]. Several studies have been performed in this regard.

Huang et al. in [47, 36] utilized image ranking of neighbor markers and worked on the optimization of the learning functions to predict the severity of cataracts by slit-lamp images. The proposed method (Learned Ranking function (LRF)) achieved better and less ranking errors among all other learning to rank methods (proposed method = 2 errors; RankBoost = 16 errors; AdaRank = 9 errors; RankingSVM—4 errors) [47, 36] . Li et al. in [48] utilized human crystalline lens' color as the critical feature for nuclear cataracts diagnosis, achieved 95% accuracy rate for feature extraction, and 0.36 mean errors for nuclear cataract grading. Besides, Fan et al. in [49] presented an algorithm for the classification of cataracts stages by extracting hand-crafted features based on the intensities of the landmarks of the visual axis in the human lens and then performing nuclear cataract grading for images taken by a slit-lamp. The basic rule was extracting global/local features of the human eye lens and feeding them into SVM or Support Vector Regression (SVR) to conduct the classification task [50]. The accuracy was up to 90%. However, in the years later, the feature extraction algorithms moved from traditional methods to CNN-based methods. X. Liu et al.[51], used the CNN-based method for extracting the features of pediatric slit-lamp images. They used SVM algorithm for automatic cataract classification with a mean accuracy of 97.07%, the sensitivity of 97.28% and specificity of 96.83%. The suggested deep learning approach was approved to be more effective than conventional methods (with the accuracy rate up to 90%) [52]. Xu et al. [15] proposed a nuclear cataract grading method; they employed Faster Region-based Convolutional Neural Networks (R-CNN) for locating the nuclear region and considered it as an input for a classifier based on ResNet-101 network. Although the automated cataract classification method is constantly proposed, its accuracy needs to be improved. Furthermore, the lack of eye clinicians and slit lamp cameras in rural regions, particularly in a developing country, are considered limitations of the cataracts diagnosis process [6]. Therefore, researchers have developed application systems based on digital image processing techniques using smartphone cameras that assist in the early detection of cataracts.

2.3 Cataract grading using smartphone

Since many people in most urban areas are frequently using smartphones, smartphones can simplify the cataracts self-diagnosing, which is associated with low expenses and is less time-consuming. [43]. More than 100,000 therapeutic and medical applications can be currently installed and used in smartphones with several other external devices like an external attachable device that simulates the slit-lamp in the clinics [52]. In addition to simplifying the cataracts screening process such applications, reduce the misdiagnosis rate and improve the treatment accessibility [52].

Smartphone-based cataracts screening systems are categorized into two groups: 1) Using smartphone images and attaching an external device. 2) Using a smartphone image

and the photo-taking function of the smartphone camera. In the first group, a micro-lens, a portable slit-lamp, is designed and attached to the smartphone's camera [52]. The eye lens's images replace the desktop slit-lamp observation system, which helps people in distant and deprived areas with no medical equipment in the early screening of cataract diseases. It also solves the problem of the shortage of specialized clinicians in those areas.

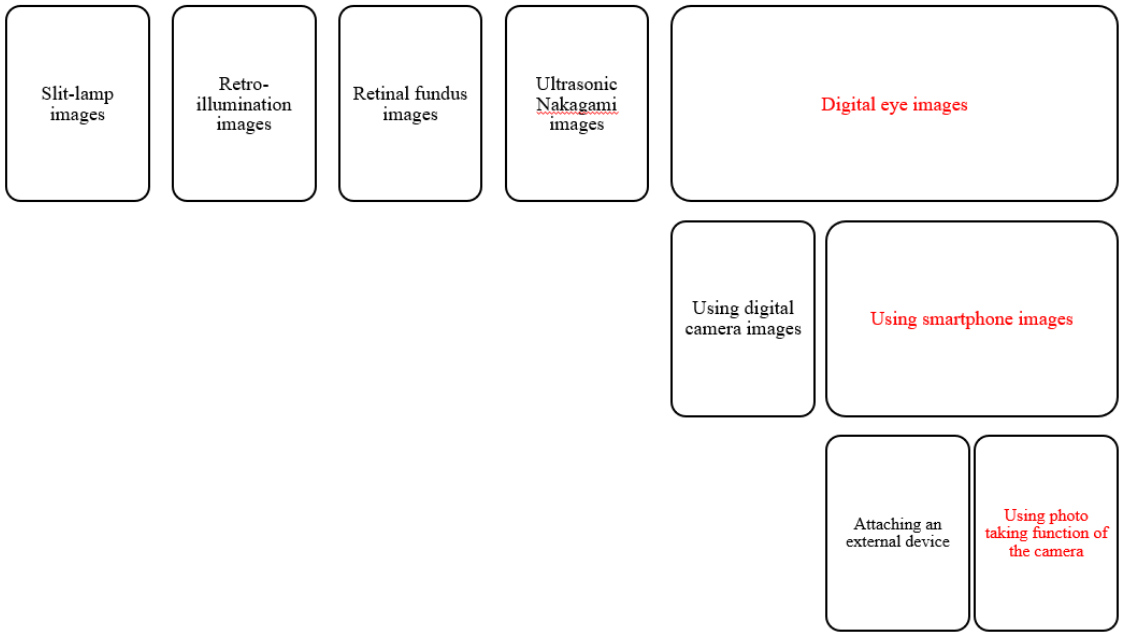

Figure 2.1: Automated cataracts grading methods using various imaging modalities. The modality and the procedures written in red shows the chosen method in the current study.

Peterson et al., in [42], used an iPhone X camera for cataracts grading of 50 subjects and attached an external device to the camera as the flashlight with auto-focus capability and maximum resolution. The eye region images were recorded and evaluated to measure luminance reflection and color features of the lens combined with a CNN as a classifier for cataracts grading. Accuracy, specificity (the true negative rate), and sensitivity (the true positive rate) of the method in diagnosing the affected eyes and distinguishing them from healthy eyes were respectively 98.2%, 97.8%, and 97.2%. In [43], a micro-lens was attached to a smartphone to simulate a slit-lamp. Images taken by the smartphone were fed into a deep learning system, which resulted in real-time and effective screenings. The obtained images included photos of cataracts and normal eyes. Also, the anterior/posterior capsules and cortex of the eye's lens were straightforward to differentiate in these images [43]. Unlike the slit-lamp, which requires professional skills for operation, this approach can be used by less-skilled or inexperienced people. In another study, a portable and recordable

slit-lamp device that could be attached to the smartphone, known as Smart Eye Camera (SEC), was invented and presented [53]. In this study, both SEC and conventional/non-portable slit-lamp microscope were used to assess NUC, and the results were compared. A total number of 64 Japanese patients (mean age: 73.95 ± 9.28 years; range: 51–92 years; female: 34) were examined. According to WHO cataract grading system, NUC is categorized into four grades (grade 0 to 3) based on three standard photos of nuclear opacities. No new algorithm was used in this study, and a single ophthalmologist performed the grading. Results revealed that NUC grading using the approaches mentioned above correlated significantly; in other words, SEC was too similar to the conventional non-portable slit-lamp microscope for NUC evaluation in terms of reliability. However, there are some limitations and drawbacks; the micro-lens should always be attached to smartphones, it is too expensive for patients, and the patients need to be taught to use the attached slit-lamp device. A review of mobile applications for cataracts detection based on image processing techniques is presented in [52].

In the second group, smartphone images are directly captured, and no additional equipment is required. This approach solves the problem of accessing micro-lens slit-lamp, which can be quite challenging and impractical in rural areas. Therefore, if the second group can provide precise and significantly better accuracy, it may replace the portable slit-lamp device.

Using a smartphone for cataract grading requires several steps, including data acquisition, image preparation, and eye region detection. In this process, the eye region is considered the region of interest, and useful features must be extracted to classify the input image into normal and various cataracts stages: early cataracts, pre-mature and mature cataracts. The following section reviews recent works on cataract grading using smartphone images without any additional device. After presenting each approach's advantages and disadvantages, some suggestions are made to improve the previously available algorithms.

2.3.1 A review on eye region detection methods

Due to the development of many potential applications such as biometrics, iris recognition, eye tracking, and diagnosis of ocular diseases, automatic human eye detection has received particular attention during the past decades. A review of recent eye detection methods is presented in [54].

Zhu et al. [55] classified existing eye detection techniques into two categories: active infrared (IR)-based methods [55, 56, 57, 58] and traditional image-based passive techniques [54]. The first approach is based on the red-eye effect in flash photographs and uses

an IR-sensitive Charged Coupled Device (CCD) for image acquisition special IR and an illuminator [54]. The most important drawback, however, is that lighting conditions should be stable and the camera must be set close to the subject [55].

Some special features of eyes, including color distribution, intensity, appearance, and shape, are often used in image-based passive eye detection methods[54]. Template matching [59, 60], eigenspace [61], and Hough transform-based methods are among the most common ones in this category. Template matching technique compares different segments of an input image with those of the template utilizing correlation values to assess the similarity [54]. Despite that, this approach is not very robust because it cannot deal with eye variations in rotation, scale, illumination, and expression. In their study, a deformable template for face feature extraction is introduced by Yuille et al. [62]. Accordingly, a parameterized template of an eye was presented and an energy function was developed to connect the peaks, edges, and valleys of input image to corresponding features of template [54]. Besides location, this method could simultaneously detect more eye features, like its shape and size. However, this approach is time-consuming, and its success rate is subject to change according to the template's primary position. Pentland et al. [61] used an eigenspace method and used training data that covered eye variations in orientation, appearance, and lighting conditions. However, the performance of this technique is highly dependent on the selected training set. Hough transform method is another widely employed technique that deals with the binary valley or edge maps and is based on the shape of the iris [63]. This method's problem is that the performance relies on the threshold values chosen for binarization of valley or edge maps. Although much effort has been made, automatic eye detection still has some limitations and drawbacks which need to be addressed. There are many factors, including face rotation in-plane and depth, facial expression, lighting conditions, and occlusion, that could certainly impact the performance of eye detection algorithms [64]. However, most of the existing eye detection approaches are basically concentrated on face images with a frontal view.

While automation of the cataract grading process is highly required, there are still some challenges: the reliable localization of the eye region in a relatively complex background is one of the challenges. Several studies cropped the eye regions manually [42]. Besides, previous works like [65, 36, 66] presented a new technique and approximately localized a Region of Interest (ROI) in slit-lamp images by thresholding 20% to 30% of the brightest pixels in a gray-scale image. However, the background noise could not be avoided with such an intensity-based binarization [15]. In their study, Fan et al. [49] used a whole image and acknowledged that the nuclear region covered less than 3% of the pixels in slit-lamp photos and digital images; thus, grading based on the whole image is not reliable enough.

Addressing the ROI challenge in nuclear cataract grading, for the first time, Xu et al. proposed a technique based on deep learning. They localized the nuclear region by Faster R-CNN [67] which, according to Ren et al. [67, 15], contains two subnetworks including

Region Proposal Network (RPN) accompanied by a detection network. The former proposes several bounding boxes with the highest likelihood to contain ROIs. Moreover, the detection network is responsible for distinguishing foreground from background and processing the predicted location and size of ROIs. As a slit-lamp photo contains only one nuclear region, ROI with the highest rank can be selected as x_{nuc}.

It is worth mentioning that Faster R-CNN could not directly be utilized for cataract grading of the detected nuclear regions [15]. Consequently, a grading model based on ResNet-101 [68] was applied and received the nuclear region as input. Unlike previous studies, ResNet-101 resulted in a unified framework of feature extraction and grading phases. Therefore, the proposed solution was more functional and computationally more efficient [15].

In a recent study in smartphone-based semi slit-lamp images [52], the nuclear region of the lens was detected by YOLOv3, which is the latest variant of an object detection algorithm. Compared with the previous state-of-the-art method, Faster-RCNN (accuracy of 50.13%), YOLOv3, with the accuracy of 52.36%, could effectively detect the ROIs in the captured images and simplify the cataract grading process. Therefore, difficulties in examination and diagnosis of cataracts were remarkably decreased by optimizing the framework.

2.3.2 Feature Extraction Methods in the previous works

After successfully extracting the ROI in the images, another challenge still exists; how to obtain a vectorized description of ROI based on which a grading (or regression) model can be constructed [15]? According to the available dataset, two types of images are used in cataract grading; smartphones take the first types, and the second group includes images captured by digital cameras.

Methods using hand-crafted features

The problem with the images captured by a compact camera is that they do not have good quality and illumination. For instance, eye images taken in bright environments have high-intensity values inside the pupil, making the analysis of color information quite difficult and challenging. Three features were normally used in previous methods, including texture uniformity, specular reflection appearance, and average intensity inside the pupil [69, 70, 71, 72]. Supriyanti attempted to address the existing cataract diagnosis problems and studied these three features [69, 70, 71, 72].

Using specular reflection feature could solve the illumination and low-quality image problems. Supriyanti et al. [69] used the obtained data on specular reflection to establish a connection between severe and non-severe cataract conditions. They applied the specular reflection feature only in 75 images and found that normal eye images showed two types of pupil reflections in the eye. In contrast, images of the eyes with cataract showed only one reflection that was always coaxial. This finding helped them distinguish cataractous and normal eye images more efficiently. Specular reflection always seems to be brighter than its surrounding area, independent of illumination conditions. However, to prove the validity of this technique, it has to be applied on a larger scale, and other factors, including camera focus, angle, and distance, should be taken into account for a robust screening system. The best camera angle would be 50° - 70°, and the best distance from the object would be 30 cm - 60 cm [69]. Furthermore, an advanced algorithm is required to calculate the pupil and front-side reflection localization.

In her next study, Supriyanti also added a texture analysis [70, 71] to identify more characteristics of severe and non-severe conditions and to expand the performance of the cataract screening system. She also exploited the uniformity and average intensity. Whitish colors inside the lens are distributed in two ways: smoothly and unevenly. A thin layer of whitish color is observed in a first manner that gradually covers the whole lens surface till it becomes thick. When all gray levels are equal, uniformity is maximized. Approximately all non-severe conditions are smoothly textured and highly uniformed. Supriyanti explains that the average intensity is measured inside the pupil and is obtained by summing up the pupil region's gray levels and dividing the result by the total number of pixels in the pupil. Eyes with cataracts have brighter intensities than normal eyes. Since she could not effectively process small or moderate samples in her previous study, Supriyanti used a larger database this time. She used specular reflection and texture analysis features together and found that they were promising in cataracts screening. Supriyanti proposed an algorithm with a True Positive Rate (TPR) equal to 92% and a False Positive Rate (FPR) equal to 18%.

Although Supriyanti et al. considered only the circular shape of specular reflections in most of their studies, in a study conducted in 2011, [73], different shapes of specular reflections were focused for cataracts screening. They found that specular reflection could be in different shapes, including ellipse, circle, cube, or rectangle; however, using an oval shape is recommended to achieve the best results.

Anayet et al. [74] extracted colors of each block of cataract images by extracting the average of their R, G, and B values. They used these extracted features to classify the eye images into normal eyes plus grades 1, 2, 3, 4, and 5 of cataract. Nayak and Jagadish [75] also used Tag Image File Format (TIFF) optical images of pupils to divide them into normal, cataractous, and post-cataract groups. According to their method, the pupil and cornea areas were detected using an edge detection method known as Canny.

White pixels in every image were then counted, and the cataract perimeter was detected using erosion. Finally, the images were classified with an average accuracy of 88.39%. In their study, Fuadah et al. cropped the pupil areas manually and converted the extracted regions into grayscale images. They used Gray level Co-occurrence Matrix (GLCM) for feature extraction to distinguish between normal and cataract images. In their study on early diagnosis and grading of cataract, Tawfik et al. [44] combined the wavelet transform method with 2D Log Gabor Wavelet transform. This method was state-of-the-art for a while and obtained a high success rate of 96.8% due to Log Gabor and wavelets' strength in detecting features.

2.3.3 Classification in the previous methods

After extracting good features of the eye as the ROI, the last step is distinguishing normal lenses from cataractous ones. For this purpose, input eye images need to be classified based on their cataracts stages. Some classification techniques used in the existing literature about cataract grading are presented and discussed in the following section.

Various cataract detection and classification systems have been designed and implemented by many researchers in the field. Neural network classifiers were developed as a diagnostic tool to help ophthalmologists detect and grade cataracts. For instance, Artificial Neural Network (ANN) is a supervised learning algorithm consisting of layers that include a number of neurons connected with an activation function. In this algorithm, data are fed to the network through an input layer, which passes them to one or more hidden layers where weighted connections will process them. Outputs are then sent to the output layer through hidden layers. Some advantages of ANN include nonlinear performance, working consistency even when an element of the neural network fails, implementation in any application, and its design that can fit almost any type of data or problems. Acharya [76] used ANN classifier to classify TIFF natural eye images based on different eye diseases, including cataracts, corneal haze, corneal arcus, and normal eye with an accuracy of 90%. In another study, Acharya [77] also used ANN to classify pupils' images based on their extracted features by fuzzy k-means into three classes, including cataracts, post-cataracts, and normal images with an accuracy of 90%. In [44], Tawfik et al. used ANN (with an accuracy of 92.3 %) to differentiate between normal, early, and advanced stages of cataracts using a dataset of 120 eye images divided into training (78) and testing (42) sets. In their study, [44], ANN used 70 neurons in the hidden layer.

K-Nearest Neighbor (K-NN) is another classification technique that performs based on previously classified data. K-NN was presented by Fix and Hodges (1951) and is commonly used due to its simplicity and low computation time [78]. This classification method works by searching the nearest distance between testing and training data. Training data

are reflected in a multidimensional space where each dimension indicates their extracted features. Accordingly, that space is divided into different classes (normal and cataractous) based on the classification of training data [78]. During the classification process, feature extraction results related to testing data are shown as some vectors in multidimensional space; the similarity of test and training data is then calculated by spotting the nearest distance using euclidean distance. K value in the K-NN technique represents the nearest distance of testing and training data that will be utilized for classification and cataract grading. The k-value that gives the highest accuracy performance is k=1. Fuadah et al. [79] used K-NN to classify the pupil areas into cataract or normal areas with an accuracy of 94.5%. In a study on smartphone-based cataract grading in 2015, Yunendah Nur Fuadah [78] used K-NN to classify the proposed optimal combination candidate of statistical texture features and performed cataract grading with the highest accuracy of 97.5%. This approach is currently considered as the state-of-the-art in smartphone-based cataract grading systems.

SVM is a machine learning algorithm that can successfully overcome any training error [80]. With a polynomial kernel function, SVM can achieve effective results with a small training sample, compared to other kernel functions [81]. Its learning output is robust, and its prediction accuracy is high [81]. A polynomial kernel detects similar training samples in a feature space over polynomials of the original variables and allows the learning of non-linear models[44]. Nayak et al. [75] used TIFF optical images of pupils and SVM classifier to classify the eye images into normal, cataracts, and post-cataracts with an average accuracy of 88.39%. In their study [44], Tawfik et al. used SVM in order to classify a dataset of digital images (120 images) of iris into normal, early, and advanced stages of cataracts with a success rate of 96.8%.

In [74], Anayet et al. classified digital images of healthy and cataractous (NC and CC) eyes into normal eyes plus grades 1, 2, 3, 4, and 5 of cataract using K-means algorithm with an accuracy of 92.5%. U. Patwari et al. detected, categorized, and assessed eye cataract using digital image processing with an accuracy of 94.96%. In another study [82], researchers used digital images of healthy and cataractous eyes (taken by microscope during surgery) and removed the noise of pupil images and then converted them into binary images for edge detection. To determine the cataract type, they measured the circularity of cataracts and compared them with a circularity threshold of nuclear cataract (NC) binary shape or with a cortical cataract (CC) binary shape. As a result, cataracts were classified into two classes; nuclear cataracts with 94.96% accuracy and cortical cataract with 95.14% accuracy.

2.3.4 Evaluation metrics in the previous methods

Accuracy, precision, recall, specificity, F-measure value are metrics that are utilized to assess the performance of proposed algorithms in the cataract grading literature [52].

All these metrics are calculated according to following equations:

$$Accuracy = \frac{(TP + TN)}{(TP + FN + TN + FP)} \tag{1}$$

$$Specificity = TNR = \frac{TN}{FP + TN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Sensitivity = TPR = Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5}$$

TP, FP, TN, and FN respectively indicate the number of true positives, false positives, true negatives, and false negatives in grading results.

To assess the predictability of a model, Receiver Operating Characteristic ROC curves and Area Under the ROC Curve (AUC) can be used [43]. The ROC curve indicates the True Positive Rate (TPR) against the False Positive Rate (FPR) according to various threshold settings [43]. The TPR is similar as sensitivity, recall, or probability of detection [43]. The FPR is similar to the probability of false alarm and is equal to (1 specificity) [43]. Larger AUC shows better predictability, which is measured by sensitivity and specificity of test datasets.

2.4 Conclusion

Some several challenges and factors play essential roles in choosing a suitable method for cataract grading. These challenges and factors are the following.

Different types of modalities can help eye care clinicians to grade cataracts. Type of modalities determines the characteristics of databases, which might vary due to some

factors, including illumination, angle, the orientation of captured images, and facial pose, particularly for smartphone-based images. The difference between all types of datasets utilized in the literature is how the eye region is captured. Do we have the other parts of the face, or just the eye region is captured? Therefore, according to the selected dataset and the number of images, different grading and classification techniques may be used. All of these factors affect the final decision. Moreover, the lack of a benchmark dataset in smartphone-based cataract grading studies should be highlighted.

Unlike ANN, SVM can process small training samples and datasets effectively [80, 75, 44]. In a recent study in which both SVM and ANN approaches were used for the classification of 120 eye images into normal, early, and advanced cataracts [44], the former showed a success rate of 96.8%, while ANN success rate result was 92.3%. Nevertheless, SVM technique needs to be tested on a larger dataset.

A comprehensive review of recent literature on cataract grading systems showed the application of a device that simulates slit-lamp in the clinics. This device attaches to the smartphone camera for capturing images directly from the lens [52, 43]. Although this device captures good-quality images too similar to slit-lamp, patients' access to such equipment, especially in rural regions, can be a big challenge. Therefore, if smartphone-based methods that utilize only the smartphone's photo-taking functionality could offer more precise and accurate screening capacity, they can replace the attachable slit-lamp device. This reason was a motivation for the current study.

Based on the literature review for cataract grading using smartphone images, feature engineering methods were utilized to extract the features previously. The previous methods were mostly based on hand-crafted feature extraction methods. These methods were based on the images of the pupil rather than periocular images in the IIITD dataset. Moreover, the number of images was small, around 150 – 120 images. In [75], Nayak used edge detection methods to extract the pupil's center for cataractous images and measured the area of the extracted region. Other hand-crafted features such as Big Ring Area (BRA), Small Ring Area (SRA), Edge Pixel Count (EPC), and Object Perimeter were extracted and fed as the input to SVM classifier for classification and cataract grading. The classification rate was nearly 90%. In most of the previous methods, only two-class classification is addressed so far. Therefore, cataracts and various stages of cataracts need to be considered and investigated[6].

Traditional machine learning methods try to learn and build a model for each task from scratches. The problem is that it is expensive or sometimes impossible to recollect the required data and reconstruct the models from the beginning. In this case, knowledge/learning transfer between task domains would be beneficial [83]. Transfer learning techniques transmit the knowledge from previous tasks to a target task when the target task possess less high-quality training data [83].

In contributing to previous methods and improving previous works, the present study introduces a novel computer-aided automatic cataract grading method for detecting various cataracts stages using smartphone images. Since most previously-used approaches were based on hand-crafted features of extracted eye regions, the current study attempts to feed ROIs to a CNN model to improve the feature extraction and grading process and extract useful features using the end-to-end nature of CNN.

In the current study, the IIITD dataset [1] will be used, classified into four stages: normal, early, pre-mature, and mature cataracts. Smartphones captured eye images before and after cataracts surgery. Since most of the traditional methods are based on frontal-view facial poses, the eye detection process in IIITD images is difficult because, in most of the captured images, only a part of the face is visible and is not a frontal view. The current study addresses this eye detection challenge using an eye detection method practical for non-frontal facial poses.

Additionally, the proposed method will implement an automated end-to-end method to a larger database.

The current study will compare available pre-trained convolutional neural networks (CNN) like VGG, ResNet, and Inception v3 together and also use the best accuracy to carry out automatic cataract classification. The pre-trained CNN models are trained and tested on a wide-scale ImageNet dataset [84]. For feature extraction and classification tasks, pre-processed images are passed into a deep learning model, the output of which is then passed into the next layer as input [84]. Finally, the last layer produces the results. The last layer will be fine-tuned with digital cataractous images collected from IIITD dataset, divided into four stages of normal, early, pre-mature, and mature cataracts with the help of an optometrist.

Like all neural networks, the layer size and network depth are considered hyper-parameters. Generally, deeper models show better performance than shallow models in extracting richer features; however, too much depth does not always guarantee the best performance for all types of tasks. Furthermore, the more data we have for our task, the more we can un-freeze the original model layers and fine-tune them for our specific task. After selecting an effective optimizer, data augmentation is used to avoid over-fitting and neutralize data disparity. Augmented data will reproduce a more comprehensive set of potential data points, reducing the difference between training, validation set, and any future testing sets [85].

Chapter 3

Methodology

3.1 Overview

Early detection of cataracts is considered a necessary first step. Preliminary research shows that it is possible to detect cataracts using image processing. As smartphones become sensibly universal in most urban areas, cataracts screening with smartphones reduces the screening cost and travel/time bothers for patients. Accordingly, a novel computer-aided automatic cataract grading method is presented in the current dissertation to detect various cataracts stages, such as normal, early, pre-mature, and mature cataracts, from the smartphone images.

3.2 Contributions of this thesis

1. A novel computer-aided automatic cataract grading method is presented to detect various cataracts stages, namely normal, early, pre-mature, and mature cataracts, using the photo-taking function of smartphones.

2. The recent, state-of-the-art studies on cataract grading utilize a device simulating slit-lamps. This device is attached to the smartphone camera [52]. Although they provide good quality images that are very close to the slit-lamps in the clinics, it can be a burden for the patient to access these types of equipment in rural areas of developing countries. Therefore, by proposing a precise and accurate screening automated method, the current investigation aims to compete with the portable slit-lamp-based approaches.

3. By proposing an eye detection method practical for non-frontal facial poses, the current thesis aims to deal with the non-frontal poses and eye detection challenge in the utilized dataset.

4. In the current study, as first steps, a few feature engineering methods such as the Canny edge detection method and Hough transform were applied to some of the eye regions after the eye detection step. Because the images are not focused on the pupil, and the pupil is a very small region of the images, the edge detection and the Hough transform method were not successful in extracting the pupil, and the results do not make sense. Therefore, we did not report the feature engineering results. In the IIITD dataset, the total number of images was 1324. A general feature extraction method was required to extract reliable features from images with inconsistent orientations, facial poses, backgrounds, illumination, and distances between camera and subject. Therefore, we implemented multiple CNN-based models such as a four-layer CNN and pre-trained networks such as VGG-16, Inception V3, and ResNet-101 for cataract grading to extract general features and do cataracts classification in an end-to-end framework.

5. The proposed method also aims to compare the available pre-trained Convolutional Neural Networks (CNNs) such as VGG network, ResNet, Inception v3, and use the best accuracy to carry out automatic cataracts classification.

6. In the earlier state-of-the-art methods, the classification result on a dataset with 120 images indicates that the SVM classifier cannot process effectively and needs a larger dataset tawfik2018early. The current research will implement ResNet-101 feature extractor combined with SVM classifier on a larger dataset, and also it aims to compare the result with an end-to-end ResNet-101. The aim is to remove the shortcoming of the small number of input images.

7. Most previous research has only dealt with the two-class classification problems - cataract and no-cataract - by smartphone [6] and the cataract stages were limited into early and advanced cataracts [44, 78, 74]. Therefore, degrees of cataracts, including early, pre-mature, and mature cataracts, will be automatically graded in the current study.

3.3 Proposed Methodology

A novel computer-aided automatic cataract grading system proposed in the current study consists of several vital steps. After data acquisition and preparing the images, it is necessary to detect the eye region. By considering the eye region as the region of interest,

proper features must be extracted to classify the input images into normal and various cataracts stages. The comprehensive description of each step is presented in the following sections. Figure 3.1 illustrates the block diagram of the cataract grading system proposed in the current dissertation.

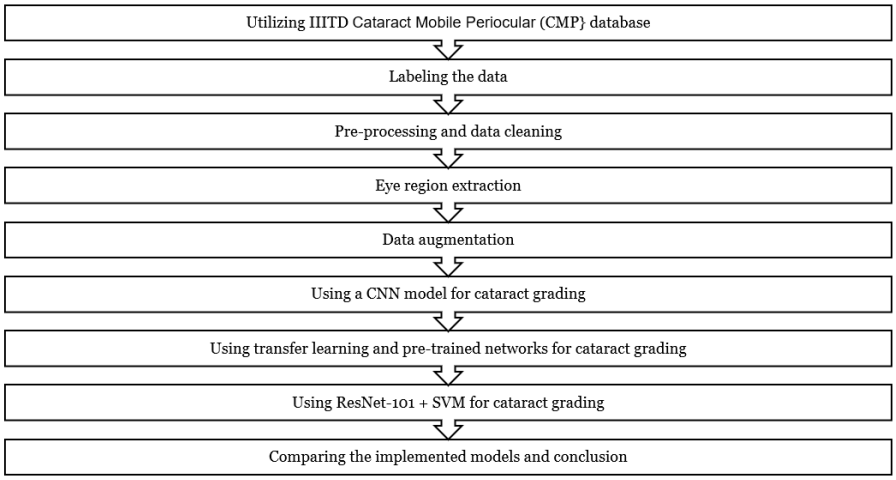

Figure 3.1: The proposed block diagram for the cataract grading system using smartphone-based images.

3.3.1 Dataset

The IIITD Cataract Mobile Periocular (CMP) database [1] was utilized as the cataractous and normal data images [1] in the current study. This database contained images capturing ocular information utilizing the photo-capturing functionality of the smartphones in the visible spectrum.

The IIITD dataset [1] in the current study is a part of a large-scale biometric recognition systems project. Large scale biometric recognition systems often rely on the iris as a specific modality. Based on the literature on iris recognition, the publications have proposed using periocular biometrics when iris recognition fails. It has been established that periocular recognition can be more accurate for recognition at a distance than iris recognition [1]. Periocular biometrics is ocular biometrics where the ocular region includes the eyebrow, pupil, sclera vasculature, and iris [1]. The IIITD dataset consists of periocular images for recognizing the identity of individuals [1]. A significant number of individuals registered in large-scale identification programs belong to a population over the age of 50 years [1].

Sessions	2
Pre-operative subjects	145
Post-operative subjects	99
Total number of images	1804
Resolution of an image	4608 ×3456
Mobile camera resolution	16 megapixels
Eliminated images due to noises	480
Total number of images after elimination	1324

Table 3.1: The characteristics of the IIITD Cataract Mobile Periocular database [1].

Statistically, eye-based diseases such as cataracts have been quite prevalent among this age group. According to the National Eye Institute (NEI), 14.36% age group 50-60, 40.18% age group 60-70, 85.98% age group 70-80). The IIITD dataset is captured in phases: pre- and post-cataract surgery sessions by a smartphone camera in uncontrolled illumination, complex background, and geometric distortions [1]. The images had challenges, including translation, rotation, and blur. Therefore, compared to the medical offices for cataract grading, the IIITD dataset does not have controlled illumination with controlled orientation and distance between the camera and the subject.

The images were obtained using a MicroMax A350 Canvas Knight mobile phone equipped with a 16-megapixel camera [1]. The eye images were labeled and diagnosed by an optometrist. Table 3.1 and figure 3.3 illustrates the dataset characteristics and sample images of the IIITD dataset respectively.

There is no information about the age groups and the number of each gender in the IIITD dataset. As one of the steps before cataract grading, we detect the eye regions, and we did cataract grading independent of the gender and age group. Figure 3.3 illustrates the challenges regarding the IIITD dataset.

Pre-processing and extra data cleaning

Due to the unconstrained nature of the images, in the first step, some data pre-processing and data cleaning were applied to prepare the input images and make them usable for CNN network input. Therefore, the images were checked one by one manually, and 480 images from a total of 1804 images were eliminated because of blurriness, low quality, or closed

Figure 3.2: Sample images of the IIITD database.

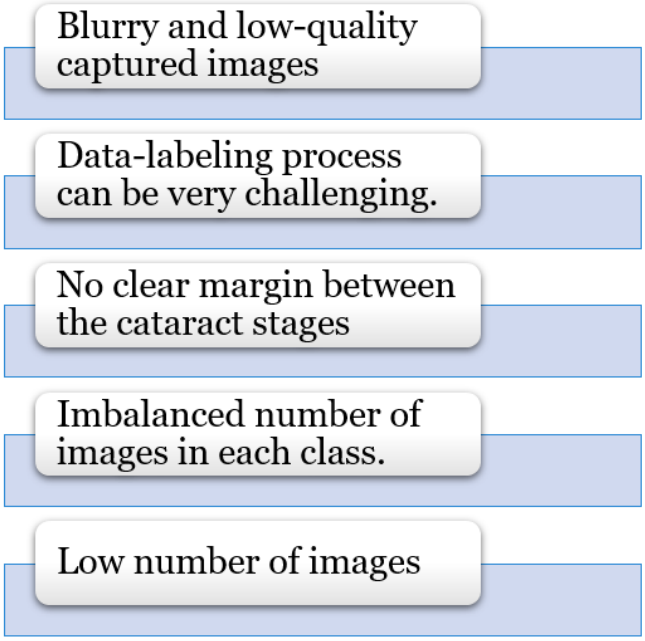

Figure 3.3: The challenges regarding the IIITD dataset

eyes. The outliers and the scattered data were eliminated; for the remaining data. After resizing, the eye regions were detected and were fed as the input to the CNN networks. Some examples of eliminated images are shown in Fig 3.4. Table 3.2 shows the number of available data for each label after data cleaning.

Table 3.2: Number of available data for each label after data cleaning

Label	Number of images
Normal	70
Early cataracts	121
Pre-mature cataracts	404
Mature cataracts	243

Figure 3.4: Corrupted images.

3.3.2 Eye region detection

As described in the previous section, due to the high resolution and the large size of images, direct use of this data considerably prolongs network training time. Due to this issue, these images must be resized before feeding to the CNN network. One easy way to fulfill it is to rescale the images directly. However, this method reduces the photos' quality, and some parts of the data will be lost. An alternative solution is to extract the subjects' eyes in every photo using an eye detection method and then feeding the detected eyes to the classification/grading model. For fulfilling this purpose, there are many different methods and tools. In this project, the dlib face landmark detection tool is used to extract the subjects' eyes [86].

Dlib face landmark detection method finds human faces in an image and estimates the facial pose in the images [87]. This robust approach employs component landmark detection that performs accurately for all poses changing from side to frontal view [86, 87]. To obtain robust detection for extreme poses, it uses a set of independent poses, and specific landmark detectors [87]. The failure rate for this technique is lower than the commercially available software [87]. The pose considers 68 landmarks. These landmarks are facial points such as the corners of the mouth, the eyebrows, the eyes, etc. A sample of the landmarks has been shown in figure 3.5.

The face detector employed the classic Histogram of Oriented Gradients (HOG) features accompanied by a linear classifier, an image pyramid, and a sliding window detection strategy [88]. The pose estimator was built utilizing the dlib method proposed by [89, 88] and was trained on the iBUG 300-W face landmark dataset.

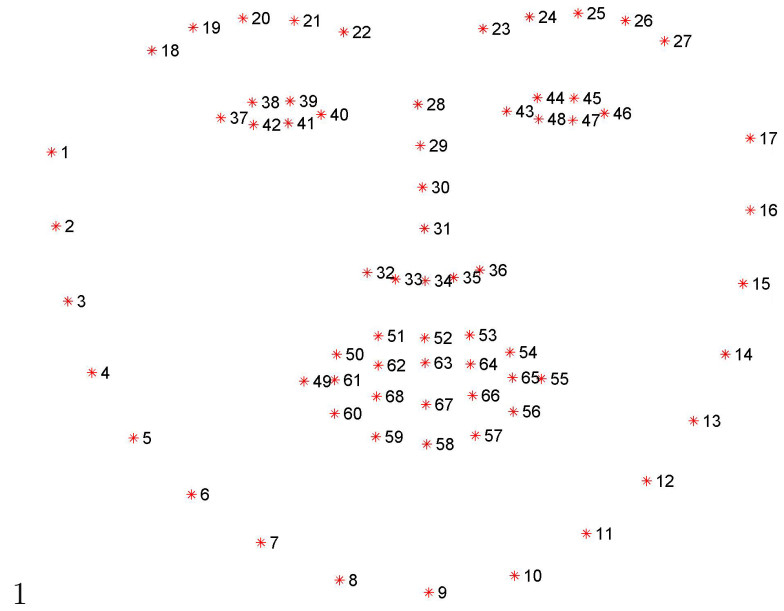

Figure 3.5: The full set of facial landmarks that can be detected via dlib [3].

In the first part of the implementation, for eye region extraction, after rescaling the images, the landmarks were extracted using dlib landmark detection technique. Then, according to the samples' extracted landmarks, the acceptable range of pixels was selected to obtain the right eye bounding box.

It has to be noted that most of the existing eye detection methods use models to identify a person's full face and then obtain the eyes or other parts from the extracted face. Since most of the IIITD dataset photos include only a part of the face, the direct use of the landmarks extraction method in many cases does not match the true landmarks correctly. To get a better and more accurate result, a method was used to recover the person's entire face and then extract the eyes. For this purpose, each image was concatenated to its flipped image horizontally to create a complete image of the person's face. Next, with the described method, the existing landmarks were identified, and the person's eye region was clipped from the photos.

After extracting the eye regions, all the outputs were checked manually, and the output was eliminated if the eye region was not extracted wholly and correctly.

The centralized eye region in the detected bounding box was another issue that had to be checked. If there was a little deviation, it was ignored because there were other samples from various angles for each eye region, and the samples from different angles could cover

these small deviations. The critical point in extracting the eyes from the original photos is that the location of the eye's landmarks may be such that the photo's size may be smaller than the required size. In this case, zero-padding was added to the photos.

In the next step, the extracted eye region was fed into the CNN model to be graded. However, before this step, the augmentation process was implemented on the available images.

3.3.3 Data augmentation

Data augmentation was also performed on the dataset. It prevents the network from memorizing the exact details of the training images and overfitting. By adding transformed versions of images in the dataset, the image augmentation method artificially expands the size of the training dataset. Training the CNN model on more variable images results in more skillful models and generalizing better from what they have learned from new samples. Image transformations include a range of operations from image manipulation, such as shifts, flips, zooms, rotates, and much more.

It has to be mentioned that another procedure that solves the centralization in the detected bounding boxes is the horizontally and vertically shifting technique, which were utilized as a part of the augmentation techniques. The horizontal and vertical shifting techniques can reduce non-centralized eye regions' effect in the detected bounding boxes.

In the next step, the augmented data was used to train the CNN network.

3.3.4 Proposed Convolutional Neural Network for automated cataract grading

Network Architecture

After data pre-processing, different models can be used for classification. This project used CNNs to categorize the images into four classes of normal, early cataracts, premature cataracts, and mature cataracts. For this purpose, a network with the following specifications was used.

Input

In this step, to speed up the training process, the input images were pre-processed and resized into half of the initial size.

Network Layers

Four layers of convolutions, in which [32,32,64,64] filters with kernel size [(5,5), (5,5), (3,3), (3,3)] were used in each layer, respectively. Also, a max-pooling layer was used between each pair of convolution layers. After convolutional layers, a flattening layer was used to convert the data into a 1-dimensional array to pass it as the input to the next layer. Finally, two Dense layers were used to dimension reduction and determine the label of the images. Figure 3.6 illustrates the architecture of the implemented network and its layers.

Output

Existing data labels include the following six values:

1. NC/ NVC (No view of Cataracts)
2. EC (Early Cataracts)
3. PMC (Pre-Mature Cataracts)
4. MC (Mature Cataracts)
5. NV (No clear view)
6. IOL (Intra-ocular lens)

NV-labeled data was removed from the processed data due to noise. Also, IOL(Intra-Ocular Lens) was considered as the normal and no cataract condition. Therefore, four classes were considered for training the dataset. To feed these labels into the CNN network, a one-hot encoding was used. A sample of one-hot encoding is shown in Figure 3.7.

Weight initializers

Generally, the neural network starts with some weights, and then in an iterative process, the weights are updated to better values. The term kernel initializer is a term that is utilized for the statistical distribution or the function to initialize the weights. The library will create numbers from that statistical distribution and employ them as starting weights. In this project, a random normal distribution was used as the kernel initializer.

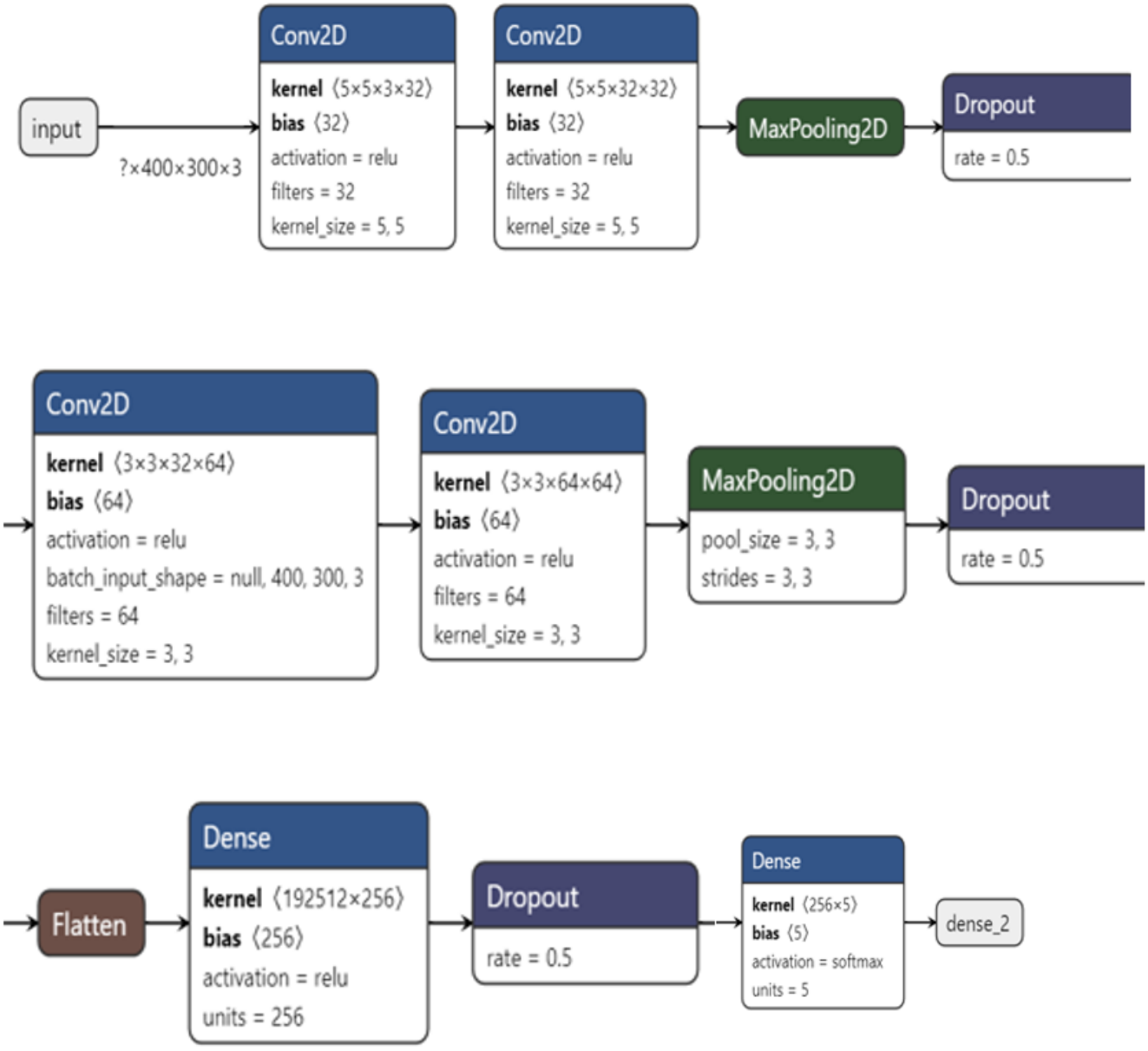

Figure 3.6: The architecture of the implemented 4-layer CNN network.

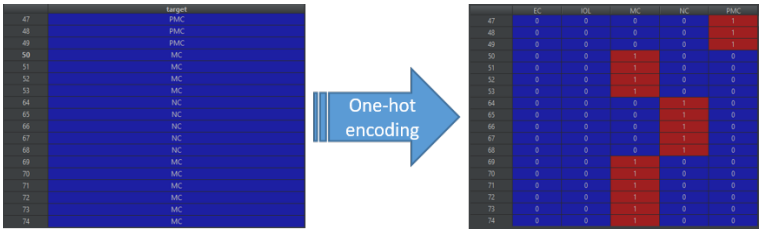

Figure 3.7: sample of one-hot encoding.

Activation Functions

The Relu and Leaky Relu function were used as the activation function for convolution layers and existing dense layers. Finally, for the last dense layer (output layer), the softmax function was used as the activation function.

Dropouts

Deep learning neural models are prone to over-fit quickly on the training data with few samples. By randomly dropping out some of the neurons during the training process, a single model can simulate having a large number of examples and avoid over-fitting. This method is called dropout. It provides a very computationally efficient and notably effective regularization strategy to reduce over-fitting and improve generalization error in deep neural networks. In the implemented network, a dropout layer was used after each convolution and dense layer.

Weight Regularizations

The second principal approach to control the complexity of a model is utilizing regularization, which involves adding a penalty term to the error function. Weight regularization is a technique to decrease the overfitting of a deep learning neural network model on the training data and make the model's performance better on new data. In the implemented network, an l2 norm regularization was used for the convolutional layers.

Optimizer

The Adam optimization algorithm is a replacement for the classic stochastic gradient descent. It is computationally efficient and straightforward in terms of implementation. As

another advantage, it is suitable for tasks with noisy gradients. In this project, Adam optimizer was used as a network optimizer.

3.3.5 Training

The classification model is responsible for assigning the items in the dataset to the determined classes. Consequently, the classifier must be assessed to determine its accuracy, error rate, and error estimates. The holdout method was used as one of the most primitive methods to evaluate the classification models in terms of their accuracy, error rate, and error estimates in the current study. After shuffling the dataset in this technique, the data set was randomly divided into two sets of the training set and the test set. The maximum data is randomly selected in the holdout method and belongs to the training set, and the remaining data belongs to the test set. The partitioning rate was 80:20 for the training and test sets. After partitioning the dataset into two sets, the training set was used to create the classification model. After building the model, we used the data examples in the test set to examine the accuracy, error rate, and error estimate of the cataract grading model. If maximum possible data examples are considered for the training set in the holdout method, the error rates would be very low, and accuracy would be high. It can be interpreted as a good classification model.

Inspired by various works, due to the holdout method's simplicity, the holdout method was selected to train and test the cataract grading model in the current study, but other techniques such as cross-validation could be utilized as well [3]. The holdout method is suitable to use when we are on a time crunch because cross-validation uses multiple train-test splits. Cross-validation needs more computational power and time to run than using the holdout method. On the other hand, cross-validation is beneficial because it offers the model the chance to train multiple train-test splits.

The network was trained for 500 epochs, and the model with the best validation accuracy (the least loss) was saved and considered as the final model.

3.4 The evaluation of cataract grading by various pre-trained CNN's

In general, transfer learning is a technique in deep learning and is a process of training a neural network on one problem and using it somehow on a second related task [83, 90]. There are several available pre-trained networks namely, VGG-16, ResNet-101, Inception

v3. These networks have been trained on ImageNet dataset with more than a million images and can classify images into 1000 object categories [83, 90]. Then, one or more layers of the trained model are utilized in a new model and fine-tuned on the desired task [83, 90]. Transfer learning is beneficial in terms of decreasing the training time for a CNN model and reducing the generalization error[83, 90]. This method is also used when the number of available data is low, and the quality of data is not good enough to extract reliable features for the machine learning task [83, 90].

There are some usage patterns of transfer learning for pre-training task [83, 90] that includes:

1. Classifier: The pre-trained network is utilized directly to categorize new images [91].
2. Standard feature extractor: The pre-trained model, or some portion of the pre-trained network, is utilized to do some sorts of image processing and obtain related features [91].
3. Integrated feature extractor: The pre-trained model, or some part of the pre-trained network, is united into a new model, but during the training process, the layers of the pre-trained model are frozen [91].
4. Weight initialization: The pre-trained model, or some part of the network, is merged into a new model, and the layers in the pre-trained model are trained at the same time with the new model [91].

Each strategy can effectively and time-savingly help to develop and train a deep convolutional neural network model. In the current study, to compare the pre-trained networks and evaluate the proposed CNN model, we used VGG-16, Inception v3, ResNet-101 pre-trained models, which contain 16, 48, and 101 layers, respectively. The relevant recent works were our motivation to select these networks [15]. The image input layer requires 224x224x3 , 299×299x3, 224x224x3 input size for VGG-16, Inception v3 and ResNet-101 respectively. To use pre-trained models, first, the data was pre-processed for a second time based on each model's requirements. For training, all layers of the base model were frozen(except Batch-Normalization layers for ResNet-101 and Inception v3), and only the top layers(classification layers)of the network were trained(which were randomly initialized). As the last step, some network layers (according to network architecture) were fine-tuned for some epochs. For more information on VGG, ResNet, and Inception v3 and their architectures, you can refer to appendices in the current dissertation.

In the utilized four-layer CNN model and the pre-trained networks, the number of layers and nodes between the models was different, but they were the same for each model's available input images.

3.4.1 Cataract grading using SVM classifier

To evaluate the current classification technique with the standard supervised classifiers, support vector machine SVM was utilized to classify the available dataset into four classes of normal and three cataracts stages, including early, pre-mature, and matured cataracts. As the feature extractor, pre-trained ResNet-101 was utilized. In the next step, the extracted features were fed as the input to the SVM classifier. The Radial Basis Function (RBF) kernel was utilized, and C and gamma parameters were tuned to get the best result.

Chapter 4

Results

In this chapter, the IIITD dataset's details and the evaluation criteria utilized in the experiments are presented, and then the experimental results and the related analysis are shown.

4.1 Experimental setup

4.1.1 Dataset

Among the 1804 periocular images available in IIITD dataset, 480 scattered and noisy images were eliminated, and a total number of 1324 images were remained to fed into the cataract grading system. The images were labeled and graded by an experienced optometrist into four class of normal, early, pre-mature and mature cataracts. After data cleaning, in the IIITD images, 70 images had non-cataract labels, while 121 , 404, and 243 images were diagnosed as early, pre-mature, and mature cataracts, respectively. The dataset was divided randomly into two non-overlapping subsets of patients, i.e., training and test, at a rough ratio of 8:2, and the validation split rate was 0.2. Figure 4.1 shows few samples of the photographs in the dataset.

All of the methods were implemented using Python 3.8 Keras library and tested on a personal computer (Microsoft Windows 10, CPU Intel Core i7 - 6700K, GPU: Nvidia Geforce GTx TITAN X RAM: 64GB).

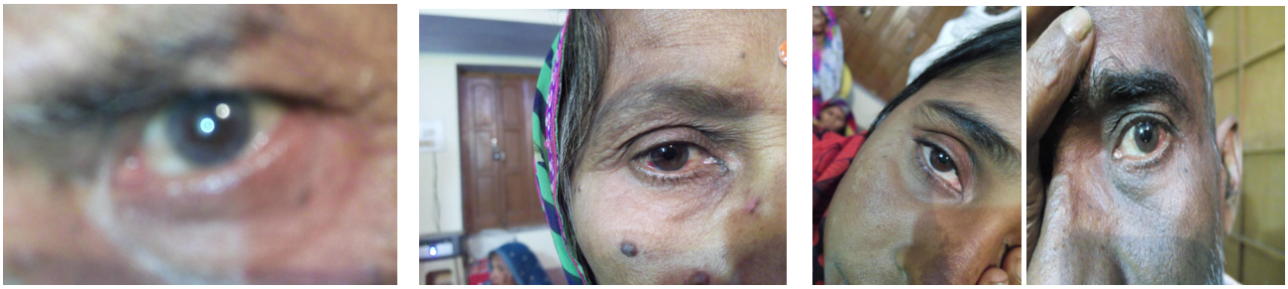

Figure 4.1: Sample images of the database.

4.1.2 Evaluation criterion

The cataract grading processes implemented in the current study were object detection for eye region extraction and a classification problem in the grading phase. According to the equations described in 2.3.4, the evaluation metrics such as accuracy, precision, recall, sensitivity, specificity, and F-measure were calculated and reported in table 4.3 to measure the performance of three-level cataract grading.

In the ROC curves illustrated in Fig. 4.8, each figure is for one model and shows the ROC curves of that model for each group of normal, early cataracts, pre-mature cataracts, and mature cataracts. Thus, the model, number of layers, and nodes are consistent between the groups, but they are changed for different implemented models.

4.2 Eye region detection results

Since the dataset images had a big size of 4606*3456 pixels and they had a complex background with irrelevant objects such as sclera, pupil, eyelid, nose, and other parts of the face; in order to alleviate this problem, different attempts were made to extract the eye regions in the IIITD images. Since the images were captured in an unconstrained condition with different facial poses, illumination condition, occlusion, and also because, in most cases, the whole face was not visible in the images, after implementing most of the conventional eye detection techniques, the implemented methods could not detect the eye region correctly, and the regions were in most cases false positive. The conventional and the eye detection techniques that were implemented include hough transform, haar cascade and Multi-Task Cascaded Convolutional Neural Networks (MTCNN) [54, 92]. To extract subjects' eyes, the images were first re-scaled by two, and then the landmarks were extracted using dlib landmarks detection. Fig. 4.2 and Fig. 4.3 are the results of detected landmarks using dlib.

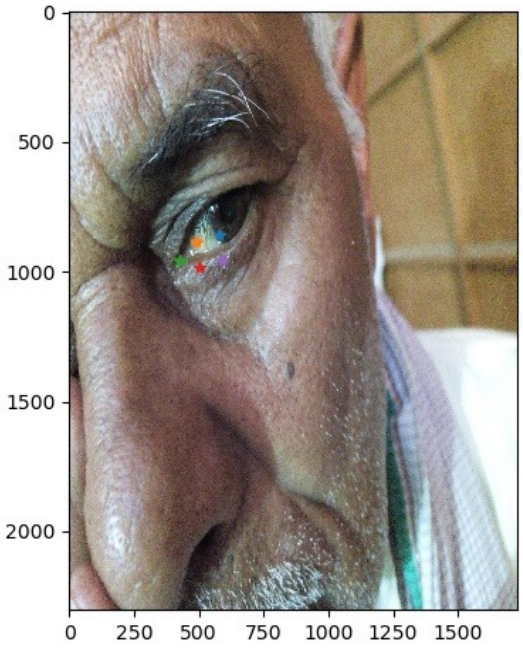

Figure 4.2: A sample of detected landmarks using dlib

The pixels in range of (37:40, 38:42) were considered as the left eye (the right eye in the dataset), and the pixels (43:46, 44:47) were considered as the right eye (the left eye in the dataset). The proposed method's outputs were the bounding boxes for the extracted eye regions (600×800 due to the use of borders for each point). To get a better and more accurate result, the person's entire face was regenerated, and then the eye regions were extracted using dlib landmarks detection method. Figure 4.4 illustrates a sample of the detected eyes using the modified dlib method and shows the difference of extracted eye regions with the previous method (using dlib without rebuilding the whole face). The second method can detect true landmarks better.

After extracting the eye regions, all the outputs were checked manually, and the output was eliminated if the eye region was not extracted wholly and correctly. The original image size for the dataset was 3456×4608, and the extracted eye region size was 600×800. Figure 4.5 shows the whole process from resizing the original image to the extracted eye region.

Eye detection results clearly show the validity of our approach. A correct eye detection rate of 97 % was achieved using the modified method.

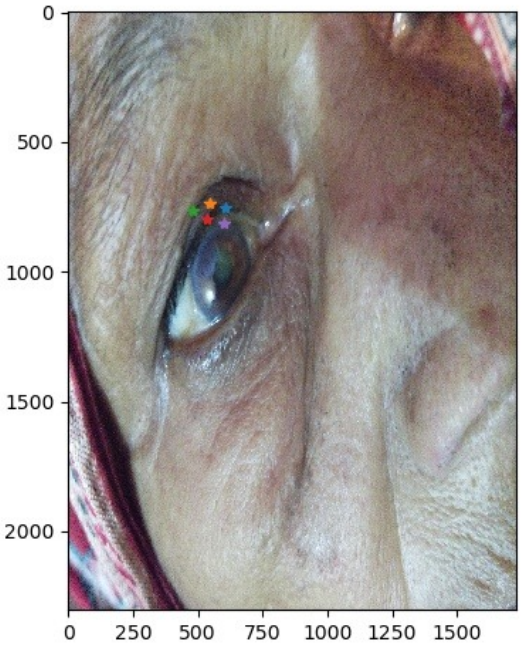

Figure 4.3: A sample of detected landmarks using dlib

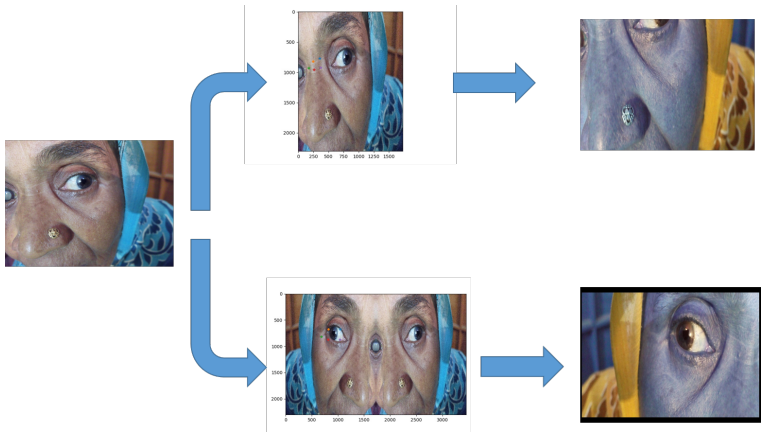

Figure 4.4: The top branch shows the eye detection process using the face landmarks. The bottom branch illustrates the eye detection in two steps of rebuilding the whole face and extracting the face landmarks using dlib.

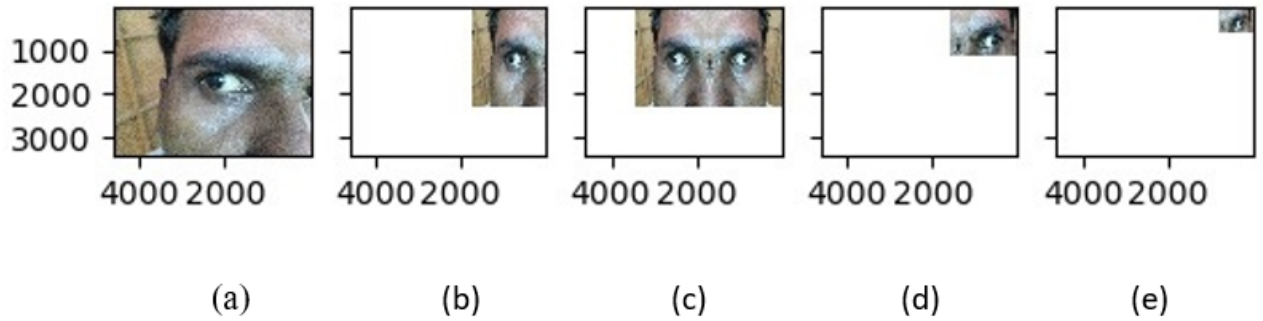

Figure 4.5: The process of resizing images from original image to the extracted eye region. (a) the raw image (3456 × 4608 pixels). (b) The image is resized into half of the initial size, (c) The whole face is rebuilt. (d) The eye region bounding box is extracted. (e). The eye region is resized into half (600 × 800 pixels).

4.3 Image augmentation

Figure 4.6 is a sample of rotation and shift up-sampling. The augmented data was utilized for training the CNN networks. The total number of images after augmentation reached 2427 images.

4.4 Cataract grading results

4.4.1 Procedure 1: CNN model

The results of the four-class classification using a four-layer CNN model are presented in table 4.2. Table 4.1 shows the number of parameters and running time for the proposed four-layer CNN model. We also presented the classification accuracy for each of the individual classes using the CNN model in Table 4.3.

The confusion matrix and the ROC curve is shown in figure fig:cm and Fig 4.8. According to the confusion matrix, the CNN model can grade images into four classes of normal, early cataracts, pre-mature, and mature cataracts with an accuracy rate of 60 %, 95%, 95%, and 63%. The best classification and grading occurred for early cataracts and pre-mature cataracts.

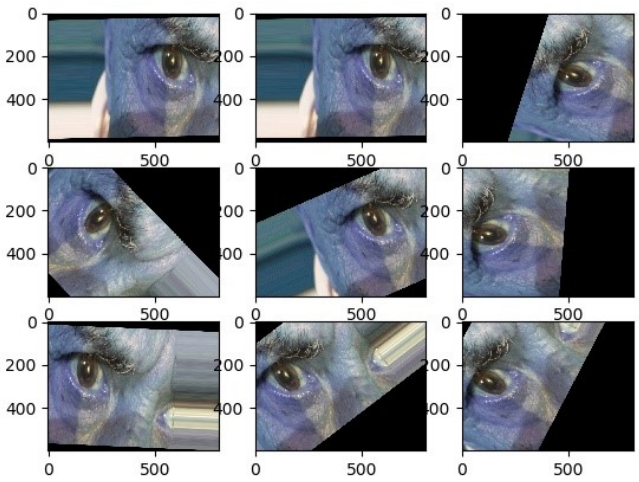

Figure 4.6: Augmentation sample (shift and rotation).

The ROC curve of the CNN model is illustrated in Figure 4.8. A ROC curve is visualized by plotting the TPR rate against the FPR rate and presents the trade-off between sensitivity (or TPR) and specificity (1 – FPR). The TPR shows the proportion of accurately predicted observations to be positive among all positive observations. Similarly, the FPR, the horizontal axis in this curve, is the proportion of falsely predicted observations to be positive among all negative observations. Classifiers that result in curves closer to the top-left corner show better performance. Furthermore, the closer the curve is to the 45-degree diagonal of the ROC space, the less accurate the test is. According to the ROC curve in figure 4.8, ROC curve of class 1 (mature cataracts) with area = 0.88 shows less accuracy, and ROC curve of class 0 (early cataracts) with the area = 0.97 is closer to the top left corner and has better accuracy.

4.4.2 Procedure 2: Transfer learning

In the second procedure, first, the data was again pre-processed based on each model's requirements. Then, all layers of the base model were frozen(except Batch Normalization layers for ResNet-101 and Inception v3), and only the top layers(classification layers)of the network were trained(which were randomly initialized). After some epochs of training of the classifier, these layers including Inception v3 = 249:315, ResNet = 338:349, VGG-16 = 19:23 were re-trained respectively. First, for the fine-tuning and training of the classifier,

50 epochs were selected. For the training of the other layers, the epoch was increased up to 500. The other parameters utilized to fine-tune the pre-trained models include rmsprop as the optimizer for the first training and Stochastic Gradient Descent (SGD) for the second round of training. The learning rate (lr) was 0.0001, and momentum was selected 0.9.

The average accuracy rates of the pre-trained networks i.e. VGG-16, ResNet-101, Inception v3 are shown in table 4.2.The results show that among the pre-trained models implemented and fine-tuned for cataract grading, ResNet-101, with the mean accuracy of 89.62 %, achieved the best accuracy in cataract grading. The ResNet-101 pre-trained model has been trained on more than a million images of ImageNet and contains 347 layers in total, corresponding to a 101 layer residual network, and can classify images into 1000 object categories. The first layer inputs 224x224x3 image size. Table 4.1 shows the number of parameters and running time during the training process for each pre-trained model.

The ROC curve of the VGG-16 network is illustrated in figure 4.8. According to the ROC curve in the figure. 4.8, ROC curve of class 2 (no cataract/normal) with area = 0.95 shows less accuracy and is closer to the 45-degree line. The ROC curve of class 0 (early cataracts) and class 1 (mature cataracts) are equal, and with the area = 0.98 are closer to the top left corner and have better accuracy.

The confusion matrix of the VGG-16 network is shown in the figure. 4.7. According to the confusion matrix, the accuracy rate of 67%, 84 %, 95%, and 83% is achieved to classify the images into a normal, early cataracts, pre-mature cataracts, and mature cataracts, respectively. The best classification and grading occurred for pre-mature cataracts.

The ROC curve of the Inception v3 network is illustrated in figure 4.8. According to the ROC curve in the figure. 4.8, ROC curve of class 1, mature test samples, and class 3, normal images, with equal area = 0.95 show less accuracy and are closer to the 45-degree line. The ROC curve of class 0 (early cataracts) with the area = 0.98 is closer to the top left corner and has better accuracy.

The confusion matrix of the Inception v3 network is shown in figure 4.7. According to the confusion matrix, the accuracy rates of 76%, 78 %, 89%, and 83% are achieved to classify the images into a normal, early cataracts, pre-mature cataracts, and mature cataracts, respectively. The best classification and grading occurred for pre-mature cataracts. Moreover, the least accuracy was for the classification of normal images.

The ROC curve of the ResNet-101 network is illustrated in figure 4.8. According to the ROC curve in the figure 4.8, ROC curve of early cataract, mature cataract, and the pre-mature cataracts images (class 0, 1, and class 3) with equal area = 0.98 show less accuracy comparing with normal lenses (class 2) with area =0.99 and are closer to the 45-degree line. The ROC curve of normal test samples (class 2) is closer to the top left corner and has better accuracy.

The confusion matrix of the ResNet-101 network is shown in the figure 4.7. According to the confusion matrix, the accuracy rates of 87%, 86 %, 93%, and 87% are achieved to classify the images into a normal, early cataracs, pre-mature cataracts, and mature cataracts, respectively. The best classification and grading occurred for pre-mature cataracts. The least accuracy was for the classification of early cataracts images.

4.4.3 Procedure 3: Cataract grading using SVM classifier

In the last phase of the experiment, the end-to-end ResNet-101 was compared with the SVM classifier. For C in the range of [1,10,100,500], the best achieved mean accuracy is illustrate in table 4.2. Table 4.3 compares the evaluation metrics and accuracy rates for each of the classes. According to SVM's results as the classifier, the Fully Connected (FC) layer in the end-to-end pre-trained ResNet-101 could outperform the cataract grading model using SVM as the classifier.

Fig 4.8, illustrates the ROC curve for ResNet-101 using SVM. The ROC curve of the ResNet-101 network as a feature extractor and SVM classifier used both together as the cataract grading model is illustrated in figure 4.8. Accordingly, the normal images' ROC curve (class 2) with area = 0.98 shows less accuracy and is closer to the 45-degree line. The mature cataracts group (class 1) with area =0.91 is closer to the top left corner and shows better accuracy.

The confusion matrix of the ResNet-101 network as the feature extractor and SVM classifier is shown in the figure 4.7. According to the confusion matrix, the accuracy rates of 65%, 75 %, 95%, and 86% are achieved to classify the images into normal, early cataracts, pre-mature cataracts, and mature cataracts, respectively. The best classification and cataracts grading are for pre-mature cataracts, and the least classification accuracy was for the classification of the normal images.

4.5 Conclusion

According to Table 4.2, ResNet-101 is outperforming the other models implemented in this study.

To see how the model performs for each class, we can take a look and use other calculated metrics in Table 4.3. According to the definitions, precision answers that how many of those whom we labeled as, for example, early cataract have actually early cataract? Recall looks for that from all the participants who are EC, how many of those are correctly

Table 4.1: Execution time and total number of parameters for each implemented model.

Model	Total parameters	$\text{Train}_1 parameters$	Time per epoch(S)	Fine-tuning	Time per epoch(S)
Proposed CNN	5,398,884	5,398,884	78		
VGG16	15,244,100	529,412	31	1,347,530	38
Inception v3	23,905,060	2,119,492	37	13,227,844	43
ResNet-101	44,760,452	2,207,620	61	5,618,052	62

Table 4.2: The mean accuracy of the implemented models

	CNN	VGG-16	Inception v3	ResNet-101	ResNet-101 with SVM
Mean Accuracy	84.67	87.64	84.67	89.62	87.14

predicted? The recall is the same as sensitivity. F measure is the harmonic average of the precision and recall and considers both of them. According to all the previously mentioned metrics and also F-measure, which is the trade-off between recall and precision, the results show that for each individual class, ResNet-101 outperforms the other models and has better grading results for IIITD with imbalanced data.

Table 4.3: The metrics of the implemented models

		CNN	VGG-16	Inception v3	ResNet-101	ResNet-101 with SVM
Accuracy	EC	0.95	0.84	0.78	0.86	0.75
	MC	0.63	0.83	0.83	0.87	0.86
	NC	0.60	0.67	0.76	0.87	0.65
	PMC	0.95	0.95	0.89	0.93	0.95
Precision	EC	0.71	0.91	0.87	0.87	0.88
	MC	0.89	0.93	0.82	0.89	0.90
	NC	1	0.84	0.73	0.91	0.97
	PMC	0.87	0.85	0.87	0.90	0.85
Recall	EC	0.95	0.84	0.78	0.86	0.75
	MC	0.63	0.83	0.83	0.87	0.86
	NC	0.6	0.67	0.76	0.87	0.65
	PMC	0.95	0.95	0.89	0.93	0.95
Sensitivity	EC	0.95	0.84	0.78	0.86	0.75
	MC	0.63	0.83	0.83	0.87	0.86
	NC	0.6	0.67	0.76	0.87	0.65
	PMC	0.95	0.95	0.89	0.93	0.95
Specificity	EC	0.93	0.99	0.98	0.98	0.98
	MC	0.97	0.97	0.92	0.96	0.96
	NC	1	0.99	0.98	0.99	1
	PMC	0.84	0.83	0.88	0.90	0.83
F-measure	EC	0.82	0.87	0.82	0.87	0.81
	MC	0.74	0.88	0.83	0.88	0.88
	NC	0.75	0.75	0.74	0.89	0.78
	PMC	0.91	0.89	0.88	0.91	0.89

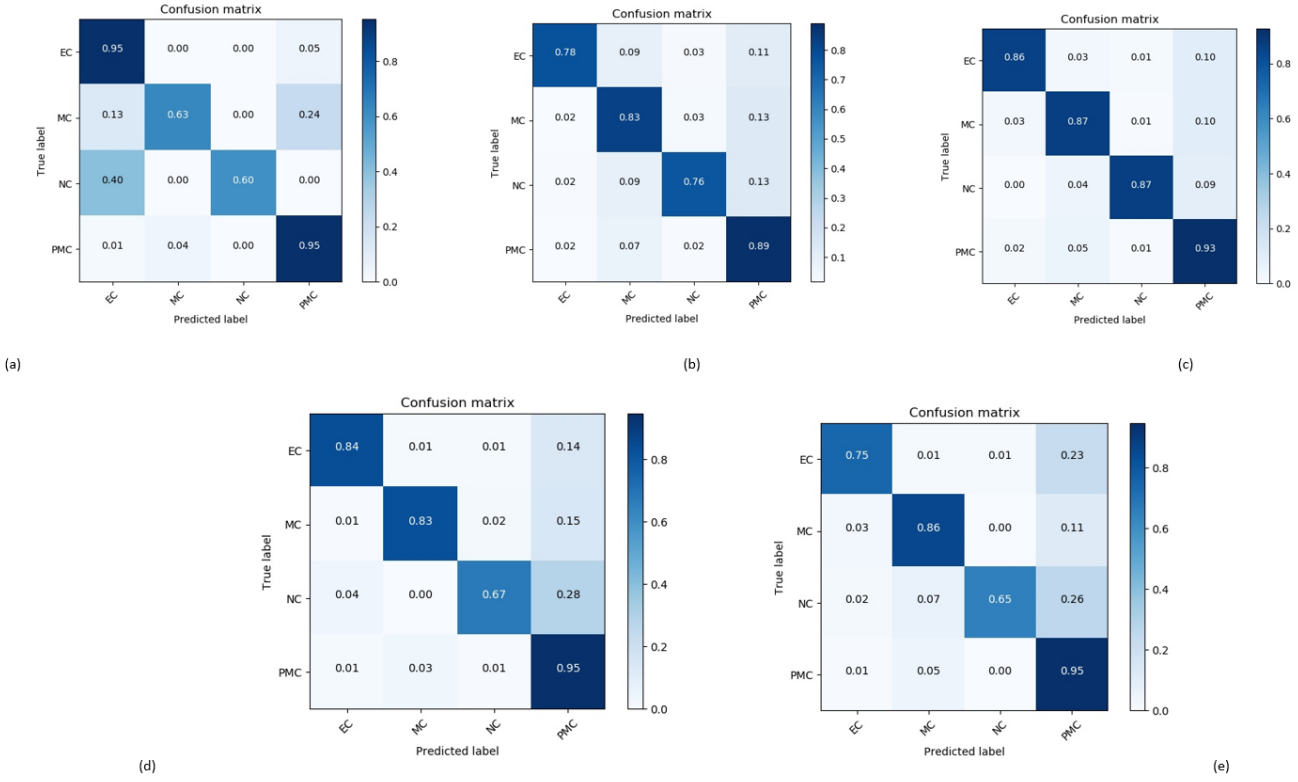

Figure 4.7: The confusion matrix of the implemented models. (a). The confusion matrix of the CNN model. (b). The confusion matrix of the Inception v3 model. (c). The confusion matrix of the ResNet-101 model. (d). The confusion matrix of the VGG-16 model. (e). The confusion matrix of the ResNet-101 feature extractor and SVM classifier model.

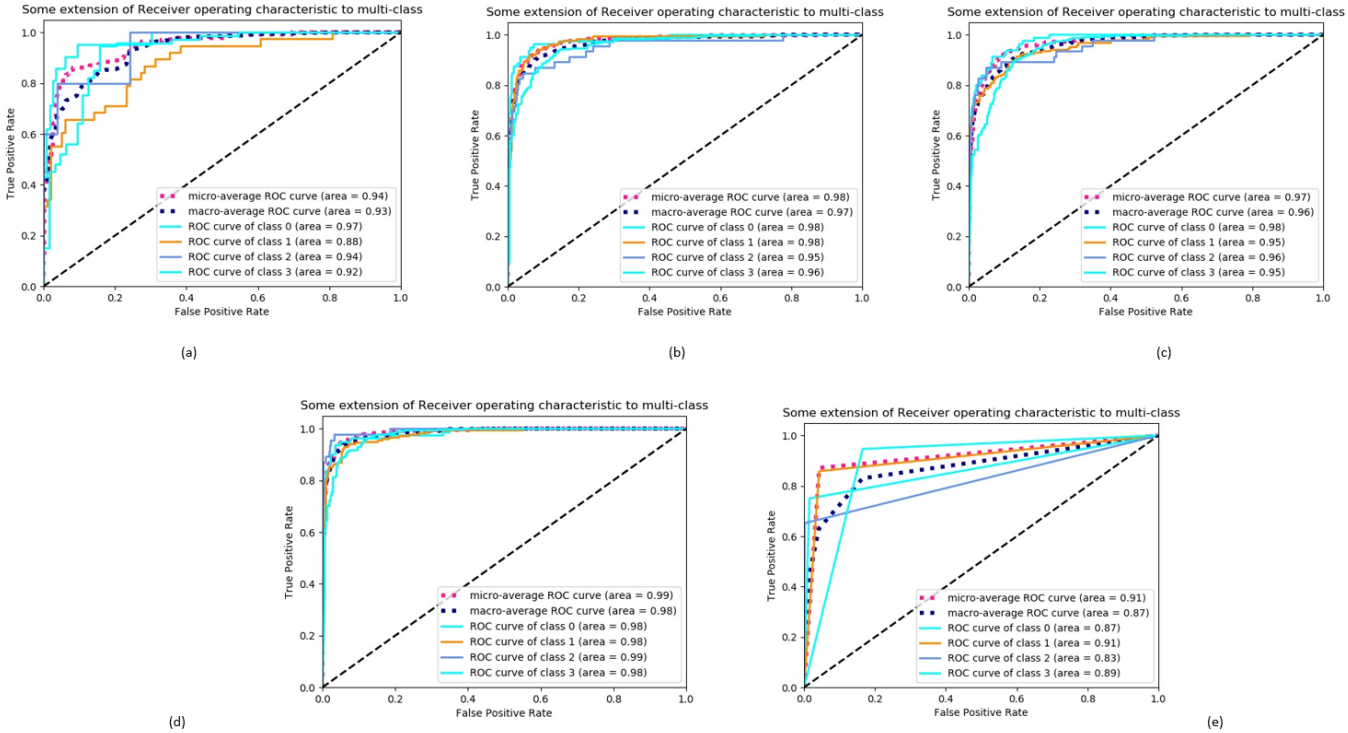

Figure 4.8: The Receiver Operator Characteristic (ROC) curve of the implemented models. (a). The ROC of the CNN model. (b). The ROC of the VGG-16 model. (c). The ROC of the Inception v3 model. (d). The ROC of the ResNet-101 model. (e). The ROC of the ResNet-101 feature extractor and SVM classifier model. class 0 = early cataract (EC), class 1 = mature cataracrt (MC), class 2 = no cataract (NC), class 3 = pre-mature cataract (PMC)

Chapter 5

Discussion and conclusion

In this project, a framework is presented to diagnose and categorize different degrees of cataracts using incomplete facial photographs captured by the smartphone camera. The current study aims to distinguish between the normal and the cataractous lenses in the captured images. After detecting cataracts, the system will grade the severity of the cataract. Therefore, based on the priority, we should focus on a model with more TPs and fewer FPs. It is always a priority to distinguish cataracts earlier; therefore, we do not want to miss the cataract cases. Thus, choosing a model with fewer FPs for the Normal group and more TPs for the Normal cases is preferable.

Moreover, when the cataract reaches the mature and pre-mature stages, based on the patient's symptoms and how much the patient has vision loss, the cataractous lens can be a surgery candidate. Thus, we need a model with fewer FPs for pre-mature cataracts and a more significant number of TPs. Accordingly, we go for the model with the ROC cure closer to the top left corner for the pre-mature and mature cataract.

One of the differences between the proposed methodology in the current study and the recent similar cataract grading systems [43, 53] is that in the very recent related works, an external device is attached to the camera in the smartphones and simulates a portable slit-lamp to monitor the human crystal lens. Although this device captures good-quality images similar to slit-lamps in the clinics, patients' access to such equipment, especially in rural regions, can be a big challenge. The current study's main objective was to propose a smartphone-based method and offer an inexpensive and practical self-diagnosing automated approach that can compete with the expensive attachable slit-lamp device.

Since the images in the IIITD dataset were collected with the smartphone cameras and are not captured in a constrained condition; factors such as camera rotation, variations

due to rotation, translation, blurriness, the distance between subject and camera, and not having a complete facial view in the images made the eye region extraction and cataract grading a challenging task. Hence, to deal with the unconstrained condition, the total face view was regenerated. Then, the eye regions were extracted with an accuracy rate of 97%. The eye region detection method implemented in this study could deal acceptably with the non-frontal poses in the images.

The extracted eye regions were utilized for training a designed CNN network. Also, in the next procedure, transfer learning and pre-trained models were used. In this step, three different pre-trained models (VGG-16, ResNet-101, and Inception v3) were selected, fine-tuned, and compared. The comparison shows that the ResNet-101 fine-tuned model has the best outcome and can classify cataracts with reasonable accuracy.

Comparing Inception v3 and ResNet-101 with VGG-16, big advancements were made in the architecture of ResNets and Inception v3 leading to boosts in accuracy and performance. Since the training process in deep neural networks is time-consuming, and these models are prone to overfitting, the residual learning framework in ResNets (short for Residual Networks) has improved the training of significantly deeper networks than those used previously. Using the residual mapping technique in ResNets solves the problem of saturation and degradation in neural networks caused by an increase in the networks' depth. Therefore, ResNet-101 can be used to address many problems. ResNet-101 is easier to optimize and achieve higher accuracy when we increase the depth, producing better results than previous networks. Looking at the average accuracy in ResNet-101 in the current study admits this claim.

Putting the improvement in the performance and accuracy of the ResNets aside, the time per epoch for ResNet-101 is the most among the three implemented pre-trained networks. The number of total parameters for ResNet-101 is more than the other implemented networks. On the other hand, compared with pre-trained ResNet-101, the proposed four-layer CNN in the current study has fewer parameters, and in some levels of cataracts, it was compatible with ResNet-101. Since CNN is a lighter model, therefore, in future works, more investigation is required to evaluate whether the ResNet-101 or the four-layer CNN with some modifications can be a better candidate for smartphone-based cataract grading loaded on a cloud platform.

In order to compare and evaluate the classifier in the ResNet-101, the extracted features by the ResNet-101 were fed to SVM classifier. The experimental results indicated that the end-to-end ResNet-101 yields better classification accuracy when compared to ResNet-101 with SVM classifier. Two reasons have been discussed in depth in this section. As the first reason, the number of images in each class was not balanced, and pre-mature cataracts images have the most number among all classes. In comparison with ResNet-101, SVM could not handle the imbalanced data. According to the confusion matrix for ResNet-101

combined with the SVM classifier, the accuracy rate for each class is not balanced, and it is biased toward the pre-mature cataract class with the most number of images. The confusion matrix for the end-to-end ResNet-101 shows that the end-to-end ResNet-101 could better handle the imbalanced number of images than SVM.

As another disadvantage for the SVM classifier, SVM performs better when there is a clear margin of separation between classes. For the cataracts stages considered in the current cataract grading study, the margin between pre-mature cataracts and mature cataracts is not clear and separable enough. It is also hard to distinguish between normal and early cataracts cases. All these reasons caused the ResNet-101 network to better grade the cataracts stages with the efficient features extracted by the convolutional layers of the end-to-end ResNet-101.

In conclusion, this study shows the capability and advantage of using an end-to-end pre-trained ResNet-101 over an end-to-end four-layer CNN model, pre-trained ResNet-101 with the SVM classifier and three other fine-tuned pre-trained networks for automated cataract grading..

To acknowledge the difference between the environment in which the eye clinician operates and the environment in which the IIITD dataset has been obtained, the eye clinician will review the medical history and symptoms and perform an eye examination to diagnose the cataract. The clinician may conduct several tests, including visual acuity test, slit-lamp examination, retinal exam. This comprehensive eye examination includes pupil dilation [93]. It means eye drops such as Tropicamide will be utilized to widen the pupil. In a visual acuity test, the eye clinician uses the Snellen chart to measure how well the patient can read a series of letters with progressively smaller letters. In slit-lamp examination, a microscope called slit-lamp is utilized. The microscope uses an intense line or slit of light to illuminate the cornea, iris, lens, the space between the iris and cornea and examine the eye's anterior structure under magnification [93]. Therefore, we have a controlled illumination in the room with the established equipment. This test helps the clinician detect any opacification in the lens and any abnormalities in the eye's anterior section. In the retinal exam, while the eye is dilated, the clinician sees the back of the eye using the slit lamp, an ophthalmoscope, or both. The clinician looks for any sign of opacification and cataract. The specialist will also examine the retina and the optic nerve head.

Cataracts have some symptoms that the eye clinician looks for them. Besides the patient's history during his or her previous visits, these symptoms can be beneficial for diagnosing cataracts. These symptoms are blurry and double vision, sensitivity to light, or glare. The patient may also have trouble seeing in bright sunlight, indoor lights, and driving at night. Also, they experience frequent changes in eyeglass or contact lens prescription and a decrease in visual acuity. All these above-mentioned eye examinations, equipment, and symptoms are determining and beneficial in precise cataracts grading operated by the eye clinicians in the clinics.

The IIITD dataset in the current study is periocular images. In these images, the ocular region includes the eyebrow, pupil, sclera vasculature, iris, and pupil. The IIITD dataset is captured in two pre-and post- cataracts surgery sessions by a smartphone camera in uncontrolled illumination, complex background, and geometric distortions. The images had challenges, including translation, rotation, and blurriness. Therefore, compared to the medical offices for cataract grading, the IIITD dataset does not have controlled illumination with controlled orientation and distance between the camera and the subject. All these unconstrained conditions make the smartphone-based cataract grading a challenging task and prone to various errors.

Solving the IIITD dataset's unconstrained condition, a few changes can be applied to the photo capturing step. In future studies, to obtain images with similar parameters such as facial poses, the distance between the camera and the face, and controlled illumination, we can use a chin rest, and the eye clinician can make the facial pose and orientation consistent by asking the patient to put the chin on the chin rest and look at the camera in the specific distance. All this equipment is located in an individual room with controlled illumination in the clinic. We can get more similar images and decrease the uncontrolled variations and parameters in the face detection and cataract grading process. Reducing the unconstrained condition can also help us obtain standard images and reduce the number of eliminations due to noise and non-standard photo capturing situation. A larger dataset can significantly increase the output classification accuracy in the proposed automated cataract grading system.

There are some other challenges with the IIITD dataset. In the following, the solutions for these challenges are discussed as suggestions for future studies.

According to the literature, it is well established that the pupil size decreases with aging [94, 95]. From an investigation of pupil size measurements conducted on 222 subjects from 20 years of age to 89 , it was ended in that there was a considerable reduction in pupil size with age in both light and dark illuminations [94, 95]. It is essential to be noted that age-related changes in pupil size can be an experimental artifact in investigations of other aspects of aging of visual functions such as cataracts [94, 95]. Therefore, in cataract grading using the proposed method in this study, we will have a smaller pupil size in the images as the age increases in the subjects; therefore, the number of FPs and FNs will increase in the proposed model for those images. In the eyeball, the iris muscles control the pupil size. Using some medicines can influence the muscles that control the pupils and prevent the reduction of the pupil size. Therefore, as a suggestion for future work, to solve the reduced pupil size in the images, the optometrists or the ophthalmologists can dilate the pupil before capturing the image. It can help us get a better view of the lens through the pupil for cataract grading.

To improve the proposed model, due to the wide range of existing CNN architectures,

more models with fewer layers can be tested in the future to solve the problem of multi-class cataracts diagnosis.

Moreover, since each model was performing well in the grading of a specific cataract severity level, to improve the accuracy and cover this problem, the ensemble learning technique can be used in future works. Ensemble methods aggregate multiple classifiers to get better predictive performance than could be achieved from any individual classifier. Ensembles can reach better results when some significant diverse models are available. Many ensemble techniques aim to increase diversity among the models they combine. Instead of constructing one learner from one training data and suggesting various versions of an algorithm, the more practical idea is to combine various strong and week learning algorithms. Then, the most popular strategies can be utilized for aggregating the outputs of the base learners, which is finding out the majority vote in a classification task and finding the mean in the regression task. This method can be considered as the next step for future works.

In the next step, to build and launch a smartphone-based cataract grading application based on the proposed model in the current study, using a cloud platform and uploading the proposed model to the cloud is one way for online applications. Then the clients can upload the images they captured by their smartphone and get the cataract grading result. The program will determine whether the client has cataracts or not and if it is yes, it will determine the cataracts level and its severity in four levels of early cataracts, pre-mature cataracts, and mature cataracts.

For cataract grading and passing the uploaded image as the input into the ResNet-101, the images go through several steps, and it is resized from an initial size of 3456×4608 to 227×227. Therefore, it has resized significantly enough, and it is ready to be used as the input for the grading process. Because the input image size is small enough for uploading, it seems that further compressions are not required. More required compressions must be investigated after future implementations, using a cloud platform, and uploading the proposed model to the cloud as the future works.

References

[1] R. Keshari, S. Ghosh, A. Agarwal, R. Singh, and M. Vatsa. Mobile periocular matching with pre-post cataract surgery. In *2016 IEEE International Conference on Image Processing (ICIP)*, page 3116, Phoenix, AZ, 2016. IEEE.

[2] P. A. Asbell, I. Dualan, J. Mindel, D. Brocks, M. Ahmad, and S. Epstein. Age-related cataract. *The Lancet*, 365(9459):599, 2005.

[3] I. José. Facial mapping (landmarks) with dlib + python. `https://towardsdatascience.com/facial-mapping-landmarks-with-dlib-python-160abcf7d672`, Accessed = 2020-11-11, Jun 2018.

[4] E. Amor. 4 cnn networks every machine learning engineer should know. `https://www.topbots.com/important-cnn-architectures/#:~:text=The%20most%20straightforward%20way%20of,Relu%20activation%20function%20from%20AlexNet.`, Accessed = 2020-11-11, February 2020.

[5] Raimi K. Illustrated: 10 cnn architectures. `https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d`, Accessed = 2020-11-11, July 2019.

[6] I. Shaheen, A. Tariq, F. Khan (Eds), M. Jan, and Alam M. Survey analysis of automatic detection and grading of cataract using different imaging modalities. In *Applications of Intelligent Technologies in Healthcare, EAI/Springer Innovations in Communication and Computing*, page 35. Springer, 2019.

[7] H. Hashemi, E. Hatef, A. Fotouhi, A. Feizzadeh, and K. Mohammad. The prevalence of lens opacities in tehran: the tehran eye study. *Ophthalmic epidemiology*, 16(3):187, 2009.

[8] C. Cedrone, F. Culasso, M. Cesareo, R. Mancino, F. Ricci, G. Cupo, and L. Cerulli. Prevalence and incidence of age-related cataract in a population sample from priverno, italy. *Ophthalmic Epidemiology*, 6(2):95, 1999.

[9] T. Li, T. He, X. Tan, S. Yang, J. Li, Z. Peng, H. Li, X. Song, Q. Wu, F. Yang, et al. Prevalence of age-related cataract in high-selenium areas of china. *Biological trace element research*, 128(1):1, 2009.

[10] GE. Nam, K. Han, SG. Ha, BD. Han, D. H. Kim, Y. H. Kim, YG. Cho, K. Hand Park, and BJ. Ko. Relationship between socioeconomic and lifestyle factors and cataracts in koreans: The korea national health and nutrition examination survey 2008–2011. *Eye*, 29(7):913, 2015.

[11] R. Varma, M. Torres, Los Angeles Latino Eye Study Group, et al. Prevalence of lens opacities in latinos: the los angeles latino eye study. *Ophthalmology*, 111(8):1449, 2004.

[12] J. M. Yu, D. Q. Yang, H. Wang, J. Xu, Q. Gao, L. W. Hu, F. Wang, Y. Wang, Q. C. Yan, J. S. Zhang, et al. Prevalence and risk factors of lens opacities in rural populations living at two different altitudes in china. *International journal of ophthalmology*, 9(4):610, 2016.

[13] S. R. Flaxman, R. RA Bourne, S. Resnikoff, P. Ackland, T. Braithwaite, M. V. Cicinelli, A. Das, J. B. Jonas, J. Keeffe, J. H. Kempen, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *The Lancet Global Health*, 5(12):e1221, 2017.

[14] H. Hashemi, R. Pakzad, A. Yekta, MR. Aghamirsalim, M. Pakbin, S. Ramin, and M. Khabazkhoob. Global and regional prevalence of age-related cataract: a comprehensive systematic review and meta-analysis. *Eye*, page 1, 2020.

[15] C. Xu, X. Zhu, W. He, Y. Lu, X. He, Z. Shang, J. Wu, K. Zhang, Y. Zhang, X. Rong, et al. Fully deep learning for slit-lamp photo based nuclear cataract grading. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 11767, page 513, Shenzhen, China, 2019. Springer.

[16] J. V. Forrester, A. D. Dick, P. G. McMenamin, F. Roberts, and E. Pearlman. *Anatomy of the eye and orbit*, chapter 1, pages 32–34. Elsevier Health Sciences, UK, 2015.

[17] S. Bassnett, Y. Shi, and G. FJM. Vrensen. Biological glass: structural determinants of eye lens transparency. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1568):1250, 2011.

[18] K. L. Moreau and J. A. King. Protein misfolding and aggregation in cataract disease and prospects for prevention. *Trends in molecular medicine*, 18(5):273, 2012.

[19] D. Allen and A. Vasavada. Cataract and surgery for cataract. *British Medical Journal*, 333(7559):128, 2006.

[20] B. Philipson. Changes in the lens related to the reduction of transparency. *Experimental eye research*, 16(1):29, 1973.

[21] JF. Aliancy, N. Mamalis, H. Kolb (Eds.), E. Fernandez, and R. Nelson. Crystalline lens and cataract. In *Webvision: The Organization of the Retina and Visual System [Internet]*. University of Utah Health Sciences Center, Salt Lake City (UT), 2017.

[22] PJ. Foster, TY. Wong, D. Machin, GJ. Johnson, and SKL. Seah. Risk factors for nuclear, cortical and posterior subcapsular cataracts in the chinese population of singapore: the tanjong pagar survey. *British journal of ophthalmology*, 87(9):1112, 2003.

[23] C. M. Mangione, R. S. Phillips, M. G. Lawrence, J. M. Seddon, E. J. Orav, and L. Goldman. Improved visual function and attenuation of declines in health-related quality of life after cataract extraction. *Archives of ophthalmology*, 112(11):1419, 1994.

[24] M. C. Leske, L. T. Chylack, and S. Y. Wu. The lens opacities case-control study: risk factors for cataract. *Archives of ophthalmology*, 109(2):244, 1991.

[25] PK. Nirmalan, A. L. Robin, J. Katz, JM. Tielsch, RD. Thulasiraj, R. Krishnadas, and R. Ramakrishnan. Risk factors for age related cataract in a rural population of southern india: the aravind comprehensive eye study. *British journal of ophthalmology*, 88(8):989, 2004.

[26] S. Krishnaiah, K. Vilas, B. R. Shamanna, G. N. Rao, R. Thomas, and D. Balasubramanian. Smoking and its association with cataract: results of the andhra pradesh eye disease study from india. *Investigative ophthalmology & visual science*, 46(1):58, 2005.

[27] T. N. Kim, J. E. Lee, E. J. Lee, J. C. Won, J. H. Noh, K. S. Ko, B. D. Rhee, and D. J. Kim. Prevalence of and factors associated with lens opacities in a korean adult population with and without diabetes: the 2008–2009 korea national health and nutrition examination survey. *PLoS One*, 9(4):e94189, 2014.

[28] RJW. Truscott. Age-related nuclear cataract—oxidation is the key. *Experimental eye research*, 80(5):709, 2005.

[29] L. Guo, J. J. Yang, L. Peng, J. Li, and Q. Liang. A computer-aided healthcare system for cataract classification and grading based on fundus image analysis. *Computers in Industry*, 69:72, 2015.

[30] Y. Xu, X. Gao, S. Lin, D. W. K. Wong, J. Liu, D. Xu, C. Y. Cheng, C. Y. Cheung, and T. Y. Wong. Automatic grading of nuclear cataracts from slit-lamp lens images using group sparsity regression. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 8150, page 468, Nagoya, Japan, 2013. Springer.

[31] H. Al Hajj, M. Lamard, P. H. Conze, S. Roychowdhury, X. Hu, G. Maršalkaitė, O. Zisimopoulos, M. A. Dedmari, F. Zhao, and J. Prellberg. Cataracts: Challenge on automatic tool annotation for cataract surgery. *Medical image analysis*, 52:24, 2019.

[32] R. Pararajasegaram. Community eyehealth. *Community Eye Health*, 15(44), 2002.

[33] G. Davis. The evolution of cataract surgery. *Missouri medicine*, 113(1):58, 2016.

[34] J. B. Alvarez, G. Kintz, D. Mintz, and S. Wong. Methods for robotic assisted cataract surgery, August 18 2020. US Patent 10,744,035.

[35] R. Raskar, V. Pamplona, E. Passos, and J. Zizka. Methods and apparatus for cataract detection and measurement, June 10 2014. US Patent 8,746,885.

[36] W. Huang, K. L. Chan, H. Li, J. H. Lim, J. Liu, and T. Y. Wong. A computer assisted method for nuclear cataract grading from slit-lamp images using ranking. *IEEE Transactions on Medical Imaging*, 30(1):94, 2010.

[37] L. T. Chylack, J. K. Wolfe, D. M. Singer, M. C. Leske, M. A. Bullimore, I. L. Bailey, J. Friend, D. McCarthy, and S. Y. Wu. The lens opacities classification system iii. *Archives of ophthalmology*, 111(6):831, 1993.

[38] R. Srivastava, X. Gao, F. Yin, D. W. Wong, J. Liu, C. Y. Cheung, and T. Y. Wong. Automatic nuclear cataract grading using image gradients. *Journal of Medical Imaging*, 1(1):014502, 2014.

[39] M. Caixinha, E. Velte, M. Santos, and J. B. Santos. New approach for objective cataract classification based on ultrasound techniques using multiclass svm classifiers. In *2014 IEEE International Ultrasonics Symposium*, page 2402, Chicago, IL, 2014. IEEE.

[40] H. E. Gali, R. Sella, and N. A. Afshari. Cataract grading systems: a review of past and present. *Current Opinion in Ophthalmology*, 30(1):13, 2019.

[41] J. H. L. Goh, Z. W. Lim, X. Fang, A. Anees, S. Nusinovici, T. H. Rim, C. Y. Cheng, and Y. C. Tham. Artificial intelligence for cataract detection and management. *The Asia-Pacific Journal of Ophthalmology*, 9(2):88, 2020.

[42] D. Peterson, P. Ho, and J. Chong. Detecting cataract using smartphone. *Investigative Ophthalmology & Visual Science*, 61(7):474, 2020.

[43] S. Hu, H. Wu, X. Luan, Z. Wang, M. Adu, X. Wang, C. Yan, B. Li, K. Li, Y. Zou, et al. Portable handheld slit-lamp based on a smartphone camera for cataract screening. *Journal of ophthalmology*, 2020, 2020.

[44] H. RM. Tawfik, R. AK. Birry, and A. A. Saad. Early recognition and grading of cataract using a combined log gabor/discrete wavelet transform with ann and svm. *International Journal of Computer and Information Engineering*, 12(12):1038, 2018.

[45] L. Zhang, J. Li, H. Han, B. Liu, J. Yang, Q. Wang, et al. Automatic cataract detection and grading using deep convolutional neural network. In *2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)*, page 60, Calabria, Italy, 2017. IEEE.

[46] V. Lakshminarayanan, J. Zelek, and A. McBride. Smartphone science" in eye care and medicine. *Optics and Photonics News*, 26(1):44, 2015.

[47] W. Huang, H. Li, K. L. Chan, J. H. Lim, J. Liu, and T. Y. Wong. A computer-aided diagnosis system of nuclear cataract via ranking. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 5762, page 803, Imperial Coll, London, England, 2009. Springer.

[48] H. Li, J. H. Lim, J. Liu, D. Wing, K. Wong, and T. Y. Wong. Feature analysis in slit-lamp image for nuclear cataract diagnosis. In *2010 3rd International Conference on Biomedical Engineering and Informatics*, page 253. IEEE, 2010.

[49] S. Fan, C. R. Dyer, L. Hubbard, and B. Klein. An automatic system for classification of nuclear sclerosis from slit-lamp photographs. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 2878, page 592. Springer, 2003.

[50] S. Hu, X. Wang, H. Wu, X. Luan, P. Qi, Y. Lin, X. He, and W. He. Unified diagnosis framework for automated nuclear cataract grading based on smartphone slit-lamp images. *IEEE Access*, 8:174169, 2020.

[51] X. Liu, J. Jiang, K. Zhang, E. Long, J. Cui, M. Zhu, Y. An, J. Zhang, Z. Liu, and Z. Lin. Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network. *PloS one*, 12(3):e0168606, 2017.

[52] S. Hu, X. Wang, H. Wu, X. Luan, P. Qi, Y. Lin, X. He, and W. He. Unified diagnosis framework for automated nuclear cataract grading based on smartphone slit-lamp images. *IEEE Access*, 8:174169, 2020.

[53] H. Yazu, E. Shimizu, S. Okuyama, T. Katahira, N. Aketa, R. Yokoiwa, Y. Sato, Y. Ogawa, and H. Fujishima. Evaluation of nuclear cataract with smartphone-attachable slit-lamp device. *Diagnostics*, 10(8):576, 2020.

[54] J. Song, Z. Chi, and J. Liu. A robust eye detection method using combined binary edge and intensity information. *Pattern Recognition*, 39(6):1110, 2006.

[55] Z. Zhu, K. Fujimura, and Q. Ji. Real-time eye detection and tracking under various light conditions. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, page 139, New Orleans,Louisiana, 2002. Association for Computing Machinery.

[56] A. Haro, M. Flickner, and I. Essa. Detecting and tracking eyes by using their physiological properties, dynamics, and appearance. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, volume 1, page 163, Hilton Head,Island, SC, 2000. IEEE.

[57] C. H. Morimoto, D. Koons, A. Amir, and M. Flickner. Pupil detection and tracking using multiple light sources. *Image and vision computing*, 18(4):331, 2000.

[58] N. N. San and N. Aye. Eye detection system using orientation histogram. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 2(4), 2013.

[59] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE transactions on pattern analysis and machine intelligence*, 15(10):1042, 1993.

[60] D. J. Beymer. Face recognition under varying pose. In *1994 IEEE Computer-Society Conference on Computer Vision and Pattern Recognition, Proceedings*, page 756, Seattle, WA, 1994.

[61] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *1994 IEEE computer society conference on Computer Vision and Pattern Recognition, Proceedings*, page 84, Seattle, WA, 1994. IEEE.

[62] A. L. Yuille, P. W. Hallinan, and D. S. Cohen. Feature extraction from faces using deformable templates. *International journal of computer vision*, 8(2):99, 1992.

[63] G. Chow and X. Li. Towards a system for automatic facial feature detection. *Pattern Recognition*, 26(12):1739, 1993.

[64] T. Rajpathak, R. Kumar, and E. Schwartz. Eye detection using morphological and color image processing. In *Proceeding of Florida Conference on Recent Advances in Robotics*, page 1, Jupiter, Florida, 2009. College of Engineering and Computer Science, Florida Atlantic University.

[65] X. Gao, S. Lin, and T. Y. Wong. Automatic feature learning to grade nuclear cataracts based on deep learning. *IEEE Transactions on Biomedical Engineering*, 62(11):2693, 2015.

[66] H. Li, J. H. Lim, J. Liu, P. Mitchell, A. G. Tan, J. J. Wang, and T. Y. Wong. A computer-aided diagnosis system of nuclear cataract. *IEEE Transactions on Biomedical Engineering*, 57(7):1690, 2010.

[67] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1137–1149, 2016.

[68] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, page 770, Seattle, WA, 2016. IEEE.

[69] R. Supriyanti, H. Habe, M. Kidode, and S. Nagata. A simple and robust method to screen cataracts using specular reflection appearance. In *Medical Imaging 2008: Computer-Aided Diagnosis*, volume 6915, page 69152Z, San Diego, CA, 2008. International Society for Optics and Photonics.

[70] R. Supriyanti, H. Habe, M. Kidode, and S. Nagata. Cataract screening by specular reflection and texture analysis. *Communications of SIWN*, 6:59, 2009.

[71] R. Supriyanti, H. Habe, M. Kidode, and S. Nagata. Extracting appearance information inside the pupil for cataract screening. In *IAPR Conference on Machine Vision Application.*, page 342, Tokyo, Japan, 2009.

[72] R. Supriyanti, H. Habe, M. Kidode, and S. Nagata. Compact cataract screening system: Design and practical data acquisition. In *International Conference on Instrumentation, Communication, Information Technology, and Biomedical Engineering 2009*, page 96, Bandung,Indonesia, 2009. IEEE.

[73] R. Supriyanti and Y. Ramadhani. The achievement of various shapes of specular reflections for cataract screening system based on digital images. In *International Conference on Biomedical Engineering and Technology (ICBET).*, page 75, Kuala Lumpur, Malaysia, 2011.

[74] M. A. U. Patwari, M. D. Arif, Md. N. A. Chowdhury, A. Arefin, and Md. I. Imam. Detection, categorization, and assessment of eye cataracts using digital image processing. In *The First International Conference on Interdisciplinary Research and Development, Thailand*, page 22.1, Bangkok, Thailand, 2011. The Interdisciplinary Network of the Royal Institute of Thailand.

[75] J. Nayak. Automated classification of normal, cataract and post cataract optical eye images using svm classifier. In *Proceedings of the World Congress on Engineering and Computer Science*, volume 1, page 23, San Francisco, CA, 2013.

[76] LY. Wong, EYK. Ng, and JS. Suri. Automatic identification of anterior segment eye abnormality. *Irbm*, 28(1):35, 2007.

[77] R. U. Acharya, W. Yu, K. Zhu, J. Nayak, T. C. Lim, and J. Y. Chan. Identification of cataract and post-cataract surgery optical images using artificial intelligence techniques. *Journal of medical systems*, 34(4):619, 2010.

[78] YN. Fuadah, AW. Setiawan, T. LR. Mengko, et al. Mobile cataract detection using optimal combination of statistical texture analysis. In *2015 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME)*, page 232, Institut Teknologi Bandung, Bandung, Indonesia, 2015. IEEE.

[79] Y. N. Fuadah, A. W. Setiawan, and TLR. Mengko. Performing high accuracy of the system for cataract detection using statistical texture analysis and k-nearest neighbor. In *2015 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, page 85, Surabaya, Indonesia, 2015. IEEE.

[80] Z. Qiao, Q. Zhang, Y. Dong, and J. J. Yang. Application of svm based on genetic algorithm in classification of cataract fundus images. In *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, page 1, Beihang Univ, Beijing,China, 2017. IEEE.

[81] W. S. Noble. What is a support vector machine? *Nature biotechnology*, 24(12):1565, 2006.

[82] H. Shen, H. Hao, L. Wei, and Z. Wang. An image based classification method for cataract. In *2008 International Symposium on Computer Science and Computational Technology*, volume 1, page 583, Shanghai, China, 2008. IEEE.

[83] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345, 2009.

[84] R. Sarki, K. Ahmed, and Y. Zhang. Early detection of diabetic eye disease through deep learning using fundus images. *EAI Endorsed Transactions on Pervasive Health and Technology*, 6(22), 2020.

[85] C. Shorten and T. M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, 2019.

[86] N. Boyko, O. Basystiuk, and N. Shakhovska. Performance evaluation and comparison of software for face recognition, based on dlib and opencv library. In *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, page 478, Lviv, Ukraine, 2018. IEEE.

[87] F. AthishMon, N. Narayanan, and K. Suthendran. Recognizing spontaneous emotion from the eye region under different head poses. *International Journal of Pure and Applied Mathematics*, 118(8):257, 2018.

[88] C. Galdi, L. Younes, C. Guillemot, and J. L. Dugelay. A new framework for optimal facial landmark localization on light-field images. In *2018 IEEE Visual Communications and Image Processing (VCIP)*, page 1, Taichung, Taiwan, 2018. IEEE.

[89] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: Database and results. *Image and vision computing*, 47:3, 2016.

[90] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu. A survey on deep transfer learning. In *International conference on artificial neural networks*, volume 11141, page 270, Rhodes, Greece, 2018. Springer.

[91] J. Brownlee. Transfer learning in keras with computer vision models. `https://machinelearningmastery.com/how-to-use-transfer-learning-when-developing-convolutional-neural-network-models/`, Accessed = 2020-11-11, May 2019.

[92] H. Ku and W. Dong. Face recognition based on mtcnn and convolutional neural network. *Frontiers in Signal Processing*, 4(1):37, 2020.

[93] K. Boyd. Cataract diagnosis and treatment. `https://www.aao.org/eye-health/diseases/cataracts-treatment`, Accessed = 2020-11-11, October 2019.

[94] J. E. Birren, R. C. Casperson, and J. Botwinick. Age changes in pupil size. *Journal of Gerontology*, 5(3):216, 1950.

[95] M. Guillon, K. Dumbleton, P. Theodoratos, M. Gobbe, C. B. Wooley, and K. Moody. The effects of age, refractive status, and luminance on pupil size. *Optometry and vision science*, 93(9):1093, 2016.

APPENDICES

This chapter provides a complementary explanation for VGG-16 , ResNet and Inception as three deep convolutional network architecture designs.

.1 VGG-16

By the developments made by the proposed networks for ImageNet Large Scale Visual Recognition Challenge (ILSVRC) classification, CNNs made another jump in the performance and were starting to get deeper and deeper. The most straightforward way was increasing the number of layers and the network size. VGG-16 was one of the VGG (Visual Geometry Group) inventions, which consists of 13 convolutional and three fully-connected layers. Similar to AlexNet, they are carrying the Relu activation function. Figure 1 illustrates the architecture of the VGG-16 [4].

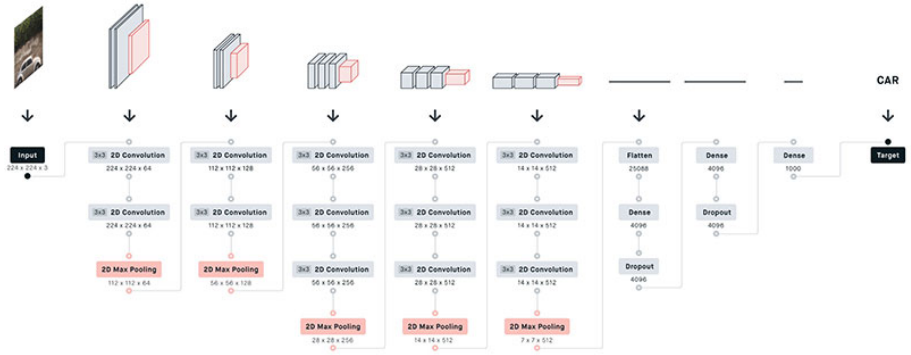

Figure 1: The architecture of the VGG-16 network [4].

In continuous to the previous networks, VGG-16 stacked the layers together but utilized a smaller size of filters (2×2 and 3×3). It has 138M parameters and needs about 500MB

of storage memory. As the next architecture, the VGG group designed a deeper variant called VGG-19 [4].

.2 ResNet

In 2015, the ILSVRC challenge winner was Kaiming He et al., who developed the residual network (ResNet). The new proposed network could achieve an astounding top-5 error rate under 3.6%. For this purpose, an extremely deep CNN composed of 152 layers was utilized. As a novelty, the skip connections were the key to train such a deep network. The feeding signal into the layer was also added to the layer's output placed a bit higher up the stack. Figure 2 illustrates the architecture of the ResNet [4].

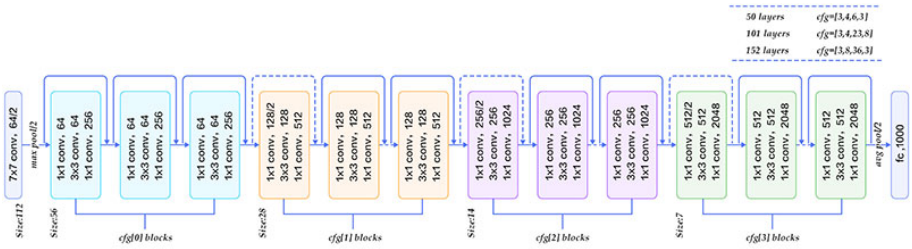

Figure 2: The architecture of the ResNet network [4].

ResNets are created out of something called a residual block. Figure 3 presents the residual block in ResNet network [4].

.3 Inception v3

Inception v3 is capable of detecting objects at different scales. It implements convolution operation, using different sized kernels, to capture variations at different scales. Moreover, a deep network may delete some features which could be useful for decoding, on the other hand shallow networks may not learn high level abstractions, while Inception module allows network choose proper convolutional operations and hamper the effects of network depth. In fact, this module lets the CNN to automatically choose the right filter size (in some few layers). Figure 4 illustrates the architecture of the Inception v3 network.

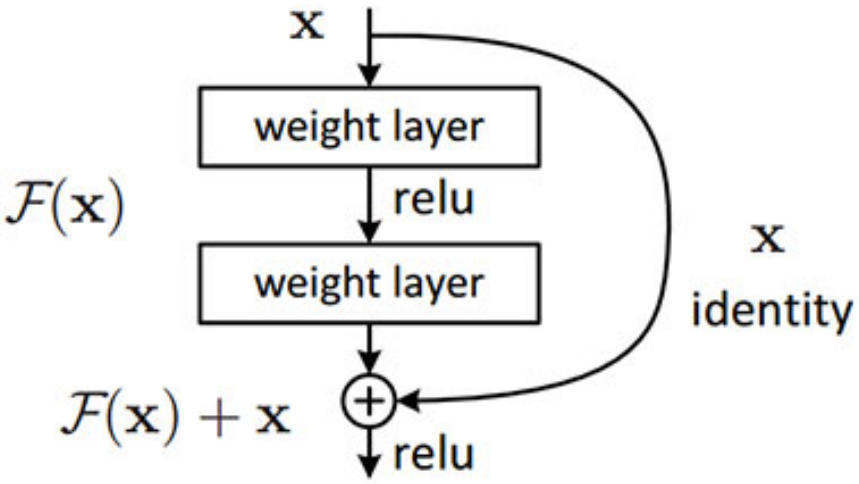

Figure 3: A residual block [4].

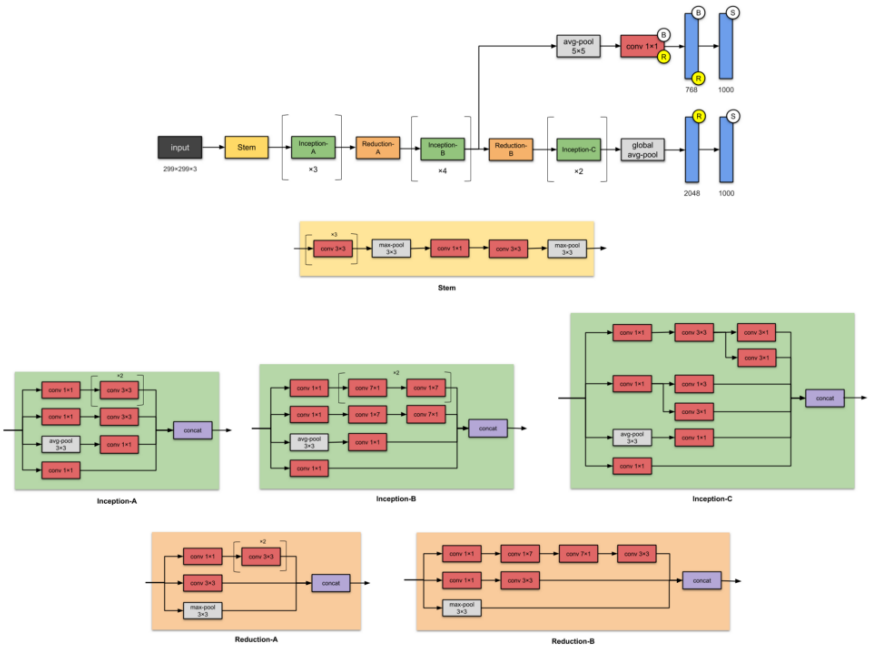

Figure 4: Inception v3 network [5].