7-28-2022 3:00 PM

Effectiveness of Machine Learning Classifiers for Cataract Screening

Ronald Cheung, *The University of Western Ontario*

Supervisor: Malvankar, Monali, *The University of Western Ontario*
Joint Supervisor: Thind, Amardeep, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Epidemiology and Biostatistics

Abstract

Cataract is the leading cause of blindness and vision loss globally. The implementation of artificial intelligence (AI) in the healthcare industry has been on the rise in the past few decades and machine learning (ML) classifiers have shown to be able to diagnose patients with cataracts. A systematic review and meta-analysis were conducted to assess the diagnostic accuracy of these ML classifiers for cataracts currently published in the literature. Retrieved from eight articles, the pooled sensitivity was 94.8% and the specificity was 96.0% for adult cataracts. Additionally, an economic analysis was conducted to explore the cost-effectiveness of implementing ML to diagnostic eye camps in rural Nepal compared to traditional diagnostic eye camps. There was a total of 22,805 patients included in the decision tree, and the ML-based eye camp was able to identify 31 additional cases of cataracts, and 2546 additional cases of non-cataract.

Keywords

Systematic review, Meta-analysis, Cost-effectiveness Analysis, Cataract, Machine Learning, Artificial intelligence, Diagnostic Accuracy, Ophthalmology, Low-to-middle income country, Global Health

Summary for Lay Audience

Cataract is an eye disease that many older adults get. A cataract is a buildup of cloudiness in the human eye lens that can result in blurry and reduced vision. Fortunately, through early and proper screening procedures, cataracts can easily be detected, and cataract surgery can be performed to gain back vision. There has been an increasing use and implementation of artificial intelligence (AI) in the healthcare field and machine learning (ML) which is a subset of AI. In the field of ophthalmology, there are many developments for the use of ML classifiers that can automatically detect eye diseases (such as cataracts) by processing images of the eye through a computer algorithm.

In this thesis, a systematic review and meta-analysis were conducted to assess the diagnostic accuracy of current machine learning classifiers for cataracts in both published databases and unpublished literature. A total of 21 studies were included in the qualitative review, and a total of nine studies were included for the quantitative analysis. From the quantitative analysis, there was observed to be high diagnostic accuracy for identifying true cataract cases and true non-cataract cases, these values are known as sensitivity and specificity, respectively. For adult cataracts, there was a 94.8% sensitivity and 96.0% specificity.

Utilizing these results from the meta-analysis, a cost-effective analysis was conducted to test the economic feasibility of a ML cataract screening program to be implemented in a rural region. In Nepal, rural Nepalis may have access to temporary village-level primary eye care centres known as “diagnostic-screening and treatment camps (eye camps)”. The objective of this second study was to conduct a cost-effectiveness analysis of the theoretical implementation of a ML-based cataract screening eye camp in rural Nepal in order to assess

if this new program is superior to the traditional eye camps. There was a total of 22,805 patients in each arm of the decision tree, and the ML-based eye camp could identify 31 additional cases of cataracts, and 2,546 additional cases of non-cataract. This suggested that the ML-based eye camp was a more cost-effective method than the traditional eye camp in rural Nepal.

Co-Authorship Statement

Chapter 03: *Diagnostic accuracy of machine learning classifiers for cataracts: a systematic review and meta-analysis*

Co-authorship: Cheung R, So S, Malvankar-Mehta MS

RC was responsible for the conceptualization of the research topic, designing and writing the protocol, conducting the database and grey literature search, screening the studies, conducting the risk of bias assessment, curating the data, extraction of data, analyzing and interpreting the data, interpreting the results, writing, and editing the paper. SS was responsible for the screening of the studies, conducting the risk of bias assessment, curating the data, reviewing, and editing the paper. MM-M was responsible for the conceptualization of the research topic, designing the review protocol, analyzing and interpreting the data, validation of results, writing and editing the paper.

Chapter 04: *The implementation of a machine learning-based cataract screening program in rural Nepal: a cost-effectiveness analysis*

Co-authorship: Cheung R, Li B, Thind A, Malvankar-Mehta MS

RC was responsible for the conceptualization of the research topic and the decision tree model, cost and effectiveness data collection, data analysis and interpretation of the study results, and writing and editing the paper. BL and AT were responsible for the conceptualization of the research topic, interpretation of study results, revision and editing of the paper, and gave content feedback throughout the study. MM-M was responsible for the conceptualization of the research topic, decision tree, interpretation of study results, and writing and editing the paper.

Dedication

This thesis is fully dedicated to the loving memory of my beloved little sister, Rhonda, who I will miss and love every day.

Acknowledgments

I first wanted to say thank you to the Department of Epidemiology and Biostatistics at Western University for providing me with a wonderful education throughout my undergraduate and master's degree. I have called this department my home for the past three years and I am so grateful for all my amazing encounters with our faculty and staff.

I would also like to first thank my co-supervisor, Dr. Amardeep Thind, for his constant support and engagement with my thesis. I would like to also acknowledge my advisory committee member, Dr. Bo Li, for his insight on my thesis. They have gladly navigated me through all the hurdles I have faced, and they have provided me with exceptional feedback on my thesis.

I wanted to also extend a big thank you to my close friends, classmates, housemates, and parents who have supported me. Thank you for all the discussion and encouragement that you all have given me over the past year, and for being there when I needed it the most.

Finally, I would like to thank and give credit to Dr. Monali Malvankar for my academic and research achievements. Dr. Malvankar is one of the most compassionate and empathetic professors there is at Western University, and she was instrumental in the completion of my thesis. Dr. Malvankar has always been supportive of my professional career goals, and she has given me countless opportunities to conduct research and develop my professional skills.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

AI	Artificial intelligence
ARDA	Automated Retinal Disease Assessment
CCPCMOH	Childhood Cataract Program of Chinese Ministry of Health
CBA	Cost-benefit analysis
CEA	Cost-effectiveness analysis
CUA	Cost-utility analysis
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CNN	Convolutional neural network
DALY	Disability-adjusted life years
DLS	Deep learning system
DOR	Diagnostic odds ratio
DR	Diabetic retinopathy
fn	False negative
fp	False positive
HSROC	Hierarchical summary receiver operating characteristic
ICER	Incremental cost-effectiveness ratio
IOP	Intraocular pressure
LMIC	Low-to-middle income country
LOCS	Lens opacity classification system
LR	Likelihood ratio
LY	Life years
MeSH	Medical Subject Headings
ML	Machine learning
OA	Ophthalmic assistants
OHIP	Ontario Health Insurance Plan
OOP	Out-of-pocket
OT	Ophthalmic technicians
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSA	Probabilistic sensitivity analysis
QALY	Quality-adjusted life years
QUADAS	Quality of diagnostic accuracy studies
SROC	Summary receiver operating characteristic
SVM	Support vector machine
tn	True negative
tp	True positive
UV	Ultra-violet
WHO	World Health Organization
WTP	Willingness-to-pay

CHAPTER 1

1 Introduction

According to a report by the World Health Organization (WHO) in 2010, cataracts account for approximate 50% of the world's first cause of blindness.[1,2] In fact, the rate and prevalence of cataracts globally and nationally is on the rise due to the world's aging population, thus making cataracts a health priority and a disease of concern for health and aging. [3–5] There are significant social and economic costs associated with vision loss for both the patient population and the healthcare system, and new technologies are emerging to help meet the high patient demands.[1,6]

Within the field of ophthalmology, the rise of artificial intelligence (AI) and machine learning (ML) has grown substantially in the past decade.[6,7] Many novel algorithms and applications of AI are currently being used in routine clinical practice that aid ophthalmologists and healthcare practitioners with the diagnosis and grading of certain eye diseases. Various technology companies, such as Google and IBM, have invested in the growth and development of machine learning in ophthalmology, and most of the research has been conducted on diabetic retinopathy (DR).[8] The success of these technologies has given the potential and hope for researchers to apply similar techniques to other common eye diseases such as glaucoma, age-related macular degeneration, and cataracts.[9–11]

For cataracts specifically, research in the use of AI and ML has shown the potential for these algorithms to be used for multiple purposes throughout the course of a patient's cataracts diagnosis to cataracts treatment. Mainly, ML classifiers have been used to screen cataracts through training and validating fundus or slit-lamp eye images to provide a fast and accurate diagnosis.[12–14] Additionally, AI-based methods have also been used to create intraocular lens power calculations as part of the cataracts surgery process to determine a predictive error post-cataracts surgery.[15,16] However, there are gaps in the literature on the aggregated diagnostic accuracy of these ML diagnostic programs, and its cost-effectiveness compared to in-person screening procedures. High diagnostic accuracy is important, but cost is also a substantial part of the decision-making process if these algorithms are to be routinely implemented in hospital settings, ophthalmology clinics, or rural diagnostic and screening eye camps. It is important to assess the current literature and body of evidence on cataracts, the cataracts screening and diagnosis process, and the applications of AI or ML in cataracts care and management.

1.1 Structure of thesis

This thesis is written in the integrated article format within the standards of Western University School of Graduate and Postdoctoral studies. Chapter 02 is a literature review on the background and current knowledge of cataracts, and machine learning applications in healthcare and ophthalmology. The literature review also discusses the methodology of a systematic review and meta-analysis used in Chapter 03 and the methodology of a cost-effectiveness analysis used in Chapter 04. Chapter 02 also includes the thesis rationale and objectives.

The thesis consists of two manuscripts. Chapter 03 is comprised of the first manuscript titled "*The diagnostic accuracy of machine learning classifiers for cataracts: a systematic review and meta-analysis*". Chapter 04 is comprised of the second manuscript titled "*The implementation of a machine learning-based cataract screening program in rural*

Nepal: a cost-effectiveness analysis". Chapter 05 includes an integrated discussion of the results of the thesis.

1.2 References

1. Pascolini D, Mariotti SP. Global estimates of visual impairment: 2010. British Journal of Ophthalmology [Internet]. 2012 May 1;96(5):614 LP – 618. Available from: http://bjo.bmj.com/content/96/5/614.abstract

2. WHO. Blindness and vision impairment [Internet]. 2021 [cited 2021 Nov 17]. Available from: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

3. Hatch W v, Campbell E de L, Bell CM, El-Defrawy SR, Campbell RJ. projecting the growth of cataract surgery during the next 25 years. Archives of Ophthalmology [Internet]. 2012 Nov 1;130(11):1479–81. Available from: https://doi.org/10.1001/archophthalmol.2012.838

4. Bourne R, Steinmetz JD, Flaxman S, Briant PS, Taylor HR, Resnikoff S, et al. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study. The Lancet Global Health [Internet]. 2021;9(2):e130–43. Available from: https://www.sciencedirect.com/science/article/pii/S2214109X20304253

5. Steinmetz JD, Bourne RRA, Briant PS, Flaxman SR, Taylor HRB, Jonas JB, et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. The Lancet Global Health [Internet]. 2021 Feb 1;9(2):e144–60. Available from: https://doi.org/10.1016/S2214-109X(20)30489-7

6. Wiens J, Shenoy ES. Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology. Clinical Infectious Diseases [Internet]. 2018 Jan 6;66(1):149–53. Available from: https://doi.org/10.1093/cid/cix731

7. Lamminen H, Voipio V, Ruohonen K, Uusitalo H. Telemedicine in ophthalmology. Acta Ophthalmologica Scandinavica [Internet]. 2003 Apr 1;81(2):105–9. Available from: https://doi.org/10.1034/j.1600-0420.2003.00045.x

8. Google Health. Seeing potential [Internet]. 2019 [cited 2021 Nov 15]. Available from: https://about.google/intl/ALL_us/stories/seeingpotential/?utm_source=about.google&utm_medium=social&utm_campaign=copy-link

9. Cheung R, Chun J, Sheidow T, Motolko M, Malvankar-Mehta MS. Diagnostic accuracy of current machine learning classifiers for age-related macular degeneration: a systematic review and meta-analysis. Eye [Internet]. 2021; Available from: https://doi.org/10.1038/s41433-021-01540-y

10. Murtagh P, Greene G, O'Brien C. Current applications of machine learning in the screening and diagnosis of glaucoma: a systematic review and meta-analysis. International journal of ophthalmology [Internet]. 2020 Jan 18;13(1):149–62. Available from: https://pubmed.ncbi.nlm.nih.gov/31956584

11. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA [Internet]. 2016;316(22):2402–10. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med13&NEWS=N&AN=27898976

12. Zheng J, Guo L, Peng L, Li J, Yang J, Liang Q. Fundus image based cataract classification. In: 2014 IEEE International Conference on Imaging Systems and Techniques (IST) Proceedings. 2014. p. 90–4. Available from: https://ieeexplore.ieee.org/document/6958452?arnumber=6958452

13. Ran J, Niu K, He Z, Zhang H, Song H. Cataract detection and grading based on combination of deep convolutional neural network and random forests. In: 2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC) [Internet]. 2018. p. 155–9. Available from: 10.1109/ICNIDC.2018.8525852

14. Wu X, Huang Y, Liu Z, Lai W, Long E, Zhang K, et al. Universal artificial intelligence platform for collaborative management of cataracts. The British Journal of Ophthalmology [Internet]. 2019;103(11 PG-1553–1560):1553–60. Available from: https://bjo.bmj.com/content/103/11/1553

15. Melles RB, Kane JX, Olsen T, Chang WJ. Update on intraocular lens calculation formulas. Ophthalmology. 2019 Sep;126(9):1334–5. Available from: https://www.aaojournal.org/article/S0161-6420(19)30284-2/fulltext

16. Olsen T. Calculation of intraocular lens power: a review. Acta Ophthalmologica Scandinavica [Internet]. 2007 Aug 1;85(5):472–85. Available from: https://doi.org/10.1111/j.1600-0420.2007.00879.x

CHAPTER 2

2 Literature Review, Thesis Rationale, and Thesis Objectives

2.1 Literature Review

2.1.1 Natural History of Cataracts

A cataract is the opacification of the lens in the human eye which results in poor visual acuity and transparency. [1] Cataract can occur in one eye or both eyes.[1] The opacity of the lens is caused by oxidative stress, and it primarily affects the growth and development of the lens epithelial cells.[2] The lens is located positionally behind the iris and in front of the vitreous body and retina. The lens helps focus light into the eye to produce sharp images, but as the cataract develops, it blocks the light passing through the lens and prevents a sharp image from reaching the retina.[1] As a result of the cataracts, the patient loses optical clarity and has a clouded vision.[1,3]

Symptoms

Common symptoms that patients with cataracts experience can include clouded, blurred, or dimmed vision.[3] A visual decline that can span over weeks, months, or years. Patients may be sensitive to light and glare, and halos can form around lights.[1,4,5] Additionally, changed experience in vision can result in the yellowing of images and decreased colour intensity. Corrective glasses do not help improve eyesight if the cataract is left untreated.

Classification of Cataracts

Cataracts is often categorized by their cause-types which include age-related cataracts, congenital cataracts, and cataracts secondary to other causes.[3,5,6] First, age-related cataracts can be divided into three types: nuclear, cortical, and posterior subcapsular.[1] Often, patients can present with just one type or a combination of types of age-related cataracts. Nuclear cataract occurs when new fiber layers from the lens epithelial cells migrate towards the lens equator, and the lens nucleus becomes compressed (nuclear sclerosis) which creates an opacification. Cortical cataract starts at the cortex of the lens where cortical spokes can develop and cause discrete opacities.[7] Cortical cataract can be located posteriorly or anteriorly of the lens, and it is often wedged-shaped.[8] Posterior subcapsular cataract is in the axial posterior cortex where plaques and deposits develop.[1,9,10]

In addition to age-related cataracts which are the most common, pediatric cataracts is also prevalent in infant populations. In congenital cataracts, the lens opacity would have been present at birth, but then manifested and developed within one year of the infant's life.[11] Approximately one-third of patients with pediatric cataracts are due to inheritance. Pediatric cataracts can be classified as unilateral or bilateral cataracts.[12]

2.1.2 Risk Factors

Cataractogenesis – meaning the process of cataract formation – may be caused and influenced by a multitude of risk factors. The current evidence of other direct risk factors for cataracts can range from limited to strong evidence. These risk factors can be modifiable or non-modifiable.[9,13,14]

Non-Modifiable Risk Factors

Age - The most common form of the development of cataracts is age-related, which makes this a disease a priority in the health and aging context.[1] The onset of cataracts often begins at the age of 45-50 years because of oxidative stress, solubilization, and cross linking by the lens fibers.[2] There is conclusive evidence that age is a personal risk factor for nuclear, cortical, and posterior subcapsular cataract.

Sex - Many studies have shown that females are at a greater risk for cataracts than males, and females experience a higher cataract burden.[15]

Genetics - Genetic effects are important and contributing factors to the development of cortical cataracts. Genetic modelling suggests that additive and dominant genes can suggest the causation of cortical cataracts based on the Twin Eye Study.[16] Additionally, genetic factors can account for up to 50% of the variation in severity in nuclear cataract cases.

Modifiable Risk Factors

Diabetes - Patients with diabetes have an increased risk for developing cataracts. There is increased risk specifically for cortical and posterior subcapsular cataracts, but there is no significant association with nuclear sclerosis.[17] Increased glucose levels in the lens are converted into sorbitol which can cause the lens to be opaquer and cloudier. Additionally, non-enzymatic damage to the lens protein (glycation) may be involved in the development of cataracts.[3]

Alcohol intake - Strong evidence of increased risk for cataracts is found for individuals who are heavy drinkers. In a meta-analysis conducted by Gong et al., heavy alcohol consumption significantly increases the risk of age-related cataracts, though moderate consumption of alcohol has been revealed to have some protective effects against cataracts.[18] Heavy alcohol consumption was defined as consuming more than 20g of alcohol in a day.

Trauma - Direct, blunt trauma to the eye is another cause of cataracts due to damaged lens fibres.[3] Blunt trauma can cause the eye to swell, and fibres in the lens to thicken which causes increased opacity. Other sources of traumatic cataracts can be due to infrared lights, electric sparks, or head injuries.[9,19]

Hypertension - A meta-analysis by Yu et al. found that high blood pressure can increase the risk of cataracts by 8-28%.[20] Studies have suggested that the link between hypertension and cataract development is in part due to anti-hypertension medications that can disturb electrolyte balance around the lens fiber membrane. Additionally, hypertension may also cause lens capsules to have conformational changes which interferes with potassium ion transport.[21,22]

Ultra-violet (UV) ray exposure - Another risk factor for cataracts is exposure to UV rays over time. UV light can damage lens proteins and cells in the lens, and it can continue to be damaged by oxidative stress. The Canadian Ophthalmological Society recommends that people and children at a young age should get into the habit of wearing sunglasses as preventative measures.[23]

Toxins and chemicals in smoke can increase the risk of cataract development and increase opacities are reported. There is an increased odds of 1.41 for nuclear cataracts with individuals who have smoked in their life.[24]

In pediatric cataracts, there is also a plethora of reasons for the development of cataracts in infants. As noted previously, the main driver of pediatric cataracts is due to hereditary factors and the inheritance of genetic factors that cause the opacification of the lens. [11] However, this claim is still scarcely researched, even if it is the commonly agreed stance.

2.1.3 Epidemiology of cataracts

Global Epidemiology

According to the WHO, there are at least 2.2 billion people in the world who experience near or distance impaired vision, and 94 million of those cases are as a result of cataracts.[25,26] More than 50% of the world's first cause of blindness is due to cataracts and it is the leading cause of blindness. Based on a meta-analysis by Hashemi et al., the pooled prevalence estimate of any cataracts is approximately 17.20% in the world, with nuclear cataract leading with a prevalence of 8.22%, and cortical cataract with 8.05%.[27] Hashemi et al. also found the pooled prevalence estimate of cataracts in females and males was around the same, with females at 33.67%, and males at 32.57%. The authors noted the geographic location (based on the six WHO regions) that had the highest prevalence of cataract was the South-East Asia region, followed by the Western Pacific, and Europe.[27]

Globally, cataract has contributed to 17.7 million disability-adjusted life years (DALY) in adult populations. DALY is a common metric to describe the total number of years lost to disability or premature death.[28] In Southeast Asia alone, the global health burden of cataract vision loss was approximately 125 disability-adjusted life years (DALYs) per 100,000 people — the highest crude DALY rate out of all WHO regions.[29,30] Unfortunately, the DALYs are expected to increase for the cataract population due to the world's aging population, and lack of access to early care. Many people who live in rural or underserviced areas often receive a delayed cataract diagnosis or lack the facilities to receive cataracts surgery, which leads to blindness (a measure of disability).

While cataract is commonly found in adult and elderly patients, pediatric cataract has an estimated prevalence of 4.24 per 10,000 live births, and it is the major causes of childhood blindness.[31]

2.1.4 Clinical Assessment of Cataracts

An early diagnosis of cataracts is often ideal because preventative measures can be immediately taken to delay and slow down the deterioration of vision.[32,33] Correction glasses, anti-glare glasses, or magnifying lenses can be used when cataracts begin to interrupt daily activities before surgical intervention is needed. The standards for the cataract screening procedure with an optometrist or an ophthalmologist will vary based on location due to government policies, insurance coverage, economic status, healthcare costs, and access to care.[6,34] Many factors influence the thoroughness and completeness of an eye examination depending on the availability and demand of such services. The clinical assessment and process of examining a patient's eye in a high-income country will be very different than a patient living in a rural or low-to-middle income country (LMIC) due to the scarcity of resources and trained personnel.[35,36]

The Canadian Ophthalmological Society has published clinical practice guidelines for cataracts surgery which also outlines the suggested ophthalmic evaluation for the diagnosis of cataract.[34] The guidelines for Ontario will be used to illustrate the full procedure of an eye examination for cataracts, although it may be not fully feasible in other under-serviced regions. It is noted that there is no single test or examination that adequately describes the effect that the cataract has on the patient's visual status and functioning ability.[33]

In Ontario, Canada, the Ontario Government has published a "*Quality-Based Procedures Clinical Handbook: Cataract Day Surgery*" which provides guidelines and standards that healthcare professions are to follow in the cataract setting.[37] As part of the assessment and referral pathway in the handbook, annual eye exams should be performed in individuals who are over the age of 65, and anyone with current conditions of diabetes, glaucoma risk, or other eye conditions. It is important note that although everyone should receive annual exams, the current Ontario Health Insurance Plan (OHIP) only insures annual eye exams for people younger than 20 or over 65, if they have one or more ocular conditions, or if their primary healthcare provider necessitates an annual exam.[37] This public insurance coverage is found to be similar in other provinces in Canada, though not all.

To diagnose a patient with cataracts, the ophthalmologist will ask about the patient's medical history and any symptoms that they are experiencing. A patient's assessment will include their current medications and medical conditions, previous ophthalmic surgery, and risk factors that can affect the surgical plan.[6,32,38] Additionally, an eye examination should be conducted by an ophthalmologist, primary care physician, or optometrist. A complete eye exam should include a visual acuity test, slit lamp imaging exam, dilated eye exam, and a tonometry.[6,34,37,39] The full ophthalmic evaluation for the diagnosis and treatment of cataract is displayed in **Table 2.1**.[34]

Visual acuity tests

Visual acuity tests are used to test the patient's ability to discern shapes and details in their vision and to rate the patient's recognition of small details with precision. Eye charts — also known as optotypes — are used to test for visual acuity. If the patient is unable to read the eye charts at any distances, alternatives to the visual acuity test include counting fingers, hand motion, and light perception tests.[40]

Most commonly, the Snellen chart is used for visual acuity tests, and it is widely used at the clinical level. It is especially commonly used for patients with myopia, hyperopia, or astigmatism, and for assessing vision problems in young children.[41,42] The Snellen chart is a simple and effective way for physicians, ophthalmologists, or optometrists to recognize signs of vision loss and diagnosis of cataracts because it also considers the patient's perspective for their visual function.[42,43] The Snellen chart is a multi-letter chart with a variation in number of letters on each line of the chart; this presents the examinee with increasing difficulty to identify the letters. The examinee typically reads the chart 6 meters (20 feet) away, testing one eye at a time.[42,43] The common term "20/20 vision" or "6/6 vision" refers to the patient's ability to see clearly at 20 feet what should normally be seen at that distance.[42]

Despite the Snellen chart's ease and universal use, its limitations include the lack of reproducibility and reliability in its results.[41] There have been articles that criticizes the Snellen chart's failure to test visual acuity at the right distance and under the recommended levels of illumination. As a result, newer charts such as the logMAR (logarithm of the minimum angle of resolution) became available to negate the disadvantages of the Snellen chart.[44,45] The logMAR chart has letters that are of equal legibility, with the same number of letters on each row and uniform letter and row spacing. There is a logarithmic progression in the letter size which ensures test task standardization. The logMAR chart is less used in clinical practice, though it has become the standard in research settings.

Slit-lamp biomicroscope imaging

Slit-lamp imaging is an important element of the eye examination because it captures physical elements of the eye that a visual acuity test cannot determine. Slit-lamp imaging must be done because the ophthalmologist needs to rule out other ocular diseases before diagnosing cataracts and suggesting cataract surgery. Slit-lamp microscopes get their name from the thin sheet of high-intensity light source that focuses and shines into the eye.[32,34] Slit-lamp microscopes are able see the details of the transparent, translucent, and opaque structures of the anterior and posterior segment of the human eye. Slit-lamp microscopes uses a variety of magnifications and angles to observe the patients and ophthalmologists can decide the type of cataracts based on what is exhibited in the lens.[34] In most patients, ophthalmologists can determine if cataracts are responsible for the patient's visual loss by comparing the slit lamp images with the patient's symptoms.

The slit lamp can see what is manifested in the anterior segments of the eye, and examine the lens, vitreous, macula, peripheral retina, and optic nerve through a dilated pupil. For example, cortical cataracts can be diagnosed by observing the formation of vacuoles, clefts, wedges, or lamellar separations.

Tonometry Test

A tonometry test is a procedure to determine the intraocular pressure (IOP) inside the eye.[46] IOP is the fluid pressure of the eye, and it is formed from the balance between aqueous humour formation and outflow on the internal surface area of the anterior eye. The Goldmann applanation tonometry, using the Goldmann equation, is the most used method in the clinical setting. The Goldmann equation uses the aqueous flow rate, aqueous outflow, and episcleral nervous pressure to measure the change in the IOP.[47] This method uses a flat-tipped probe that presses against the surface of the eye; however, other types of methods are available called non-contact tonometry that uses air pressure.[47,48] Pressures between 11 and 21 mmHG are generally considered to be normal. A tonometry test is important in a full ocular examination because an elevated IOP is an indicator for glaucoma. By identifying the IOP, ophthalmologists or optometrists can confirm or rule out glaucoma.

Cataract classification systems

Currently, there are several systems that are commonly used to classify and grade lens opacities. Grading systems are important for clinical and research use, and for the communication between patient and physician.[49] However, the variation in classification systems makes it very difficult for comparative studies to be performed due to the slightly differing definitions and grading systems that are being used. In 1989, West and Taylor called for a standardized method of grading and classifying cataracts, thus the Oxford Clinical Grading System and the early versions of the Lens Opacity Classification System (LOCS) were created.[50,51] Since then, these grading and classification systems have evolved. There are updated versions of the LOCS (LOCS II, and LOCS III) and other grading systems that have emerged including the Johns Hopkins system, and the Wisconsin Cataract Grading System.[52] However, it is noted that the cataract classification system varies from country to country, often influenced by insurance coverage, and health priorities of the country.

In these grading systems, the following elements are considered: anterior clear zone thickness, anterior subcapsular opacity, posterior subcapsular opacity, cortical spoke opacity, water clefts, vacuoles, retro-dots, focal dots, nuclear brunescence and white nuclear scatter.[50,53] The LOCS III was updated in 1993 in order to better capture an early cataracts diagnosis by observing nuclear opalescence and nuclear colour on a scale from 1 to 6, cortical cataracts on a scale from 1 to 5, and posterior subcapsular cataracts on a scale from 1 to 5.[50, 53]

Clinical Assessment of Cataracts in Rural and Low-to-Middle Income Countries

In rural regions, villages, and LMIC, most people do not have the ability or luxury to seek out an eye care provider regularly because the closest eye care centres are in semiurban or urban areas which may be geographically far away from the villages, people may have to pay out of pocket for the healthcare services, and people may be unaware of such services.[36,54]

Additionally, the optometrist and ophthalmologist to general population ratio may be very low and below the standards set by WHO. For example, in Nepal, the ophthalmologist to population ratio is approximately 1:193,900, and 1:791,700 for optometrists.[36] The recommendation set out by WHO is 1:100,000 for both ophthalmologists and optometrists.[25] To remedy many of these concerns in both Nepal, India and other LMICs, temporary village-level primary eye care centres called “eye camps” have been implemented in the past few decades to reach to these rural patients.[35,55] The goal of these eye camps is to eliminate any financial or geographical barriers that many rural patients may have to access their health services in more semiurban settings.

These make-shift and temporary eye camps often employ only ophthalmic assistants, ophthalmic technicians, and/or nurses to operate the eye camps, and very minimal ophthalmic equipment is available for these workers.[56,57] An ophthalmic assistant often conducts the assessment and makes clinical decisions in replacement of an ophthalmologist or optometrist. Therefore, it is not feasible to conduct a full eye examination (as they do by standard in Ontario, Canada) in these eye camps, but rather a simplified process that may include a Snellen chart, a slit-lamp microscope, a pen light, and/or a portable ophthalmoscope.[58] When a rural patient is given a diagnosis of any grade and/or classification of cataract in diagnostic eye camps, they are automatically given a referral to an ophthalmologist.[35] Typically, in the Ontario and Westernized context, a patient may have multiple follow-up appointments with an optometrist or eye specialist over a period before a referral to an ophthalmologist is given.[34]

2.1.5 Artificial Intelligence in Healthcare

The use of AI has revolutionized the way that healthcare can be provided to patients in these rural areas, and it has greatly supported the efforts and developments of tele-medicine and more specifically, tele-ophthalmology.

Evolution of AI healthcare

AI is a branch of computer science that aims to simulate a human's mental process through software programs and learn to solve problems similarly to the human brain.[59] The use of AI has been on the rise in health care and biomedical research in the past decade. Substantive progress in this field of research has been made as healthcare providers, policy makers, and researchers because of the potential for its use in standard practice.[59–61] However, AI is not a completely new technology that has only recently existed. In fact, AI has been used in healthcare in the 1970s in the form of rule-based approaches. This early application of AI was used to interpret electrocardiograms, diagnose certain diseases, and provide simple clinical reasoning and interpretations for hypothesis generation.[60] However, the performance of these systems was limited by their lack of comprehensive medical knowledge, and it required humans to be involved in every decision step. Despite the inefficiencies of the early AI technologies, they provided progress towards a fully autonomous system.[60]

The healthcare sector has benefited from the emergence of AI because the increasing availability of healthcare data has made it possible for AI to flourish in this sector. Large amounts of healthcare data can be used in AI algorithms to assist in making clinical decisions that can be cost and time efficient. AI is not intended to replace human physicians or healthcare providers, but to assist in clinical decisions and replace human judgement in certain areas of healthcare.[59] Ideally, the ultimate goal for AI processes is to have a fully automated clinical system to make decisions and output results, but as these technologies are still emerging, AI can exist in the healthcare setting through conventional decision support systems, or integrative decision support systems.[59,61–63]

AI research areas in the current literature are mostly done in the field of diagnostic imaging, genetics, electrodiagnosis, and physiologic monitoring.[59,61] The leading disease types that have the most research conducted in AI in the literature are neoplasms, nervous system, cardiovascular, and urogenital. Since the 2010s, there has been a stark increase in the number of publications with the AI keyword in the databases.[59] In the past decade, the use and application of AI algorithms have spanned across a plethora of disciplines and fields, notably in ophthalmology.

Since the utilization and implementation of AI in healthcare is still relatively new, there continues to be many ethical and legal concerns surrounding its practice. There exist many issues regarding data privacy, data and algorithm bias that exists in AI which are debated by policy makers and researchers.[64,65] These concerns are further addressed in the integrated discussion (Chapter 05).

Machine Learning

Machine learning (ML) is a sub-set of AI which describes the use of computer algorithms to learn and identify patterns in the data. Through training and validation, these

ML algorithms be given large amounts of healthcare data to perform a specific task.[60] There exist many types of ML algorithms that are regularly used in healthcare applications including support vector machines (SVM), neural networks, logistic regression, random forests, and others. [60] SVM and neural networks are the two most researched and used algorithms.

2.1.6 AI and ML in Ophthalmology

Specifically in the field of ophthalmology, vast amounts of AI research have been conducted by Google Inc for diabetic retinopathy (DR).[66] The Google Health team has developed a DR screening solution in which the team recruited a large team of ophthalmologists to screen through 100,000 retinal scans in order to train their AI algorithm.[66,67] In their study conducted by Gulshan et al., the Google Health research team was able to train and validate an algorithm that had a 98.1% specific and 98% specificity for detecting referable DR.[67] The purpose of this project was to create an AI-based application called the Automated Retinal Disease Assessment (ARDA) to assist clinicians and physicians to screen through retinal images in lower-to-middle income countries (LMIC) such as Thailand and Nepal. The ARDA screening program allows the user to upload a fundus image to the platform, and the application can give an instant analysis of diabetic retinopathy.[67] In a matter of seconds, the algorithm can identify if DR exists in the image and the grade of the DR. The rise of this study gave potential to many researchers within the ophthalmology field to explore the use of AI in diagnosis and other aspects of healthcare delivery.[68,69]

There have been numerous systematic reviews and meta-analyses that have investigated the use of ML classifiers for diagnoses in different eye diseases. In a meta-analysis by Cheung et al. that included 13 studies, they found that ML classifiers were able to detect age-related macular degeneration with a 91.8% sensitivity, and 88.8% specificity.[70] In another meta-analysis conducted by Murtagh et al., the researchers assessed the accuracy of ML screening for glaucoma.[71] Similarly to Cheung et al., they found high accuracy for the screening program with an area under the ROC (AUROC) value of 0.957 for fundal photos, and 0.923 for OCT images.[71] Evidently, there is a lot of research conducted in diagnostic imaging, proving screening programs like ARDA can be translated to other eye diseases.

The use of AI and ML have demonstrated effective use for offering diagnosis services to individuals in under-developed, under-serviced, and remote areas.[72] For patients in these regions such as Indigenous communities where there are limited ocular specialists, the use of AI can provide patients with a quick and cost-effective diagnosis. An ophthalmic technician or a general practitioner can take an image of the patient's eye and diagnose the patient using the AI screening program.[72,73] This method of healthcare delivery prevents the need of patients to travel long distances to visit an ophthalmologist, and the patient is able to receive an early diagnosis. The healthcare sector benefits from this process due to reduced wait times that may exist in clinics, reduced travel times, and increased specialist referral rates.

Additionally, in the field of ophthalmology, there exists high occupational burnout among ophthalmologists.[74] Due to the increase in our aging population, there will inevitably be an increase in people with eye diseases such as glaucoma, cataracts, and AMD.[75] The workload of ophthalmologists will need to increase in order to accommodate these patient demands. In the conventional method of ophthalmologists diagnosing every patient, this process can be very time consuming and expensive. There will be continual pressures faced by ophthalmologists to keep up with the influx of patients. Thus, novel methods such as the implementation of AI should be implemented in the workplace to assist clinicians.

Machine Learning Classifiers for Cataract Diagnosis

In the last decade, research teams around the world have researched the use of ML classifiers to automatically diagnosis and screen for cataracts. There have been studies published on this research since 2009 and new studies published in 2021.[63] Acharya et al. published their study in 2009 based in India, and the research uses a backpropagation neural network as their ML classifier.[69] As newer studies were published, there was a general trend of an increase in images used to train and validate the ML algorithms. This may be due to the increasing advancement of ML algorithms, coding, and infrastructure of the neural networks. In Wu et al., a total of 37,638 slit-lamp images were used to train and validate, and a convolutional neural network (CNN) was used in the study.[76] With their CNN, they were able to accurately diagnose cataracts with a 92.0% sensitivity, and 83.9% specificity.

There is a gap in current literature for a meta-analysis on the diagnostic accuracy of all ML classifiers for diagnosing cataracts. This type of study is needed to show the potential of ML classifiers for an accurate diagnosis so these technologies can be implemented in regular clinical settings. If the existing ML classifiers prove to be inaccurate with a low pooled sensitivity and specificity, then it informs researchers and computer scientists to develop better models and algorithms for cataract diagnosis.[62]

2.1.7 Thesis Methodologies

This thesis will contain two studies (Chapter 03 and Chapter 04) that will use different methodologies. Chapter 03 will contain a systematic review and meta-analysis that will summarize the current literature on the diagnostic accuracy of machine learning classifiers for cataracts. Chapter 04 will contain a cost-effectiveness analysis using the results found from Chapter 03.

Systematic Review

A systematic review is a type of literature review that synthesizes all available scientific evidence with a specific methodology that limits bias on a certain topic.[77] Systematic reviews are a form of evidence synthesis that are reproducible and transparent in its methods and have a focused and well-defined research question – these are some elements that distinctly separates a systematic review from a narrative review.[77] Systematic reviews are important because it is an evidence-based practice that uses the best available evidence in both published and grey literature. Systematic reviews are especially useful for policy makers, healthcare personnel, and researchers because it summarizes and collates all available literature on a research topic into one document.[77] This type of review makes it very easy for individuals to be thoroughly informed on one research topic and make evidence-based decisions. This form of review increases the precision of result estimates by minimizing bias in the review, and it also judges the quality of the evidence included in the study.

Systematic reviews need to follow the PRISMA checklist which is a standardized process.[78] The process of a systematic review begins with developing a strong and well-defined research question. This is important to the review because a clear and strong objective will guide the researcher to developing useful results and analysis.[77,79] Next, a comprehensive database search of published literature must be conducted to obtain relevant literature related to the research question.[77] Keywords and MeSH terms should be formulated by the research team in order to yield the most relevant results for the selected database search. Additional searches consist of forward and backward citation tracing, manual searching, and grey literature through conference proceedings and unpublished literature. After retrieving all relevant literature, the study will go through study screening where the reviewers will do a multi-level title, abstract, and full-text screening. If the reviewers do not agree on an article's eligibility towards the study inclusion criteria, then the reviewers may resolve conflicts with one another, or a third reviewer will step in and decide.

The studies that have been included after the full-text review will then go through a risk of bias assessment to assess the bias and quality of the individual studies included.[77,79] There are many different types of risk assessments for different study types (ie. intervention studies, observational studies, diagnostic accuracy studies, etc.) which rates the individual studies. For diagnostic accuracy studies, the QUADAS-2 Tool is the most used assessment to assess each included study.[80] After the risk of bias assessment, the finalized studies can go through data extraction where relevant study information such as study population, design, objectives, and results can be collected.[77]

Meta-analysis

Meta-analysis is a form of statistical analysis that uses different techniques to pool and summarize data from different studies on a similar topic.[81,82] When conducting a systematic review, the authors may additionally conduct a meta-analysis if the retrieved and included studies contain a consistent effect size across the studies to compute a summary effect. Effect size is a unit of currency in a meta-analysis, and it is a measure in a study that represents the impact of an intervention in a study.[82] An effect size can represent any relationship between two variables in a study, or it can be an estimate of a single value. Effect sizes can be dichotomous (ie. Risk ratios, odds ratios, log ratios), continuous (ie. mean differences, response ratios), or correlational; analysis can be made with many types of data.[82]

Meta-analyses are especially powerful when summarizing and quantifying effect sizes because each study included in the meta-analysis is given a different weight in the analysis. Studies with greater precision in its results are weighted more than studies with poor precision. Precision of a study is often driven by the sample size of the study, so most often, the greater the sample size the greater the study weight.[81,82] However, precision is just one of many factors that can influence the weight of a study.

A meta-analysis can be classified under two types of models: a fixed-effect model, or a random-effects model. In a fixed-effect model, there is an assumption that there is just one common (true) effect amongst all the included studies.[82] This implies that all factors that could influence the effect size will be the same and constant in all studies; the only difference and variation from study to study only exists from sampling error. [77, 82] However, the fixed-effect model is often rarely used in practicality because the assumption of a true effect is implausible and there is often variance and heterogeneity in the studies. There can be many differences in the studies in terms of study population, study demographics, and other factors that vary. Thus, the random-effects model is most often used. In the random-effects model, there is an assumption that the true effects are normally distributed, and the model tries to deal with both the within study variance, and the between-study variance.[81,82]

To test and quantify heterogeneity, Cochran's Q test, T^2, and I^2 statistics are computed.[82] The Q score is a standardized measure which sums the squared deviation of each individual effect size from the mean, multiplied by the weighted inverse-variance for a particular individual study – this is also known as the weighted sum of squares. The use of the Q score is to compare with the expected weighted sum of squares to test the null and get an estimate of the excess variance.[82,85] T^2 (Tau-squared) measures the variance of the true effects, and it is used to assign the study weights in the random-effects model. Tau is also able to estimate the distribution of the true effects and evaluate the standard deviations.[82,85] I^2 represents the proportion of the observed variance with the real differences in the effect size. I^2 is a descriptive statistic that essentially measures the inconsistency of study results across all the studies.[82,85] Higgins et al. (2003) gives recommendations on how to interpret the I^2. They suggest that a value of 25%, 50%, and 75% may be considered as low, moderate, or high heterogeneity, respectively.[82]

Meta-analysis of diagnostic accuracy studies

In meta-analysis of diagnostic accuracy studies, the use of a hierarchical logistic regression is a common analysis. There are several methods to statistically analyze diagnostic accuracy data which includes the use of the hierarchical summary receiver operating characteristic (HSROC) model, and the bivariate model.[83,84] In many updated statistical packages that exist in statistical software, the statistical command often fits the model of both the HSROC and bivariate parameterizations called a hierarchical logistical regression, for example the “metandi” command in STATA.[85] The advantage in using hierarchical logistic regression for the meta-analysis of diagnostic studies is its ability to perform statistical distributions at two levels. [83,84] The first level accounts for the within-study variability using binomial distributions by assessing the number of true positive and true negative cases. The second level accounts for between study variance using logistical (log-odds) transformation of the sensitivity and specificity.[83,84]

In the instance of diagnostic accuracy studies, the effect size and the measurement required for the meta-analysis are the number of true positives (tp), false positives (fp), false negatives (fn), and true negatives (tn).[86] These values are known as a confusion matrix, and it is regularly used to calculate the sensitivity and specificity for diagnostic tests. True positive refers to a positive diagnosis for a patient with the disease of interest, and a fp refers to a positive diagnosis for a patient free from the disease of interest. While tn refers to a negative diagnosis for a patient who is free of the disease of interest, and a fn refers to a negative diagnosis for a patient with the disease of interest.[86]

The sensitivity refers to the proportion of individuals with a condition that received a positive result on the test. In an example considering cataract diagnosis, sensitivity is the proportion of confirmed cataract patients who received a cataract diagnosis. Whereas specificity refers to the proportion of individuals who do not have the condition of interest that received a negative result on the diagnostic test; these are the non-cataractous patients who received a no-cataract diagnosis from the screening procedure.[86] The calculations for sensitivity and specificity are shown in (1) and (2).

$$(1)\ Sensitivity = \frac{tp}{tp + fn}$$

$$(2)\ Specificity = \frac{tn}{tn + fp}$$

Using the sensitivity and specificity values, a summary receiver operating characteristic (SROC) plot can be constructed in which the sensitivity on the y-axis is plotted against the specificity on the x-axis.[85] Unlike a conventional receiver operating characteristics plot, there are no lines that connect the plots with each other because each plot is a different study rather than a different threshold within the same study. Each study in an SROC plot is indicated by a circle and the size of the circle represents the sample size of each study.[85]

The HSROC model is based on an underlying SROC plot, and it makes the same normality assumptions as in a random-effects model.[85] The parameters included in the HSROC model include the mean and variance of the accuracy parameter, a positivity parameter with a mean and variance, and a constant shape parameter. A plot of the fitted HSROC model will continue a summary curve, a summary operating point for the pooled sensitivity and specificity value, the 95% confidence region, and the 95% prediction region.[85] The 95% confidence region is the area for the point estimate of the sensitivity and specificity, while the 95% prediction region is the confidence region for the forecasted sensitivity and specificity of future studies.[85]

The bivariate model models the same parameters as the HSROC model, but it utilizes the log-odds transforms for a bivariate normal distribution between the included studies.[83,85] The output of these models will give the summary values and confidence intervals for sensitivity and specificity, and this can be graphically modelled back in the linear scale. Additionally, the diagnostic odds ratio (DOR), and positive and negative likelihood ratios (LR +/-) can be retrieved from the computed sensitivity and specificity values.[86]

Economic Evaluations

Through an economics perspective, resources are always scarce, and choices must be made towards the optimal allocation of resources.[87] A healthcare economic evaluation is the analysis of the cost and effectiveness of at least two treatments, and it is an important decision-making consideration for healthcare officials and administrators. Economic evaluations involve placing a value on a certain course of action, and it can motivate a reallocation of resources.[87,88] The goal of an economic evaluation is often to identify which program, treatment or intervention is most efficient and it begins with a desired policy objective.

In Canada, where there is publicly financed health insurance (ie. OHIP in Ontario), economic evaluations are important for optimal resource allocation and to determine which health programs or interventions are funded and covered by the government.[89] Often, based on the results of an evaluation, they determine which drugs and interventions are covered in the public health insurance plan and included in the schedule of benefits. [90] In low-to-middle-income countries (LMIC), knowing the most cost-effective intervention can inform on programs and interventions that can be implemented at a lower cost, and potentially serve a wider range of under-serviced populations.[89]

In health economic evaluations, the consequence is a more complicated measure to obtain because different people have different views on how to assign social value towards a health gain or health loss.[87] The consequences of a health program can reflect a change in health status of patient, change in health sector resources consumed, or change in non-health effects (ie. changes in productive working time, time saved).[87,88] The three types of economic evaluations reflect these measures and preferences.[92]

Within the healthcare sector, often the final policy objective is to produce the most health-related welfare by observing any changes to the health status of the patient – this may be looking at quality of life or disability adjusted life year measures.[87] However, there are other intermediate outcomes that may reflect change in other important clinical indicators pertinent to other parties such as physicians, or family and caregivers of the patient. Therefore, the perspective (or viewpoint) that is used in an economic evaluation is important because a certain treatment or program may look unattractive when other perspectives are being considered.[92] These perspectives can include the individual patient, a specific organization, the Ministry of Health, or a societal perspective.

Another methodological consideration in an economic evaluation is the time horizon of the study. Many studies will capture the entire lifetime of a patient in order to assess both intended and unintended effects of a patient's life as a result of health intervention.[87,92] For example, if a study was assessing the mortality of a patient after a certain medical treatment, a short time horizon may lead to an over optimistic view of the analysis. A longer time horizon which follows through the patient's life course would be more beneficial and fully encompassing the patient's experience.[87] However, this is no standard time horizon in economic evaluations because each study has its own goals and scope. For economic studies that investigate benefits that occur in the shorter term such as the number of cases detected by a screening program or reduction in the number of medical visits by a patient, then a shorter time horizon may also be appropriate.[87]

Methods of Economic Evaluations

There are three main types of economic evaluations: cost-effectiveness analysis (CEA), cost-utility analysis (CUA), and cost-benefit analysis (CBA). All analyses use similar monetary unit measurement, but the consequences may be differently reported and used in each type. Cost-minimization analysis is also another form of economic evaluation, although this method is more outdated and less robust than the other method.

Cost-effectiveness analysis is a type of economic evaluation that assesses the cost per unit effect achieved – it measures the consequences in natural units (ie. life-years gained, cases averted, etc.).[87,88] The purpose of a CEA can be to compare the costs of different medical treatment, programs, or interventions aimed at the same health problem, and evaluate the expected benefits, but a CEA does not place a social value to the consequence.[88] For example, if there is an alternative dialysis program that may prolong the life of a patient with renal failure, then the CEA would be interested in looking at the extra cost per life-year gained as a result of the new treatment program.[88] For cancer screening programs, the cost of a 'detected case', 'case averted', or 'patient diagnosed' by the screening intervention may be the more relevant outcome of interest.[87] However, there are limitations to a CEA because it may be difficult to evaluate the opportunity costs and benefits forgone in other programs. Taking the cancer screening program for example again, although a CEA may be able to capture the number of cancers detected, the scope of the analysis does not account for the type of cancer or the stage of cancer which can have very different health effects on the individual patient.[88]

A decision tree can be generated to compare two separate interventions, and case-base probabilities are inputted into the analysis. Through inputting the probabilities of an output with the costs and effectiveness, a simple cost-effectiveness measure can be computed to compare the interventions.[87] An incremental cost-effectiveness ratio (ICER) describes the average incremental cost associated with one additional unit of the measure of effect, and there will be a dominated, dominant, or undominated intervention. An intervention is defined as dominated when the intervention is more costly and produces lower effects or consequence. The calculation of an ICER is expressed in (3) which describes the change in incremental resources required by the intervention, divided by the change in incremental health effects gained by the intervention.[87,88]

$$(3)\ ICER = \frac{Cost\ new - Cost\ old}{Effect\ new - Effect\ old}$$

A CEA generates one of nine possible "dominance" outcomes when a new treatment or program is being compared to another (**Table 2.2**).[87,88] In instances where there is clearly one program that is less costly and more effective, then this program is said to be absolutely/strongly dominant. There are also outcomes that may be classified as “weak dominance” such that a program may be equally as effective as the comparator but costs more or less, or a program that is more or less effective but costs the same as the comparator.[88] Finally, there are outcomes that are classified as “non-dominance” in which there is a trade-off between programs to see if the added effects generated are worth the extra costs, or if the lowered costs justify the lowered effects.[88]

The results of a CEA can also be shown on a cost-effectiveness plane which plots the ICER onto a plane with four quadrants: northeast, southeast, southwest, and northwest.[93] The vertical axis represents the change in cost, and the horizontal axis represents the difference in effect. A positive ICER slope represents a trade-off for either intervention; the northeast quadrant represents that the new program is more effective and more costly than the comparator, and the southwest quadrant represents the new program is less effective and less costly than the comparator.[87] A negative ICER slope needs to be interpreted with caution because it represents two extremes: the new program is dominant (southeast quadrant), or the new program is dominated (northwest quadrant).[87,93]

A second type of economic evaluation is a cost-utility analysis (CUA) which values health outcomes and consequences in terms of a generic measure of health gain (ie. quality-adjusted life years (QALY), disability-adjusted life years (DALY), healthy years equivalent).[87,88] These estimates of health utility can quantify the quality of life and productiveness of a patient using a rating or valuation from 0 to 1.[88] Using these ratings, each case that is considered in the CUA can be adjusted by the length of time affected by the disease of interest. The benefits to utilizing the health utility of the patient is that a patient’s health status can be considered after the implementation of the new program or treatment.[88] If a patient with severe vision loss as a result of cataracts receives a successful cataract surgery and regains 20/20 vision, the QALY of the patient would inevitably increase post-surgery. If the patient does not receive cataract surgery or has an unsuccessful surgery, then their vision will continue to deteriorate and their QALY may continue to decrease over the time horizon of the study.[94]

The third type of economic evaluation is a cost-benefit analysis (CBA) which is similar to a CUA, but the consequence is valued in money or willingness-to-pay (WTP).[87,88] The result of a CBA is often stated as a form of ratio of costs to benefits, or a sum of net benefits or loss of one program compared to the other. A CBA can indicate whether a program is worthwhile at all to be implemented. In a CBA, individuals express their hypothetical WTP which is a scenario where the individual can consider their willingness-to-pay in a dollar amount to mitigate a certain health risk.[88] Based on the pre-determined WTP, the WTP can be plotted on a cost-effectiveness plane to determine if the new program or treatment should be accepted. For example, if the WTP is $40,000/QALY, then this is the cost-effectiveness threshold that would be used to draw an acceptability curve to illustrate if one intervention is favoured over the other. It is noted that CBA is used more for feasibility studies rather than full program implementation purposes.[88]

2.1.8 Conclusion

In conclusion, cataract is a global disease that requires attention and research. Cataracts is the leading cause of blindness in adults, and the aging population of Canada and the world will inevitably cause a massive increase in cataract cases.[3] The diagnosis and treatment of cataracts warrants research, and the use of AI and ML can assist clinicians in the care of cataracts. There is potential for AI screening programs for diagnosing eye diseases such as cataracts, as shown by the ARDS program by Google Health. Currently, there are many research teams who are investigating the use of ML classifiers for cataract diagnosis, and there is great potential for its regular clinical use. However, there remains a gap in the literature for a meta-analysis on the diagnostic accuracy of machine learning classifiers for cataracts, and a cost-effectiveness analysis for AI screening programs for cataracts in rural regions.

2.2 Thesis Rationale

The utilization and implementation of artificial intelligence and machine learning in the healthcare setting has been on the rise in the past few decades. Complex algorithms and software have been developed to resolve complex problems and processes within medical data and clinical decisions.[59,63] The incorporation of these new and modern technologies has the ability to improve medical care delivery, and the patient experience in our healthcare system. The use of AI for telemedicine has shown benefits for patients in terms of access to care and healthcare equity. There continues to be large amounts of studies on AI and ML within the healthcare field published every year due to the increasing trend of digital medicine.[61]

Additionally, currently in Nepal, there is approximately an 8.5% prevalence of any type of cataract among adults.[36] Thus, many individuals are impacted by cataracts, and it is a very relevant aging health condition seen across many adults. The field of ophthalmology seems to be a very attractive field for AI development perhaps due to the readily available datasets of ocular images. Large technology companies such as Google and IBM have invested in large research teams to make developments in these areas.[59] Therefore, this thesis is important because it addresses key topics within artificial intelligence, ophthalmology, and global health that remains unanswered in the current body of literature.

Many studies have claimed that the use of machine learning can provide an accurate and cost-effective alternative to regular clinical practice of treatments or interventions in the healthcare setting. However, there have not been any studies that prove this to be true for cataract diagnosis.[91] It is hypothesized that ML screening programs will in fact be superior to human assessments in both the diagnostic accuracy and direct costs.

This thesis will investigate, at a high level, the general diagnostic accuracy of all machine learning classifiers for cataracts that are currently in literature in both published and unpublished sources. To date, there are no systematic reviews or meta-analysis on the use of machine learning for cataract screening.[91] By investigating and exploring the sensitivity and specificity of these novel algorithms, it can give more information to researchers and clinicians on whether more development of algorithms is warranted to produce better diagnostic accuracy, or if current algorithms are capable to be implemented to the regular clinical setting in hospitals and ophthalmology clinics.

Equally as important to the ML diagnostic effectiveness is the financial and health economical consideration to this screening program. Health economic evaluations are also warranted to demonstrate the financial feasibility of these new interventions and to assess the long-term benefits of any investments. With human assessment of cataract diagnosis, it can involve costly personnel (ophthalmologists, eye specialists) in every step of the patient care.

There is currently a paucity of literature on the CEA of any AI-related interventions. In fact, in a systematic review by Wolff et al., the authors found only 6 studies that met their objective of summarizing cost-effectiveness studies dedicated to AI in healthcare.[91] Out of the 6 studies, no studies were identified to have comprised of a methodologically complete cost impact analysis. To the best of our knowledge, the cost-effectiveness analysis in Chapter 04 presents as one of the first cost-effectiveness analysis of a machine learning screening program for cataracts.

In Nepal, there is currently no machine learning screening program readily available for diagnosing cataracts. Given the potential of machine learning diagnosis utilized for other diseases such as diabetic retinopathy, cataract diagnosis can benefit from the same developments and provide timely patient referrals. The thesis rationale is that a ML-based screening program for cataracts may be a feasible and viable alternative over the traditional diagnostic eye camps for cataracts by assessing the diagnostic accuracy and cost-effectiveness.

2.3 Thesis Objectives

This thesis aims to evaluate the effectiveness of ML classifiers for the diagnosis of cataracts through two objectives: 1) assessing the diagnostic accuracy, and 2) determining the cost-effectiveness.

Objective 1 – Diagnostic accuracy

To systematically review and meta-analyze the diagnostic accuracy of ML classifiers for cataracts among all adult and pediatric eyes available in datasets to assess their accuracy and reliability to be implemented in real clinical settings. Chapter 03 will aim to qualitatively and quantitatively summarize the existing body of knowledge pertaining to the accuracy of novel ML classifiers developed and compute a pooled-sensitivity and specificity estimate.

Objective 2 – Cost-effectiveness

To determine the cost-effectiveness of implementing a fully automated ML-based screening program in eye camps compared to the current standard eye camps for the diagnosis of cataracts for the adult population in rural Nepal.

2.4 References

1. Asbell PA, Dualan I, Mindel J, Brocks D, Ahmad M, Epstein S. Age-related cataract. The Lancet [Internet]. 2005;365(9459):599–609. Available from: http://www.sciencedirect.com/science/article/pii/S0140673605179112

2. Vinson JA. Oxidative stress in cataracts. Pathophysiology [Internet]. 2006;13(3):151–62. Available from: https://www.sciencedirect.com/science/article/pii/S0928468006000526

3. Liu Y-C, Wilkins M, Kim T, Malyugin B, Mehta JS. Cataracts. The Lancet [Internet]. 2017;390(10094):600–12. Available from: http://www.sciencedirect.com/science/article/pii/S0140673617305445

4. Steinmetz JD, Bourne RRA, Briant PS, Flaxman SR, Taylor HRB, Jonas JB, et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. The Lancet Global Health [Internet]. 2021 Feb 1;9(2):e144–60. Available from: https://doi.org/10.1016/S2214-109X(20)30489-7

5. Michael R, Bron AJ. The ageing lens and cataract: a model of normal and pathological ageing. Philosophical transactions of the Royal Society B: Biological sciences. 2011;366(1568):1278–92. Available from: https://doi.org/10.1098%2Frstb.2010.0300

6. Government of British Columbia. Cataract - Treatment of adults [Internet]. BC Guidelines. 2021 [cited 2021 Nov 15]. Available from: https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/bc-guidelines/cataract#diagnostic

7. Group A-REDSR. Risk factors associated with age-related nuclear and cortical cataract : a case-control study in the Age-Related Eye Disease Study, AREDS Report No. 5. Ophthalmology [Internet]. 2001 Aug;108(8):1400–8. Available from: https://pubmed.ncbi.nlm.nih.gov/11470690

8. Li H, Ko L, Lim JH, Liu J, Wong DWK, Wong TY. Image based diagnosis of cortical cataract. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2008. p. 3904–7. Available from : https://doi.org/10.1109/IEMBS.2008.4650063

9. Delcourt C, Cristol J-P, Tessier F, Léger CL, Michel F, Papoz L, et al. Risk Factors for Cortical, Nuclear, and Posterior Subcapsular Cataracts: The POLA Study. American Journal of Epidemiology [Internet]. 2000 Mar 1;151(5):497–504. Available from: https://doi.org/10.1093/oxfordjournals.aje.a010235

10. Coronado AC, Zaric GS, Martin J, Malvankar-Mehta M, Si FF, Hodge WG. Diabetic retinopathy screening with pharmacy-based teleophthalmology in a semiurban setting: a cost-effectiveness analysis. CMAJ open. 2016;4(1):E95–102. Available from: https://doi.org/10.9778/cmajo.20150085

11. Song Z, Zhao D, Lv C, Pu W, Xiao W. Ten-year etiologic review of Chinese children hospitalized for pediatric cataracts. Eye science. 2014;29(3):138–42.

12. Delpero WT, Robinson BE, Gardiner JA, Nasmith L, Rowan-Legg A, Tousignant B. Evidence-based clinical practice guidelines for the periodic eye examination in children aged 0-5 years in Canada. Canadian Journal of Ophthalmology, Journal canadien d'ophtalmologie. 2019 Dec;54(6):751–9. Available from: https://doi.org/10.1016/j.jcjo.2019.09.003

13. Congdon NG. Prevention strategies for age related cataract: present limitations and future possibilities. The British Journal of Ophthalmology [Internet]. 2001 May;85(5):516–20. Available from: https://pubmed.ncbi.nlm.nih.gov/11316704

14. Chua J, Koh JY, Tan AG, Zhao W, Lamoureux E, Mitchell P, et al. Ancestry, socioeconomic status, and age-related cataract in Asians: the Singapore epidemiology of eye diseases study. Ophthalmology. 2015;122(11):2169–78. Available from: https://doi.org/10.1016/j.ophtha.2015.06.052

15. Lou L, Ye X, Xu P, Wang J, Xu Y, Jin K, et al. Association of Sex With the Global Burden of Cataract. JAMA Ophthalmology [Internet]. 2018 Feb 1;136(2):116–21. Available from: https://doi.org/10.1001/jamaophthalmol.2017.5668

16. Hammond CJ, Snieder H, Spector TD, Gilbert CE. Genetic and Environmental Factors in Age-Related Nuclear Cataracts in Monozygotic and Dizygotic Twins. New England Journal of Medicine [Internet]. 2000 Jun 15;342(24):1786–90. Available from: https://doi.org/10.1056/NEJM200006153422404

17. Li L, Wan X, Zhao G. Meta-analysis of the risk of cataract in type 2 diabetes. BMC Ophthalmology [Internet]. 2014;14(1):94. Available from: https://doi.org/10.1186/1471-2415-14-94

18. Gong Y, Feng K, Yan N, Xu Y, Pan C-W. Different amounts of alcohol consumption and cataract: a meta-analysis. Optometry and vision science: official publication of the American Academy of Optometry. 2015 Apr;92(4):471–9. Available from: https://doi.org/10.1097/opx.0000000000000558

19. Klein BEK, Klein R, Lee KE, Meuer SM. Socioeconomic and lifestyle factors and the 10-year incidence of age-related cataracts. American Journal of Ophthalmology. 2003;136(3):506–12. Available from: https://doi.org/10.1016/s0002-9394(03)00290-3

20. Yu X, Lyu D, Dong X, He J, Yao K. Hypertension and risk of cataract: a meta-analysis. PloS one [Internet]. 2014 Dec 4;9(12):e114012–e114012. Available from: https://pubmed.ncbi.nlm.nih.gov/25474403

21. Ye JJ, Zadunaisky JA. Ca2+/Na+ exchanger and Na+,K+ 2Cl- cotransporter in lens fiber plasma membrane vesicles. Experimental Eye Research. 1992 Dec;55(6):797–804. Available from: https://doi.org/10.1016/0014-4835(92)90006-e

22. Lee SM, Lin SY, Li MJ, Liang RC. Possible mechanism of exacerbating cataract formation in cataractous human lens capsules induced by systemic hypertension or glaucoma. Ophthalmic Research. 1997;29(2):83–90. Available from: https://doi.org/10.1159/000268001

23. Elizabeth Glassen. Wearing sunglasses can reduce risk of cataracts: Canadian Ophthalmological Society [Internet]. Canadian Ophthalmological Society. 2021 [cited 2021 Nov 15]. Available from: https://www.cos-sco.ca/wearing-sunglasses-can-reduce-risk-of-cataracts-canadian-ophthalmological-society/

24. Ye J, He J, Wang C, Wu H, Shi X, Zhang H, et al. Smoking and Risk of Age-Related Cataract: A Meta-Analysis. Investigative Ophthalmology & Visual Science [Internet]. 2012 Jun 25;53(7):3885–95. Available from: https://doi.org/10.1167/iovs.12-9820

25. Pascolini D, Mariotti SP. Global estimates of visual impairment: 2010. British Journal of Ophthalmology [Internet]. 2012 May 1;96(5):614 LP – 618. Available from: http://bjo.bmj.com/content/96/5/614.abstract

26. Fricke TR, Tahhan N, Resnikoff S, Papas E, Burnett A, Ho SM, et al. Global Prevalence of Presbyopia and Vision Impairment from Uncorrected Presbyopia: Systematic Review, Meta-analysis, and Modelling. Ophthalmology.

2018 Oct;125(10):1492–9. Available from: https://doi.org/10.1016/j.ophtha.2018.04.013

27. Hashemi H, Pakzad R, Yekta A, Aghamirsalim M, Pakbin M, Ramin S, et al. Global and regional prevalence of age-related cataract: a comprehensive systematic review and meta-analysis. Eye [Internet]. 2020;34(8):1357–70. Available from: https://doi.org/10.1038/s41433-020-0806-3

28. Ono K, Hiratsuka Y, Murakami A. Global Inequality in Eye Health: Country-Level Analysis from the Global Burden of Disease Study. American Journal of Public Health [Internet]. 2010 Sep 1;100(9):1784–8. Available from: https://doi.org/10.2105/AJPH.2009.187930

29. He M, Wang W, Huang W. Variations and Trends in Health Burden of Visual Impairment Due to Cataract: A Global Analysis. Investigative Ophthalmology & Visual Science [Internet]. 2017 Aug 28;58(10):4299–306. Available from: https://doi.org/10.1167/iovs.17-21459

30. Bourne R, Steinmetz JD, Flaxman S, Briant PS, Taylor HR, Resnikoff S, et al. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study. The Lancet Global Health [Internet]. 2021;9(2):e130–43. Available from: https://www.sciencedirect.com/science/article/pii/S2214109X20304253

31. Wu X, Long E, Lin H, Liu Y. Prevalence and epidemiological characteristics of congenital cataract: a systematic review and meta-analysis. Scientific reports. 2016 Jun;6:28564. Available from: https://doi.org/10.1038/srep28564

32. American Optometric Association Consensus Panel on Care of the Adult Patient with Cataract. Care of adult patient with cataract [Internet]. 2004. Available from: https://www.aoa.org/AOA/Documents/Practice Management/Clinical Guidelines/Consensus-based guidelines/Care of the Adult Patient with Catract.pdf

33. Olson RJ, Braga-Mele R, Chen SH, Miller KM, Pineda R 2nd, Tweeten JP, et al. Cataract in the Adult Eye Preferred Practice Pattern®. Ophthalmology. 2017 Feb;124(2):P1–119. Available from: https://doi.org/10.1016/j.ophtha.2016.09.027

34. Canadian Ophthalmological Society Cataract Surgery Clinical Practice Guideline Expert Committee. Canadian Ophthalmological Society evidence-based clinical practice guidelines for cataract surgery in the adult eye. Canadian Journal of Ophthalmology [Internet]. 2008;43:S7–33. Available from: https://www.sciencedirect.com/science/article/pii/S0008418208800014

35. Kharel Sitaula R, Joshi SN, Khanal S. Surgical eye camp in rural area of Nepal and its role in Vision 2020. Journal of Chitwan Medical College [Internet]. 2017 Feb 20;6(2 SE-Original Research Articles):1–5. Available from: https://www.nepjol.info/index.php/JCMC/article/view/16677

36. Nepal Netra Jyoti Sangh. Epidemiology of Blindness in Nepal [Internet]. 2011. Available from: https://www.iapb.org/learn/resources/epidemiology-of-blindness-in-nepal/

37. Provincial Vision Task Force and Ministry of Health. Quality-based procedures clinical handbook: cataract day surgery [Internet]. 2020. Available from: https://www.health.gov.on.ca/en/pro/programs/ecfa/docs/qbp_cataract.pdf

38. Floyd RP. History of cataract surgery. In: Principles and practice of Ophthalmology. WB Saunders, Philadelphia; 2000. p. 1463–76.

39. AAO PPP Committee. Comprehensive Adult Medical Eye Evaluation PPP 2020. 2020. Available from: https://www.aao.org/preferred-practice-pattern/comprehensive-adult-medical-eye-evaluation-ppp

40. Holladay JT. Visual acuity measurements. Journal of Cataract & Refractive Surgery [Internet]. 2004;30(2). Available from:

https://journals.lww.com/jcrs/Fulltext/2004/02000/Visual_acuity_measurements.2.aspx

41. McGraw P, Winn B, Whitaker D. Reliability of the Snellen chart. BMJ [Internet]. 1995;310(6993):1481–2. Available from: https://www.bmj.com/content/310/6993/1481

42. Sue S. Test distance vision using a Snellen chart. Community Eye Health. 2007;20(63):52. Available from: https://pubmed.ncbi.nlm.nih.gov/17971914

43. Peters HB. Vision screening with a Snellen chart. Optometry and Vision Science. 1961;38(9):487–505. Available from: https://doi.org/10.1097/00006324-196109000-00001

44. Hussain B, Saleh GM, Sivaprasad S, Hammond CJ. Changing from Snellen to LogMAR: debate or delay? Clinical & Experimental Ophthalmology [Internet]. 2006 Jan 1;34(1):6–8. Available from: https://doi.org/10.1111/j.1442-9071.2006.01135.x

45. Laidlaw DAH, Abbott A, Rosser DA. Development of a clinically feasible logMAR alternative to the Snellen chart: performance of the compact reduced logMAR visual acuity chart in amblyopic children. British Journal of Ophthalmology [Internet]. 2003;87(10):1232–4. Available from: https://bjo.bmj.com/content/87/10/1232

46. Kniestedt C, Punjabi O, Lin S, Stamper RL. Tonometry Through the Ages. Survey of Ophthalmology [Internet]. 2008;53(6):568–91. Available from: https://www.sciencedirect.com/science/article/pii/S0039625708001495

47. Moses RA. The Goldmann applanation tonometer. American Journal of Ophthalmology. 1958;46(6):865–9. Available from: https://doi.org/10.1016/0002-9394(58)90998-x

48. Frenkel REP, Hong YJ, Shin DH. Comparison of the Tono-Pen to the Goldmann applanation tonometer. Archives of Ophthalmology. 1988;106(6):750–3. Available from: https://doi.org/10.1001/archopht.1988.01060130820030

49. Morales-Lopez H, Cruz-Vega I, Rangel-Magdaleno J. Cataract Detection and Classification Systems Using Computational Intelligence: A Survey. Archives of Computational Methods in Engineering [Internet]. 2021;28(3):1761–74. Available from: https://doi.org/10.1007/s11831-020-09440-2

50. Chylack LTJ, Wolfe JK, Singer DM, Leske MC, Bullimore MA, Bailey IL, et al. The Lens Opacities Classification System III. The Longitudinal Study of Cataract Study Group. Archives of ophthalmology (Chicago, Ill : 1960). 1993 Jun;111(6):831–6.

51. Taylor HR, West SK. The clinical grading of lens opacities. Australian and New Zealand journal of ophthalmology. 1989 Feb;17(1):81–6. Available from: https://doi.org/10.1111/j.1442-9071.1989.tb00492.x

52. Wong WL, Li X, Li J, Cheng C-Y, Lamoureux EL, Wang JJ, et al. Cataract conversion assessment using lens opacity classification system III and Wisconsin cataract grading system. Investigative ophthalmology & visual science. 2013 Jan;54(1):280–7. Available from: https://doi.org/10.1167/iovs.12-10657

53. Bencić G, Zorić-Geber M, Sarić D, Corak M, Mandić Z. Clinical importance of the lens opacities classification system III (LOCS III) in phacoemulsification. Collegium antropologicum [Internet]. 2005;29 Suppl 1:91–4. Available from: https://pubmed.ncbi.nlm.nih.gov/16193685/

54. Murthy GVS, John N, Shamanna BR, Pant HB. Elimination of avoidable blindness due to cataract: where do we prioritize and how should we monitor this decade? Indian Journal of Ophthalmology. 2012;60(5):438. Available from: https://doi.org/10.4103/0301-4738.100545

55. Kandel RP, Rajashekaran SR, Gautam M, Bassett KL. Evaluation of alternate outreach models for cataract services in rural Nepal. BMC Ophthalmology [Internet]. 2010;10(1):9. Available from: https://doi.org/10.1186/1471-2415-10-9

56. Venkataswamy G, Lepkowski JM, Ravilla T, Brilliant GE, Shanmugham CA, Vaidyanathan K, et al. Rapid epidemiologic assessment of cataract blindness. The Aravind Rapid Epidemiologic Assessment Staff. International Journal of Epidemiology. 1989;18(4 Suppl 2):S60-7. Available from: https://doi.org/10.1093/ije/18.supplement_2.s60

57. Aravind Eye Care System. About Aravind [Internet]. Aravind Eye Care System. 2022 [cited 2022 Mar 10]. Available from: https://aravind.org/our-story/

58. Collon S, Chang D, Tabin G, Hong K, Myung D, Thapa S. Utility and Feasibility of Teleophthalmology Using a Smartphone-Based Ophthalmic Camera in Screening Camps in Nepal. Asia-Pacific journal of ophthalmology (Philadelphia, Pa) [Internet]. 2020;9(1):54–8. Available from: https://pubmed.ncbi.nlm.nih.gov/31990747

59. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nature Biomedical Engineering [Internet]. 2018;2(10):719–31. Available from: https://doi.org/10.1038/s41551-018-0305-z

60. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke and Vascular Neurology [Internet]. 2017 Dec 1;2(4):230 LP – 243. Available from: http://svn.bmj.com/content/2/4/230.abstract

61. Datta S, Barua R, Das J. Application of artificial intelligence in modern healthcare system. Alginates—recent uses of this natural polymer. 2020.

62. Lee D, Yoon SN. Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. International Journal of Environmental Research and Public Health. 2021;18(1):271. Available from: https://doi.org/10.3390/ijerph18010271

63. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthcare Journal. 2019;6(2):94. Available from: https://doi.org/10.7861/futurehosp.6-2-94

64. Zuiderveen Borgesius F. Discrimination, artificial intelligence, and algorithmic decision-making. 2018. Council of Europe, Directorate General of Democracy. Available from: https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decisionmaking/1680925d73

65. Hague DC. Benefits, Pitfalls, and Potential Bias in Health Care AI. North Carolina Medical Journal [Internet]. 2019 Jul 1;80(4):219 LP – 223. Available from: http://www.ncmedicaljournal.com/content/80/4/219.abstract

66. Google Health. Seeing Potential [Internet]. 2019 [cited 2021 Nov 15]. Available from: https://about.google/intl/ALL_us/stories/seeingpotential/?utm_source=about.google&utm_medium=social&utm_campaign=copy-link

67. Health G. Using AI to prevent blindness. Available from: https://health.google/caregivers/arda/#:~:text=Using AI to prevent blindness,clinicians in identifying other diseases.

68. Xu Y, Gao X, Lin S, Wong DWK, Liu J, Xu D, et al. Automatic grading of nuclear cataracts from slit-lamp lens images using group sparsity regression. Medical image computing and computer-assisted intervention: MICCAI . International Conference on Medical Image Computing and Computer-Assisted Intervention [Internet]. 2013;16(Pt 2 PG-468-75):468–75.

69. Acharya RU, Yu W, Zhu K, Nayak J, Lim T-C, Chan JY. Identification of cataract and post-cataract surgery optical images using artificial intelligence techniques. Journal of medical systems [Internet]. 2010;34(4):619–28. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med8&NEWS=N&AN=20703916

70. Cheung R, Chun J, Sheidow T, Motolko M, Malvankar-Mehta MS. Diagnostic accuracy of current machine learning classifiers for age-related macular degeneration: a systematic review and meta-analysis. Eye [Internet]. 2021; Available from: https://doi.org/10.1038/s41433-021-01540-y

71. Murtagh P, Greene G, O'Brien C. Current applications of machine learning in the screening and diagnosis of glaucoma: a systematic review and Meta-analysis. International journal of ophthalmology [Internet]. 2020 Jan 18;13(1):149–62. Available from: https://pubmed.ncbi.nlm.nih.gov/31956584

72. Lamminen H, Voipio V, Ruohonen K, Uusitalo H. Telemedicine in ophthalmology. Acta Ophthalmologica Scandinavica [Internet]. 2003 Apr 1;81(2):105–9. Available from: https://doi.org/10.1034/j.1600-0420.2003.00045.x

73. Saleem SM, Pasquale LR, Sidoti PA, Tsai JC. Virtual Ophthalmology: Telemedicine in a COVID-19 Era. American Journal of Ophthalmology [Internet]. 2020;216:237–42. Available from: http://www.sciencedirect.com/science/article/pii/S0002939420302142

74. Cheung R, Yu B, Iordanous Y, Malvankar-Mehta MS. The Prevalence of Occupational Burnout Among Ophthalmologists: A Systematic Review and Meta-Analysis. Psychological Reports [Internet]. 2020 Aug 31;0033294120954135. Available from: https://doi.org/10.1177/0033294120954135

75. Khairallah M, Kahloun R, Bourne R, Limburg H, Flaxman SR, Jonas JB, et al. Number of People Blind or Visually Impaired by Cataract Worldwide and in World Regions, 1990 to 2010. Investigative Ophthalmology & Visual Science [Internet]. 2015 Oct 20;56(11):6762–9. Available from: https://doi.org/10.1167/iovs.15-17201

76. Wu X, Huang Y, Liu Z, Lai W, Long E, Zhang K, et al. Universal artificial intelligence platform for collaborative management of cataracts. The British journal of ophthalmology [Internet]. 2019;103(11 PG-1553–1560):1553–60.

77. Khan K, Kunz R, Kleijnen J, Antes G. Systematic reviews to support evidence-based medicine. Crc press; 2011.

78. Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLOS Medicine [Internet]. 2009 Jul 21;6(7):e1000097. Available from: https://doi.org/10.1371/journal.pmed.1000097

79. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions. John Wiley & Sons; 2019.

80. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011 Oct;155(8):529–36.

81. Egger M, Smith GD, Altman D. Systematic reviews in health care: meta-analysis in context. John Wiley & Sons; 2008.

82. Borenstein M, Hedges L v, Higgins JPT, Rothstein HR. Introduction to meta-analysis. John Wiley & Sons; 2021.

83. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Statistics in medicine. 2001 Oct;20(19):2865–84.

84. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JAC. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics [Internet]. 2007 Apr 1;8(2):239–51. Available from: https://doi.org/10.1093/biostatistics/kxl004

85. Sterne JAC. Meta-analysis in Stata: an updated collection from the Stata Journal. StataCorp LP; 2009.

86. Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. Acta paediatrica. 2007;96(3):338–41.

87. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford university press; 2015.

88. Hurley J. Health Economics. 1st Edition. McGraw-Hill Ryerson; 2010.

89. Schoen C, Osborn R, Squires D, Doty MM, Pierson R, Applebaum S. How health insurance design affects access to care and costs, by income, in eleven countries. Health affairs. 2010;29(12):2323–34.

90. Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG. Cost-effectiveness in health and medicine. Oxford University Press; 2016.

91. Wolff J, Pauling J, Keck A, Baumbach J. The Economic Impact of Artificial Intelligence in Health Care: Systematic Review. Journal of medical Internet research [Internet]. 2020 Feb 20;22(2):e16866–e16866. Available from: https://pubmed.ncbi.nlm.nih.gov/32130134

92. Cohen DJ, Reynolds MR. Interpreting the results of cost-effectiveness studies. Journal of the American College of Cardiology [Internet]. 2008;52(25):2119-26. Available from: https://www.jacc.org/doi/abs/10.1016/j.jacc.2008.09.018

93. Klok RM & Postma MJ. Four quadrants of the cost effectiveness plane: some considerations on the south-west quadrant, Expert Review of Pharmacoeconomics & Outcomes Research [Internet]. 2004; 4:6, 599-601. Available from: 10.1586/14737167.4.6.59

94. Brown GC. Vision and quality-of-life. Transactions of the American Ophthalmological Society. 1999;97:473-511.

2.5 Table

Table 2.1 Ophthalmic evaluation for the diagnosis and treatment of cataract. (Table adapted from the Canadian Ophthalmological Society)[34]

Evaluation	**Details**
Patient History	- Patient’s assessment of functional status - Pertinent medical conditions - Current medications - Allergies to medications and latex - Risk factors that could affect the surgical plan - Previous ophthalmic surgery, including refractive surgery
Measurements	- Visual acuity with current correction at distance and at near - Best-corrected visual acuity, including under glare conditions - Intraocular pressure
Examinations	- External (lids, lashes, lacrimal apparatus, orbit) - Ocular alignment and motility - Slit-lamp biomicroscope of the anterior segment - Dilated examination of the lens, macula, peripheral retina, optic nerve, and vitreous; B-scan ultrasound of fundus if inadequate view clinically - Assessment of relevant aspects of the patient’s mental and physical status

Table 2.2 Possible results from a cost-effectiveness analysis – dominance chart

		Incremental effectiveness of new program compared to control		
		More	**Same**	**Less**
Incremental cost of new program compared to control	**More**	Non-dominance (trade-off)	Weak dominance (reject new program)	Strong Dominance (reject new program)
	Same	Weak dominance (accept new program)	Non-dominance (neutral)	Weak dominance (reject new program)
	Less	Strong Dominance (accept new program)	Weak dominance (accept new program)	Non-dominance (trade-off)

CHAPTER 3

3 Diagnostic Accuracy of Machine Learning Classifiers for Cataracts: A Systematic Review and Meta-analysis

3.1 Introduction

3.1.1 Background

A cataract is the opacification of the lens in the human eye which results in cloudiness and poor visual acuity.[1,2] The development of cataracts is often related to age, trauma, and even congenital factors.[2-5] According to a WHO report, more than 50% of the world's first cause of blindness is due to cataracts and it continues to be the leading cause of blindness, especially in low to middle-income countries.[6] While cataract is commonly found in adult and elderly patients, pediatric cataract has an estimated prevalence of 4.24 per 10,000 live births, and it is the major causes of childhood blindness.[4] On a global level, cataract has contributed to 17.7 million disability-adjusted life years, a measure that represents the total number lost to disability or premature death, and it is continuing to be increasing.[7]

In current clinical practice, ophthalmologists commonly use several diagnostic tests for cataracts. Slit-lamp imaging is the most common imaging technology that utilizes an intense line of light to illuminate the eye and to look for abnormalities. Clinicians often use the Lens Opacities Classification System III for grading images of cataract.[1,2]

The early diagnosis of patients with cataracts can often lead to improved visual outcomes because patients can quickly receive treatment and cataract surgery.[8,9,10] However, this process is often stalled or delayed when people in remote and under-serviced areas such as First Nations communities, rural regions, or low-to-middle income countries do not have access to ophthalmologists, treatment, or healthcare resources.[11,12] Often, ophthalmologists or other trained eye specialists in urban settings will travel to these under-served areas to perform diagnosis, check-ups, and treatments.[12] Considering the aging population, the increased number of cataract cases can potentially contribute to the demand of ophthalmologists to rise. Ophthalmologists are facing high prevalence of occupational burnout and they have high demands of patient care and overtime work.[13,14] There is an evident need for the use of AI as it holds great potential for its application in clinical settings.[11,12]

AI has been an emerging technology in the medical field, and it can be an influential modern technological innovation. The role of AI is to mimic and simulate a human's mental process through computers to perform complex and sophisticated tasks. Machine learning is an application of AI, and its purpose is to automatically perform tasks through training and learning processes.[15] Researchers have used ML to train computer algorithms to automatically detect eye diseases such as AMD, glaucoma, and diabetic retinopathy through processing large sets of fundus, optical coherence tomography, and slit-lamp images.[16-21] Machine learning classifiers such as support vector machines, convolutional neural networks and random forests have been used to obtain a cost-effective, simple, and fast diagnosis of eye diseases.

3.1.2 Objective

The applications of AI in the field of ophthalmology are growing rapidly and it is proven to be a powerful tool for the diagnosis of eye diseases. There have been numerous systematic reviews published on the diagnostic accuracy of AI for glaucoma and diabetic retinopathy.[22,23] It is to the best of our knowledge that this is the first systematic review of its kind. The objective of this study is to systematically review and meta-analyze the diagnostic accuracy of machine learning classifiers for cataracts among all pediatric and adult eyes available in databases to assess their accuracy and reliability to be implemented in real clinical settings.

3.2 Methods

This systematic review has been registered in PROSPERO (CRD42020219316) and it follows PRISMA guidelines (**Appendix A**).[24]

3.2.1 Search Strategy

An initial scoping search was performed using PubMed, Google Scholar, and Web of Science. A systematic and comprehensive database search included MEDLINE/PubMed, EMBASE, CINAHL and ProQuest Dissertations and Theses to find articles on current artificial intelligence technologies used in the field of ophthalmology for the diagnosis of cataracts. The search was carried out using keywords and controlled terms for the following concepts: “Artificial intelligence” AND “Diagnosis” AND “Cataracts”. The search strategy and keywords for each database is detailed in **Appendix B** and the searches were conducted until September 12, 2021. The search was limited to English and human studies only. No limits were placed on publication date and study location to maximize our eligible studies. OVID AutoAlerts for MEDLINE and EMBASE databases were used to send weekly updates for any new published literature that the search strategy encompassed.

In addition, grey literature searches were conducted in order to obtain a comprehensive search. Conferences held through the American Academy of Ophthalmology, the Association for Research in Vision and Ophthalmology, and the Canadian Society of Ophthalmology were searched in all available years. We searched through the conferences until September 12, 2021. Keywords that were used for the grey literature search consisted of “artificial intelligence” and “diagnosis”. The search strategy and search results for each conference is displayed in **Appendix C**. Forward and backward citation tracing were carried out on studies that were included after the full-text screening. Refer to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram for more details.[24]

3.2.2 Inclusion and Exclusion Criteria

This systematic review included all studies that utilized artificial intelligence to diagnose cataracts on human eyes — there was no restriction in the age of the eyes. Any ML classifiers such as neural networks, Random Forests, adaptive boosting, or support vector machines that were able to differentiate between healthy and cataract eyes were included. The ML classifier must include a learning, training or validation processes when evaluating the images. If the study mentioned a computer assisted or automated process without the mention of AI or a learning process, then the study was excluded. We included studies that used AI algorithms to make a first diagnosis of cataracts from healthy eyes, or the AI was able to differentiate between cataract eyes and healthy eyes. Studies that investigated the use of AI for grading cataracts were excluded.

Additionally, all imaging techniques were included in this review; this included fundus imaging, slit lamp imaging, or visible wavelength images. Studies that reported diagnostic performance indicators such as sensitivity and specificity were also included. The studies must also include a reference standard, confirmed, and validated by trained clinicians or ophthalmologists. Included publications must be primary studies, and there were no restrictions on study design; ophthalmology news articles, opinion pieces, and case reports were excluded. Only studies in English were included, and there was no restriction placed on study location or publication date.

3.2.3 Screening

Database search results were all imported into Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia). All duplicated articles were removed in Covidence and two levels of systematic screening were conducted by two independent reviewers (RC & SS). When consensus could not be reached between the two reviewers, all disagreements were resolved by discussion of the two reviewers. The first level of screening consisted of a broad title and abstract screening. If the study title and/or abstract mentioned the use of AI and the diagnosis of cataracts, the study was included and moved on to the second screening; the rest of the articles were excluded. The second level of screening was a full-text screening, and we examined each article to choose relevant studies that matched our research question. The article must consist of a first diagnosis of cataracts using an AI algorithm. However, the technology must be an advanced ML classifier that includes a training and processing element for the diagnosis of cataracts. The included studies moved on to the risk of bias assessment. If data from the same study was discovered to be reported in multiple papers, only the report with higher quality was retained.

Cohen's kappa (κ) statistics were calculated at both screening levels before conflicts were resolved as shown in **Appendix D**. We reported kappa values based on the recommendations of Landis and Koch: greater than 0.75 represents excellent agreement beyond chance, below 0.40 represents poor agreement, and 0.40 to 0.75 represents intermediate to good agreement.

3.2.4 Risk of Bias Assessment and Data Extraction

To assess the risk of bias of the included articles, the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool was used to assess the level of bias and concerns of applicability.[25] Two independent reviewers (RC & SS) conducted the risk of bias and any disagreements were resolved by discussion. QUADAS-2 considered 4 domains: patient selection, index tests, reference standard, and flow and timing. Each of the domain was given a risk of bias score to assess whether there was high, low, or unclear bias. The QUADAS-2 tool was not intended to provide a summary quality score. All studies were included for this systematic review.

A data extraction form was created to collect relevant data and details of each of the included studies. A pilot extraction form was first created for a subset of included studies in order to determine what information was most relevant and pertinent to this study. Study data were extracted by one reviewer (RC). The following data were extracted from each study: author, year of publication, study location, study design, ML classifier, type of imaging, number of study participants, number of images used, training and testing process, database and datasets used, diagnosis reference standard, area under curve, sensitivity, specificity, and accuracy. Additionally, other outcomes that were extracted if reported in the studies were the tn, tp, fn, fp, and total number of images classified as healthy or cataractous. Data extraction table for the meta-analysis is outlined in **Appendix E.**

3.2.5 Statistical Analysis

Data was synthesized and analyzed using STATA 15.0 (STATA Corp, College Station, Texas, U.S.A.) for the diagnostic accuracy of machine learning for cataracts. The extracted data of interest were the sensitivity, specificity, and area under curve values of the ML classifiers used. A hierarchical bivariate random effects model was conducted. Hierarchical logistic regression was used to determine the pooled estimates of sensitivity and specificity of diagnostic accuracy. The summaries of the fitted Hierarchical Summary Receiver Operator Characteristic (HSROC) model, the summary receiver operating characteristics (SROC) curve, the 95% confidence interval and the 95% prediction region were plotted graphically. Sub-group analysis was conducted based on adult and pediatric cataracts.

The positive/negative likelihood ratios (LR+/LR-) were calculated using bivariate models to generate estimates of the likelihood of a positive/negative test. From this result the diagnostic odds ratio (DOR) was calculated to determine the relative diagnostic effectiveness. DOR is the ratio of the odds of a positive screen test in a cataract case relative to the odds of a negative screen test in a non-cataract case.

Forest plots showing the within-study estimates and confidence intervals for sensitivity and specificity were plotted separately. For each study, the sensitivity and specificity were aggregated using the fixed or random-effects model based on the absence or presence of heterogeneity to estimate the summary effect. To test for heterogeneity, I^2 statistics, Q-value, and χ^2 statistics were computed. An I^2 value of less than 50% implies low heterogeneity, and in these cases, a fixed-effect model was computed. An I^2 statistics of 50% or more represents high heterogeneity, and in these cases a random-effects model was calculated. Additionally, a high Z-value, a low p-value (< 0.01) and a large χ^2 value implies significant heterogeneity and therefore, a random-effects model was computed.

3.3 Results

3.3.1 Search Results and Study Characteristics

The inclusion and exclusion process are shown by **Figure 3.1** using the PRISMA flow diagram. In total, the search strategy yielded 150 articles, and the grey literature search yielded 35 articles. However, 50 of those articles were identified as duplicates, resulting in 135 articles entering the first level title and abstract screening. In the first level, 107 articles were excluded, and 28 studies moved onto the final level (full text) screening. After the full text screening, seven studies were excluded due to wrong study outcomes and comparators, and 21 studies met the inclusion criteria. The kappa statistics score was 0.72 and 0.84 at each stage of screening — this was considered to be moderate to high agreement from both reviewers. The studies that were included after the last screening went through a risk of bias assessment using the QUADAS-2 Tool. All 21 studies were included for the qualitative synthesis[26–46] and 11 of those studies were included for the meta-analysis.[26,27,34–36,38–42,45]

The study characteristics of the 21 included studies are displayed in **Table 3.1**. All included studies were conducted in Asian countries including China [27–29,33–36,38,40–45], Singapore[30–32], India[26,37,39], and Japan[46]. Nine of the studies used slit-lamp images[26,30–32,34–36,40,45], 11 studies used fundus images [27–29,33,38,39,41–44,46], and one used visible wavelength eye images[39] for the testing and training process. Among all the studies, there was a varied use of different ML classifiers. The most common classifier used was support vector machines (SVM) which was used by six studies[28,31,32,38,39,41], and convolutional neural networks which was also used by eight studies[33,36,38,40,42,44–46]. Other ML classifiers included backpropagation neural network[26,27], discriminant analysis[29], AdaBoost (adaptive boosting)[35,43], CC-Cruiser[34], and a novel ranking classifier.[30] Each study had a unique training and testing process to teach their classifier to differentiate between the healthy and non-healthy images.

In the included studies, there were multiple studies that overlapped in the use of certain datasets including the Singapore Malay Eye Study, Childhood Cataract Program of Chinese Ministry of Health, and the dataset from Beijing Tongren Eye Center of Beijing Tongren Hospital. However, all studies differed in their choice of ML classifier, and number of images used for training and testing. In studies Li et al. (2009) and Li et al. (2010), both uses the same ML classifier, imaging technique, and dataset to detect cataract. Lin et al. (2019), Lin et al. (2020), and Liu et al. (2017) investigated specifically on pediatric cataracts and used the same available database from the Childhood Cataract Program in China. Among pediatric cataracts, three studies contributed to four sensitivity and specific pairs in total. Whereas among the adult cataracts analysis, there were nine pairs of sensitivity and specificity pairs in total used for the quantitative analysis. All images that were used had been confirmed by a human grader such as an ophthalmologist, clinician, or clinical grader to confirm the diagnosis of the eye.

3.3.2 Risk of Bias Assessment

Most of the studies that passed the full-text review had a low risk of bias in the four domains; it was low risk in patient selection (76.2%), index tests (95.2%), reference standard (95.2%), and flow and timing (76.2%). There was low concern of applicability in the patient selection (85.7%), index test (95.2%), and reference standard (95.2%). The study by Shimizu et al. was rated high risk across all domains because only its abstract was available. The risk of bias assessment and concerns about applicability for each study are summarized in **Appendix F**.

3.3.3 Diagnostic Accuracy of Machine Learning Classifiers for Cataracts in Adults

Eight studies were used for the meta-analysis to conduct the analysis of diagnostic accuracy for cataracts in adult patients. The SROC curve is represented in **Appendix G** which plots the sensitivity against the specificity of each study. The SROC curve shows that most of the included studies are scattered across the top right corner of the plot, and it demonstrates that there is a high specificity and sensitivity of various ML classifiers.

Figure 3.2 shows the HSROC plot which illustrates the study estimates indicated by the circles, the HSROC curve or summary curve, a summary operating point or the summary value for sensitivity and specificity, the 95% confidence region (inner ellipse), and the 95% prediction region (outer ellipse) for the summary operating point. The HSROC curve appears in the left upper quadrant and has a large area under the curve. This is an indication that ML classifiers are a relatively accurate method for diagnosis because the area under the HSROC curve is large. Four studies fall outside the 95% confidence interval of the summary estimate. The 95% prediction region is the estimate of future observations. The prediction region shows a wide prediction region for the true predictions of both specificity and sensitivity; there is a greater expected variability for the sensitivity.

The summary estimate for sensitivity was 0.948 [95% CI: 0.815-0.987] and specificity was 0.960 [95% CI: 0.924-0.980] for cataracts screening using an ML classifier (**Figure 3.3**). The summary estimates indicate that ML classifier correctly detects 94.8% of cataract cases and correctly classifies 96.0% of those without cataract as cataract-negative. The distribution of the studies in the plot demonstrates the variability of both specificity and sensitivity amongst studies.

The positive likelihood ratio was 23.837 [95% CI: 12.241-46.419], while the negative likelihood ratio was 0.054 [95% CI: 0.014-0.208] (**Appendix H**). This shows that the likelihood of a positive diagnosis in a cataract case is greater than the likelihood of negative diagnosis in a non-cataract case. The positive likelihood ratio is greater than one and it represents that the positive diagnosis is associated with cataract. Because the negative likelihood ratio is less than one, the ML classifier which gave a negative diagnosis is associated with the absence of cataract. The effectiveness of the diagnostic accuracy of the ML classifiers for cataract given by the diagnostic odds ratio is 442.248 [95% CI: 89.201-2192.611] (**Appendix H**). This demonstrates that the relative odds of a positive screen test in cataract cases are 442.248 times more likely than a negative screen test in a non-cataract case. Thus, the ML classifiers discriminate between the true negative and true positive cataract images correctly and accurately.

3.3.4 Diagnostic Accuracy of Machine Learning Classifiers for Pediatric Cataracts

A sub-group analysis was conducted for pediatric cataracts and a total of three studies were used for the quantitative analysis for assessing the diagnostic accuracy of ML classifiers for pediatric cataracts. **Figure 3.4** shows the HSROC plot for pediatric cataracts. All four classifiers fall within the 95% confidence interval (inner ellipse) of the summary estimate. The 95% prediction region shows wide variability for the true predictions of both specificity and sensitivity.

The summary estimate for sensitivity was 0.882 [95% CI: 0.696-0.961] and specificity was 0.891 [95% CI: 0.807-0.942] for cataracts screening using an ML classifier (**Figure 3.5**). The distribution of the studies in the plot demonstrates the variability of both specificity and sensitivity amongst studies. The positive likelihood ratio was 8.119 [95% CI: 4.068-16.206], while the negative likelihood ratio was 0.133 [95% CI: 0.045-0.392] for cataracts in children (**Appendix H**). The effectiveness of the diagnostic accuracy of the ML classifiers for cataract given by the DOR is 61.200 [95% CI: 11.656-321.328] (**Appendix H**). The relative odds of a positive screen test in pediatric cataract cases are 61.2 times more likely than a negative screen test in a non-cataract case. Thus, the ML classifiers discriminate between the true negative and true positive images correctly and accurately in child eyes.

3.4 Discussion and Conclusion

This systematic review and meta-analysis included 21 full text articles for the qualitative synthesis and 11 full text articles for the quantitative synthesis. In the systematic review, 100,134 images were used for training and validation of the ML classifiers for diagnosing cataracts in human eyes. For the adult cataract meta-analysis, 74,188 images were used included for the analysis, and 5246 images were used for the pediatric cataract subgroup analysis. To the best of our knowledge, this is the first review of its kind to assess the diagnostic accuracy of ML classifiers for cataracts. ML classifiers are advantageous at detecting true positive cases of cataracts and they have very high DOR estimates.

Given the COVID-19 pandemic, the role of telemedicine — more specifically in teleophthalmology — has demonstrated a growing importance in healthcare, and the use of AI algorithms can further assist clinicians in making clinical decisions. The use of ML has demonstrated good use for offering cataract diagnosis services to people in under-developed and remote regions.[11,12] This alternative method of receiving healthcare benefits both the patients and the healthcare system because there is reduced wait and travel times, increased specialist referral rates, and reduced patient costs. In urban settings, the use of ML for diagnosis can reduce patient load, wait times, and improve efficiency of ophthalmology clinics.[12]

The results of the pooled sensitivity and specificity estimates for diagnosing cataracts have shown that ML-classifiers perform with high accuracy for both true positive (tp) and true negative (tn) cases. It is also equally important to consider the number of cases that are classified as false negative (fn) and false positive (fp) in order to assess how many patients may be underdiagnosed and in need of cataract treatment. A missed cataract diagnosis may significantly affect the patient's quality of life and quality of vision, and it may result in a delayed cataract treatment and follow-up.[8,9] When assessing the accuracy of the ML classifiers, both the sensitivity and specificity of the diagnostic technology must be fully considered.[47] All included qualitative publications in this study reported an accuracy proportion in their article, however, single accuracy proportions do not indicate whether there is a trade-off between the sensitivity or specificity of the test. For future diagnostic accuracy studies, it is encouraged for all authors to report the sensitivity, specificity, tn, tp, fn, and fp values for researchers and clinicians to make better informed decisions.

Additionally, meta-analysis of observational studies is influenced by inherent biases.[48] Factors such as the hospital and study location, race and age of study participants, and type of imaging technique can influence the study results. The clinical diagnosis and confirmation of cataracts may also be subject to each ophthalmologist or retinal specialist and study location. All included studies had a reference standard which may be an ophthalmologist, an eye specialist, or a professional/clinical grader. However, not all studies explicitly stated the clinical guidelines or cataract classification systems that the reference standard used to provide a cataract diagnosis. An ophthalmologist's number of years of experience in the field, and an ophthalmologist's field of expertise in ophthalmology are additional factors that may influence the study results.

In training a machine learning algorithm in ophthalmology, there can be multiple imaging modalities that researchers may use including slit-lamp imaging, fundus imaging or OCT imaging. For cataracts, it is common to use slit-lamp or fundus imaging as shown in

Table 3.1. Due to the differing image modalities used by different studies, there may be potential bias in the results because each individual algorithm learns to read and process the image types differently — this may increase the between study heterogeneity. Despite the variation in ML classifiers within the slit lamp imaging and fundus cohorts, all the included studies for the quantitative analyses displayed consistent results. Based on this study's inclusion and exclusion criteria, only two studies that used slit-lamp images would have been included for a potential subgroup analysis. Therefore, a subgroup analysis was not conducted due to the small sample size and insufficient power in the analysis.

All included studies originated from three study countries: China, Singapore, and India. Due to the limited eye database and datasets available in these countries, there were database overlaps throughout the included studies. Datasets from Beijing Tongren Hospital and the Childhood Cataract Program of Chinese Ministry of Health (CCPCMOH) from China were most used. CCPCMOH was the only database used for the pediatric cataract subgroup, thus more images of pediatric eyes with cataract are needed for continued research in the future. The pediatric cataract results from this study should be interpreted with caution due to the limited number of studies available. This suggests that more expansive research is warranted in other regions and countries to retrieve more unique eyes for this analysis.

In conclusion, the diagnostic accuracy of ML classifiers for adult and pediatric cataracts is very high and the diagnostic test performance shows very promising results. The prospects of using ML for the diagnosis of cataracts in real clinical settings is a possibility, although the extent of our findings and the timeline of this implementation still needs to be established. This study demonstrates only one facet of the application of artificial intelligence in the healthcare field and ophthalmology. There are endless opportunities for the implementation of AI in medical care as novel research and new algorithms are developed.

3.5 References

1. Liu Y-C, Wilkins M, Kim T, Malyugin B, Mehta JS. Cataracts. Lancet [Internet]. 2017;390(10094):600-612. Available from: https://doi.org/10.1016/s0140-6736(17)30544-5

2. Asbell PA, Dualan I, Mindel J, Brocks D, Ahmad M, Epstein S. Age-related cataract. Lancet [Internet]. 2005;365(9459):599-609. Available from: https://doi.org/10.1016/s0140-6736(05)17911-2

3. Sharma AK, Aslami AN, Srivastava JP, Iqbal J. Visual outcome of traumatic cataract at a tertiary eye care centre in North India: a prospective study. Journal of Clinical and Diagnostic Research [Internet]. 2016;10(1):NC05-NC8. Available from: https://doi.org/10.7860/jcdr/2016/17216.7049

4. Wu X, Long E, Lin H, Liu Y. Prevalence and epidemiological characteristics of congenital cataract: a systematic review and meta-analysis. Scientific Reports [Internet]. 2016;6:28564. Available from: https://doi.org/10.1038/srep28564

5. Khairallah M, Kahloun R, Bourne R, Limburg H, Flaxman SR, Jonas JB, et al. Number of people blind or visually impaired by cataract worldwide and in world regions, 1990 to 2010. Investigate Ophthalmology & Visual Science [Internet] 2015;56(11):6762-6769. Available from: https://doi.org/10.1167/iovs.15-17201

6. Pascolini D, Mariotti SP. Global estimates of visual impairment: 2010. British Journal of Ophthalmology [Internet]. 2012;96(5):614-618. Available from: https://doi.org/10.1136/bjophthalmol-2011-300539

7. Ono K, Hiratsuka Y, Murakami A. Global inequality in eye health: country-level analysis from the global burden of disease study. American Journal of Public Health

[Internet]. 2010;100(9):1784-1788. Available from: https://doi.org/10.2105/ajph.2009.187930

8. Chylack Jr LT, Wolfe JK, Singer DM, Leske MC, Bullimore MA, Bailey IL, et al. The Lens Opacities Classification System III. The Longitudinal Study of Cataract Study Group. Archives of Ophthalmology [Internet]. 1993;111(6):831-836. Available from: https://doi.org/10.1001/archopht.1993.01090060119035

9. Gali HE, Sella R, Afshari NA. Cataract grading systems: a review of past and present. Current Opinion in Ophthalmology [Internet]. 2019;30(1). Available from: https://doi.org/10.1097/icu.0000000000000542

10. Allen D, Vasavada A. Cataract and surgery for cataract. British Medical Journal [Internet]. 2006;333(7559):128-132. Available from: https://doi.org/10.1136/bmj.333.7559.128

11. Lamminen H, Voipio V, Ruohonen K, Uusitalo H. Telemedicine in ophthalmology. Acta ophthalmologica Scandinavica [Internet]. 2003;81(2):105-109. Available from: https://doi.org/10.1034/j.1600-0420.2003.00045.x

12. Saleem SM, Pasquale LR, Sidoti PA, Tsai JC. Virtual ophthalmology: telemedicine in a COVID-19 era. American Journal of Ophthalmology [Internet]. 2020;216:237-242. Available from: https://doi.org/10.1016/j.ajo.2020.04.029

13. Cheung R, Yu B, Iordanous Y, Malvankar-Mehta MS. The prevalence of occupational burnout among ophthalmologists: a systematic review and meta-analysis. Pyschological Reports [Internet]. 2021;124(5):2139-2154.

14. Dewa CS, Loong D, Bonato S, Thanh NX, Jacobs P. How does burnout affect physician productivity? A systematic literature review. BMC Health Services Research [Internet]. 2014;14:325. Available from: https://doi.org/10.1186/1472-6963-14-325

15. Consejo A, Melcer T, Rozema JJ. Introduction to machine learning for ophthalmologists. Seminars in Ophthalmology [Internet]. 2019;34(1):19-41. Available from: https://doi.org/10.1080/08820538.2018.1551496

16. Wang Y, Zhang Y, Yao Z, Zhao R, Zhou F. Machine learning based detection of age-related macular degeneration (AMD) and diabetic macular edema (DME) from optical coherence tomography (OCT) images. Biomedical Optics Express [Internet]. 2016;7(12):4928-4940. Available from: https://doi.org/10.1364/boe.7.004928

17. Tan JH, Bhandary SV, Sivaprasad S, Hagiwara Y, Bagchi A, Raghavendra U, et al. Age-related macular degeneration detection using deep convolutional neural network. Future Generation Computer Systems [Internet]. 2018;87:127-135. Available from: https://doi.org/10.1016/j.future.2018.05.001

18. Devalla SK, Liang Z, Pham TH, et al. Glaucoma management in the era of artificial intelligence. *Br J Ophthalmol* 2020;104(3):301-311.

19. Hamzah H, Lim G, Duc Nguyen Q, Mani B, Hsu W, Lee ML, et al. Artificial intelligence using deep learning system for glaucoma suspect detection. Investigative Ophthalmology & Visual Science [Internet]. 2018;59(9):4074. Available from: https://iovs.arvojournals.org/article.aspx?articleid=2691777

20. Usher D, Dumskyj M, Himaga M, Williamson TH, Nussey S, Boyce J. Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening. Diabetic medicine: a journal of the British Diabetic Association [Internet]. 2004;21(1):84-90. Available from: https://doi.org/10.1046/j.1464-5491.2003.01085.x

21. Lynch SK, Folk JC, Abramoff MD. Autonomous artificial intelligence (AI) reliably detects diabetic retinopathy. Investigative Ophthalmology & Visual Science [Internet]. 2019;60(9):1537. Available from: https://iovs.arvojournals.org/article.aspx?articleid=2741928

22. Ullah W, Pathan SK, Panchal A, Anandan S, Saleem K, Sattar Y, et al. Cost-effectiveness and diagnostic accuracy of telemedicine in macular disease and diabetic retinopathy: A systematic review and meta-analysis. Medicine (Baltimore) [Internet]. 2020;99(25). Available from: https://doi.org/10.1097/md.0000000000020306

23. Murtagh P, Greene G, O'Brien C. Current applications of machine learning in the screening and diagnosis of glaucoma: a systematic review and Meta-analysis. International Journal of Ophthalmol [Internet]. 2020 Jan 18;13(1):149–62. Available from: https://pubmed.ncbi.nlm.nih.gov/31956584

24. Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLOS Medicine [Internet]. 2009 Jul 21;6(7):e1000097. Available from: https://doi.org/10.1371/journal.pmed.1000097

25. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct;155(8):529–36. Available from: https://doi.org/10.7326/0003-4819-155-8-201110180-00009

26. Acharya RU, Yu W, Zhu K, Nayak J, Lim TC, Chan JY. Identification of cataract and post-cataract surgery optical images using artificial intelligence techniques. J Med Syst [Internet]. 2010;34(4):619–28. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med8&NEWS=N&AN=20703916

27. Cao L, Li H, Zhang Y, Xu L. Hierarchical method for cataract grading based on retinal images using improved Haar wavelet. Information Fusion [Internet]. 2020;53:196–208. Available from: http://www.sciencedirect.com/science/article/pii/S1566253518302914

28. Dong Y, Zhang Q, Qiao Z, Yang J. Classification of cataract fundus image based on deep learning. In: 2017 IEEE International Conference on Imaging Systems and

Techniques (IST) [Internet]. 2017. p. 1–5. Available from: https://ieeexplore.ieee.org/document/8261463

29. Guo L, Yang JJ, Peng L, Li J, Liang Q. A computer-aided healthcare system for cataract classification and grading based on fundus image analysis. Computers in Industry [Internet]. 2015;69:72–80. Available from: http://www.sciencedirect.com/science/article/pii/S0166361514001754

30. Huang W, Li H, Chan KL, Lim JH, Liu J, Wong TY. A computer-aided diagnosis system of nuclear cataract via ranking. Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention [Internet]. 2009;12(Pt 2 PG-803-10):803–10. Available from: https://doi.org/10.1007/978-3-642-04271-3_97

31. Li H, Lim JH, Liu J, Wong DWK, Tan NM, Lu S, et al. An automatic diagnosis system of nuclear cataract using slit-lamp images. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference. 2009;(101243413 PG-3693–6):3693–6. Available from: https://doi.org/10.1109/iembs.2009.5334735

32. Li H, Lim JH, Liu J, Mitchell P, Tan AG, Wang JJ, et al. A computer-aided diagnosis system of nuclear cataract. IEEE Transactions on Biomedical Engineering [Internet]. 2010;57(7 PG-1690–8):1690–8. Available from: https://doi.org/10.1109/tbme.2010.2041454

33. Li J, Xu X, Guan Y, Imran A, Liu B, Zhang L, et al. Automatic Cataract Diagnosis by Image-Based Interpretability. In: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC) [Internet]. 2018. p. 3964–9. Available from: https://ieeexplore.ieee.org/document/8616669

34. Lin H, Li R, Liu Z, Chen J, Yang Y, Chen H, et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial.

EClinicalMedicine [Internet]. 2019;9:52–9. Available from: http://www.thelancet.com/journals/eclinm/issue/current

35. Lin D, Chen J, Lin Z, Li X, Zhang K, Wu X, et al. A practical model for the identification of congenital cataracts using machine learning. EBioMedicine [Internet]. 2020;51. Available from: http://www.journals.elsevier.com/ebiomedicine/

36. Liu X, Jiang J, Zhang K, Long E, CUI J, Zhu M, et al. Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network. PLoS ONE [Internet]. 2017;12(3 PG-e0168606):e0168606. Available from: https://doi.org/10.1371/journal.pone.0168606

37. Pratap T, Kokil P. Computer-aided diagnosis of cataract using deep transfer learning. Biomedical Signal Process Control [Internet]. 2019;53. Available from: https://doi.org/10.1016/j.bspc.2019.04.010

38. Ran J, Niu K, He Z, Zhang H, Song H. Cataract detection and grading based on combination of deep convolutional neural network and random forests. Proceedings of IC-NIDC 2018 [Internet]. 2018:155-159. Available from: https://ieeexplore.ieee.org/document/8525852

39. S V MK, R G. Computer-aided diagnosis of anterior segment eye abnormalities using visible wavelength image analysis based machine learning. Journal of Medical Systems [Internet]. 2018;42(7):128. Available from: https://doi.org/10.1007/s10916-018-0980-z

40. Wu X, Huang Y, Liu Z, Lai W, Long E, Zhang K, et al. Universal artificial intelligence platform for collaborative management of cataracts. British Journal of Ophthalmology [Internet]. 2019;103(11 PG-1553-1560):1553-1560. Available from: https://doi.org/10.1136/bjophthalmol-2019-314729

41. Yang JJ, Li J, Shen R, Zeng Y, He J, Bi Jing, et al. Exploiting ensemble learning for automatic cataract detection and grading. Computer Methods and Programs in

Biomedicine [Internet]. 2016;124(doh, 8506513 PG-45-57):45-57. Available from: https://doi.org/10.1016/j.cmpb.2015.10.007

42. Zhang L, Li J, Zhang I, Han He, Liu B, Yang J, et al. Automatic cataract detection and grading using deep convolutional neural network. IEEE 14th International Conference on Networking, Sensing and Control (ICNSC). 2017:60-65. Available from: https://ieeexplore.ieee.org/document/8000068

43. Zheng J, Guo L, Peng L, Li J, Yang J, Liang Q. Fundus image based cataract classification. IEEE International Conference on Imaging Systems and Techniques (IST) Proceedings. 2014:90-94. Available from: https://ieeexplore.ieee.org/document/6958452?arnumber=6958452

44. Zhou Y, Li Q, Li H. Automatic cataract classification using deep neural network with discrete state transition. IEEE Transactions on Medical Imaging. 2020;39(2 PG-436-446):436-446. Available from: https://ieeexplore.ieee.org/document/8759939

45. Xu X, Zhang L, Li J, et al. A hybrid global-local representation CNN model for automatic cataract grading. IEEE Journal of Biomedical Health Informatics. 2020; 24:556–567.

46. Shimizu E, Yazu H, Aketa N, Tanji M, Sakasegawa A, Nakayama S, et al. Innovative artificial intelligence-based cataract diagnostic method uses a slit-lamp video recording device and multiple machine-learning. Investigative Ophthalmology & Visual Science. 2021 Jun 21;62(8):1031. Available from: https://iovs.arvojournals.org/article.aspx?articleid=2773807

47. Altman DG, Bland JM. Diagnostic tests. 1: sensitivity and specificity. BMJ Journal [Internet]. 1994;308(6943):1552. Available from: https://doi.org/10.1136/bmj.308.6943.1552

48. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ [Internet]. 1997;315(7109):629-634. Available from: https://doi.org/10.1136/bmj.315.7109.629

3.6 Tables and Figures

Table 3.1 Study characteristics of the included studies

Study	Study Location	Patient Type	ML Classifier	Imaging Technique and Model	# Study Participants	# Images	Training and Testing Process	Database and Datasets	Reference Standard
Acharya et al. (2009)	India	Adult	Backpropagation Neural Network	Slit Lamp	140	2520	1620 images used for training, 900 images used for testing	Department of Ophthalmology, Kasturba Medical College Hospital, Manipal, India	Ophthalmologist
Cao et al. (2020)	China	Adult	Backpropagation Neural Network	Fundus	NR	1355	452 images used for training, 903 images used for testing	Beijing Tongren Eye Center of Beijing Tongren Hospital	Two ophthalmologists
Dong et al. (2017)	China	Adult	SVM	Fundus	NR	7851	5495 images used for training, 2356 images used for testing; repeated training and testing 50 times	Department of Ophthalmology, Tsinghua University	Professional doctors
Guo et al. (2015)	China	Adult	Discriminant analysis	Fundus	NR	445	312 images used for training, 133 images used for testing; repeating the procedure 100 times	Community clinics, remote rural hospitals and other hospitals, sharing the healthcare resources through the internet	Ophthalmic experts or ophthalmologists
Huang et al. (2009)	Singapore	Adult	Novel Ranking Classifier	Topcon DC-1 Digital Slit Lamp	1000	1000	5-fold cross validation	Singapore Malay Eye Study (SiMES)	Ophthalmologists using Wisconsin cataract grading system
Li et al. (2009)	Singapore	Adult	SVM	Topcon DC-1 Digital Slit Lamp	3280	5820	100 images used for training, 5490 images used for testing	Singapore Malay Eye Study (SiMES)	Human graders using the Wisconsin cataract grading system

Li et al. (2010)	Singapore	Adult	SVM	Topcon DC-1 Digital Slit Lamp	3280	5850	100 images used for training, 5550 images used for testing	Singapore Malay Eye Study (SiMES)	Human graders using the Wisconsin cataract grading system
Li et al. (2018)	China	Adult	CNN - ResNet50	Fundus	248	8030	7030 images used for training, 1000 images used for testing	Beijing Tongren Eye Center of Beijing Tongren Hospital	Professional graders
Lin et al. (2019)	China	Pediatric	CC-Cruiser	Slit Lamp	350	350	CC-Cruiser is an ophthalmic AI platform developed by Zhongshan Ophthalmic Centre (ZOC)	Childhood Cataract Program of Chinese Ministry of Health (CCPMOH)	Senior Consultants
Lin et al. (2020)	China	Pediatric	Random Forest, AdaBoost	Slit Lamp BX900	2005	2005	4-fold cross validation	Childhood Cataract Program of Chinese Ministry of Health (CCPMOH)	Two ophthalmologists
Liu et al. (2017)	China	Pediatric	CNN	Slit Lamp BX900	NR	886	4-fold cross validation; each test was performed with 50 iterations	Childhood Cataract Program of Chinese Ministry of Health (CCPMOH), Zhongshan Ophthalmic Centre Sun Yatsen University	Two ophthalmologists
Pratap & Kokil (2019)	India	Adult	SVM	Fundus	NR	800	400 images used for training, 400 images used for testing	High resolution fundus (HRF) image database, structured analysis of the retina (STARE), standard diabetic retinopathy database (DIARETDB0), e-ophtha: a color fundus image database, methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology	Ophthalmologic Experts

								(MESSIDOR) database, digital retinal images for vessel extraction (DRIVE) database, fundus image registration (FIRE) dataset, digital retinal images for optic nerve segmentation database (DRIONS-DB), Indian diabetic retinopathy image dataset (IDRiD), available datasets from Dr. Hossein Rabbani, and other internet resources	
Ran et al. (2018)	China	Adult	CNN - Random Forest	Fundus	NR	5408	5-fold cross validation	NR	Two ophthalmologists and three experienced graders
S V & R (2018)	India	Adult	SVM	Visible Wavelength Eye Image	64	228	129 images used for training, 99 images used for testing	Indira Gandhi Medical College and Research Institute, Puducherry	Ophthalmologist
Shimizu et al. (2021)	Japan	Adult	CNN	Slit Lamp	NR	18,596	NR	NR	Ophthalmologists
Wu et al. (2019)	China	Adult	CNN - ResNet	Slit Lamp BX900, BQ900, OVSII, PSL-Classic	16,611	37,638	30132 images for training, 7506 images for testing	Chinese cataract screening programme by the Chinese Medical Alliance for Artificial Intelligence (CMAAI)	Three ophthalmologists
Xu et al. (2021)	China	Adult	CNN	Fundus	NR	8030	5621 images or training, 2409 for testing	Beijing Tongren Eye Center of Beijing Tongren Hospital	Ophthalmologist
Yang et al. (2016)	China	Adult	SVM, Backpropagation Neural Network	Fundus	NR	1239	Images divided into 3 subsets. In each fold, one subset chosen as the testing set, the	Picture Archiving and Communication System (PACS)	Ophthalmologists

							other 2 used for training.		
Zhang et al. (2017)	China	Adult	CNN	Fundus	NR	5620	Cross validation	Beijing Tongren Eye Center of Beijing Tongren Hospital	Professional graders
Zheng et al. (2014)	China	Adult	AdaBoost	Fundus	NR	460	10-fold cross validation; images divided into 10 subsets. In each fold, one subset is testing set and another nine subsets as training set	NR	Professional ophthalmologists
Zhou et al. (2020)	China	Adult	DST-ResNet	Fundus	1000	1355	Images divided into 4 subsets. In each fold, one subset is testing set and another 3 subsets used as training set	Beijing Tongren Eye Center of Beijing Tongren Hospital	Clinical graders

AdaBoost, Adaptive Boosting; CNN, convolutional neural network; DLS, deep learning system; SVM, support vector machine; LCP, Linear Configuration Patterns; NR, not reported

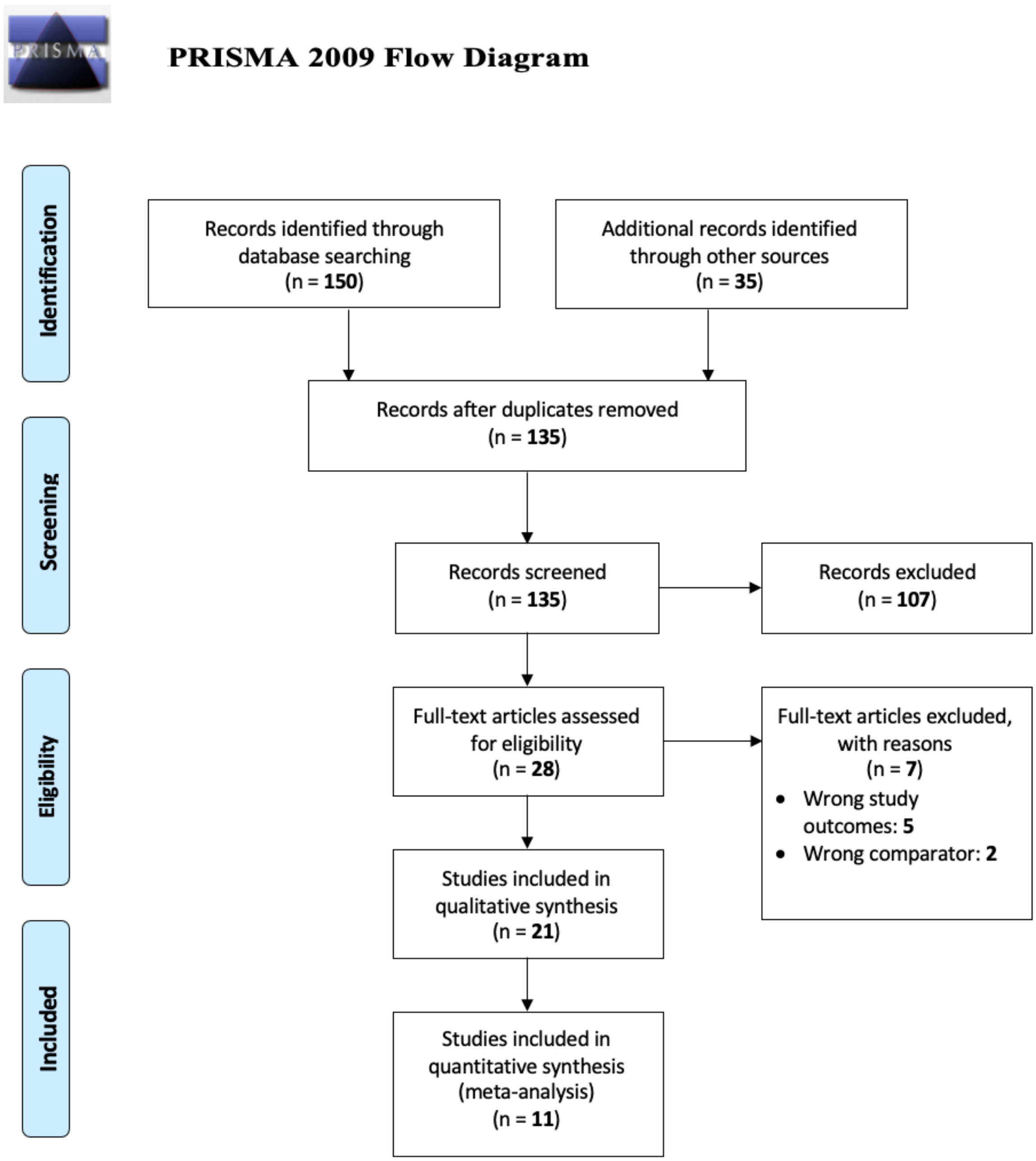


Figure 3.1 PRISMA flow diagram showing the study selection process and reasons for exclusion

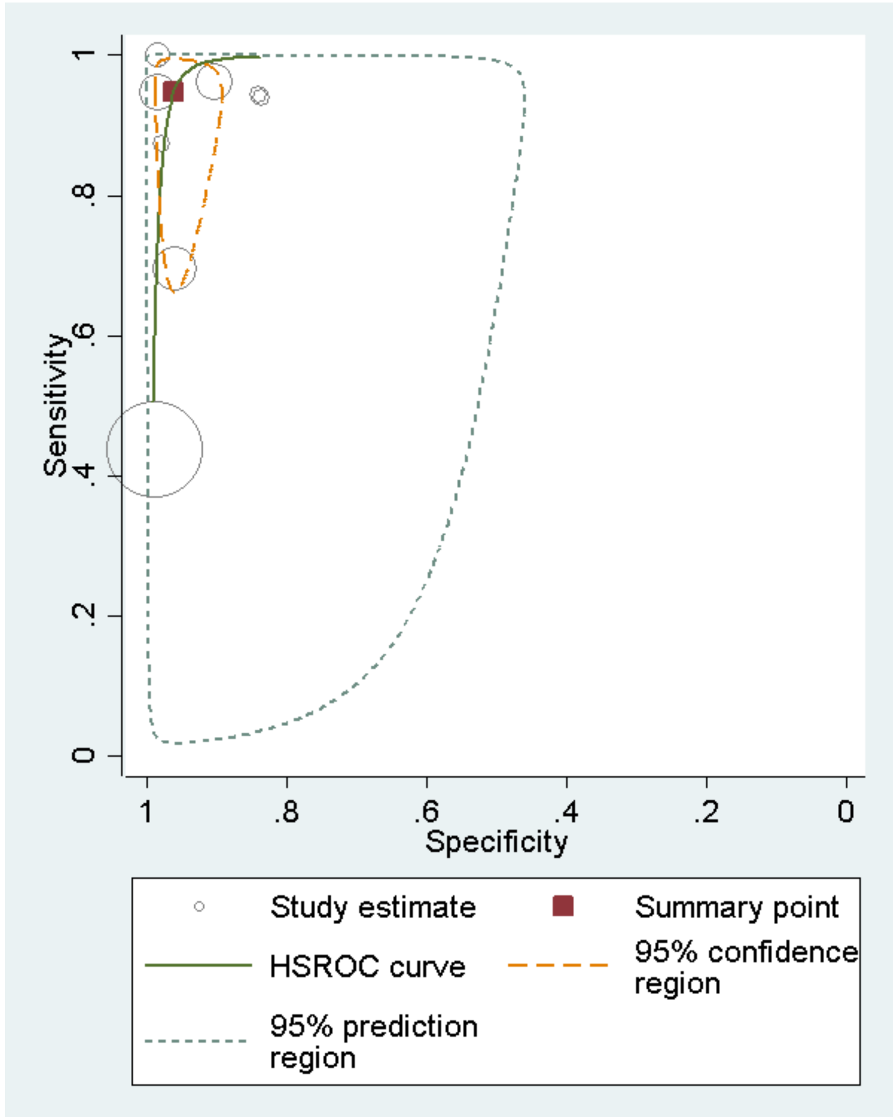


Figure 3.2 Hierarchical summary receiver operating characteristic plot for cataracts in adults

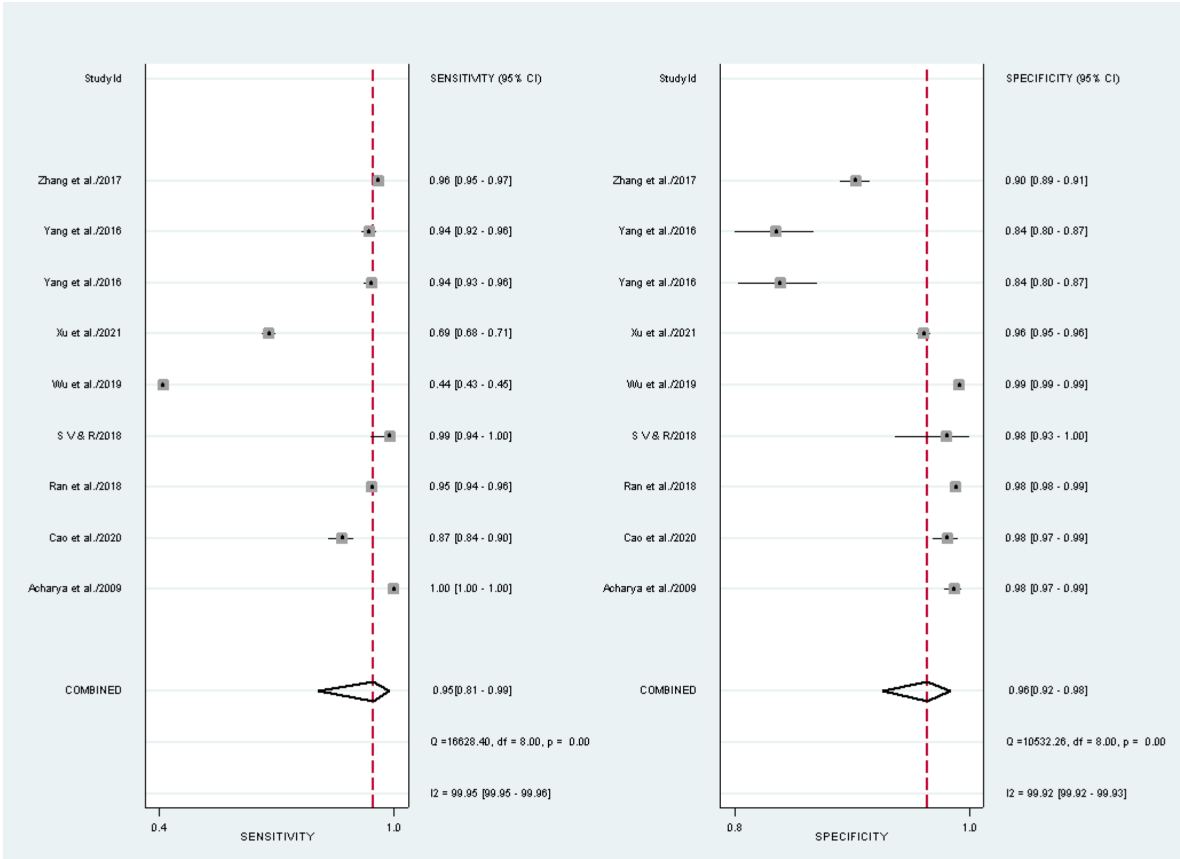


Figure 3.3 Forest plot of the pooled sensitivity and specificity estimates for the ML classifiers for cataracts in adults

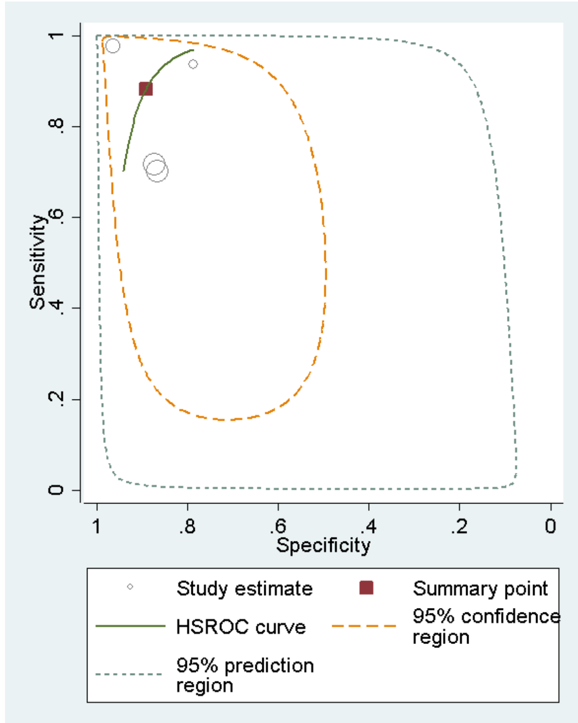


Figure 3.4 Hierarchical summary receiver operating characteristic plot for pediatric cataracts

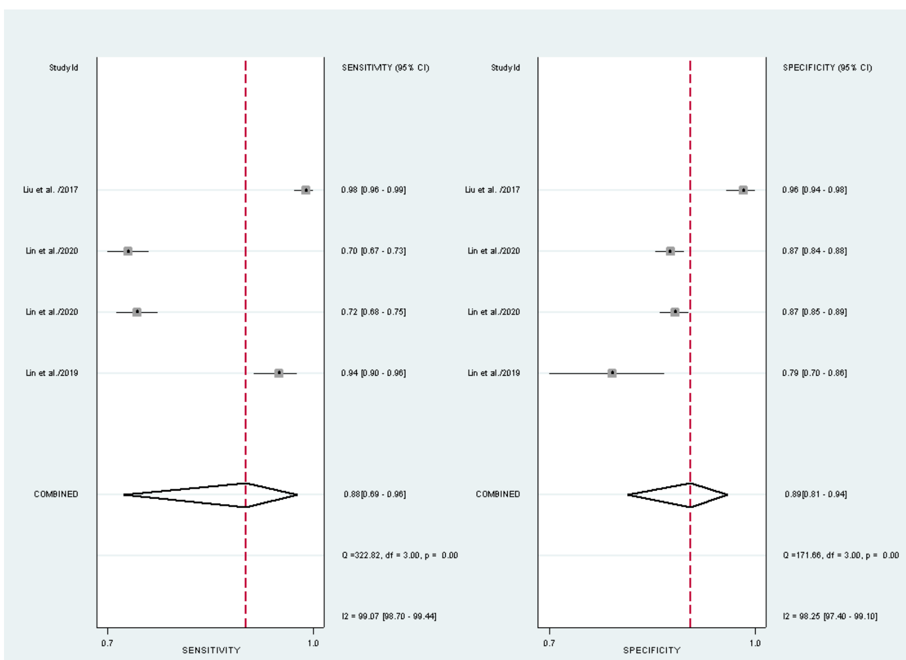


Figure 3.5 Forest plot of the pooled sensitivity and specificity estimates for ML classifiers for pediatric cataracts

CHAPTER 4

4 The Implementation of a Machine Learning-Based Cataract Screening Program in Rural Nepal: A Cost-Effectiveness Analysis

4.1 Introduction

Cataract is the opacification of the human lens in the eye, and it causes cloudiness and blurriness in the patient's vision.[1–3] Cataract is the leading cause of blindness and vision loss in many low-to-middle income countries (LMIC) due to many barriers to accessing eye care services.[4–6] In Southeast Asia alone, the global health burden of cataract vision loss was approximately 125 disability-adjusted life years (DALYs) per 100,000 people — the highest crude DALY rate out of all WHO regions.[7,8] In Nepal, the infrastructure to support eye care services and examinations have been growing in the past few decades, and there have been many improvements to provide Nepalis with an accurate cataract diagnosis.[9–11]

In 2015, Nepal promulgated a new constitution which replaced their original unitary government with a federal system of government consisted of three levels: federal, provincial, and local governments.[12,13] In revamping the constitution, the Ministry of Health and Population in Nepal was also restructured to follow the new federal structure in hopes of improving the federal health infrastructure. However, the crux of federalization and the challenge in Nepal's healthcare system is the means of financing health care.[14] Based on statistics by the WHO in 2019, the annual health spending per capita in Nepal was $53 USD.[15] Nearly 60% of the total health expenditure came from out-of-pocket payments (OOP) and it was the principal means of financing health care in Nepal.[15] This is commonly seen in other low-to-middle income countries (LMIC) where there is a higher proportion of OOP and lower proportion of federal spending. Nepal's healthcare has also heavily relied on foreign aid such as non-governmental organizations and international aid,[14]

Additionally, under this new constitution, Nepal has stated and addressed that health is a fundamental human right, and all citizens have the right to access basic healthcare services that are free of cost.[12] Evidently in reality, there are many geographic and socio-economic barriers towards access to care in Nepal which makes it very difficult to provide equitable care. In the total population, approximately only 60% of the population have access to a health facility within 30 minutes. However, disparity is observed when we compare urban populations (85.9%) with rural populations (59%) for access to a reasonably close health facility.[14] It often may take hours for patients in villages or mountainous regions to travel to their closest health clinic and some patients may even need to travel more than 100 kilometers away.[10,16] Most private hospitals and pharmacies are concentrated in the central region of Nepal (most developed region), while the western region (less developed region) has no private hospitals.[14] If Nepalis need access to a specific healthcare service, they will likely pay out-of-pocket at a private institution which creates a financial barrier to many rural populations.[14]

Further, eye specialists are not common in general practice, and they often work in specialized eye hospitals or clinics, making access to eye care even more difficult. The ophthalmologist and optometrist to population ratio is very low in Nepal (ophthalmologist 1:193,900; optometrist 1:791,700) and eye care is very underserviced.[10] The WHO goal and norm for eye care is 1:100,000 for both ophthalmologists and optometrists.[10]

Therefore, outreach services to rural Nepal, often funded by non-governmental organizations, have arisen and many Nepalis may have access to temporary village-level primary eye care centres known as “diagnostic-screening and treatment camps” or “eye camps”.[10,17,18] The implementation of eye camps in villages aim to reduce both the geographic and economic burden for these rural populations.

These eye camps follow the Aravind Model which make use of trained healthcare professionals or tele-ophthalmology with eye hospitals and ophthalmologists in Nepal.[17,19,20] A team of ophthalmic assistants (OA), ophthalmic technicians (OT), or nurses make a day trip to a village in Nepal and provide basic eye care and examinations to patients. Often, the OA will identify and diagnose patients with cataracts and refer them to see an ophthalmologist at the base hospital.[10,17] However, the agreement of a cataract diagnosis between an OA with an ophthalmologist is moderate (kappa = 0.623)[11], and the OA’s diagnostic accuracy for identifying cataracts is also moderate (specificity = 0.838)[19]. As a result, there may be many patients with cataracts who do not receive a timely diagnosis or referral to an ophthalmologist in Nepal. Additionally, current eye camps are often described as “hectic” because the eye camps service hundreds of patients a day and many workers experience occupational burnout in these conditions.[11]

Currently, there are various machine learning (ML) classifiers that can automatically detect and diagnose cataracts.[21] ML classifiers have been proven to show high diagnostic accuracy for correctly identifying cataract and non-cataract cases, and these algorithms have proven to be a powerful diagnostic tool that clinicians have begun to use.[22–24] In urban contexts, there are many benefits to using a ML screening program as it can reduce the workload of healthcare providers, reduce the burden and wait times of the eye clinics, and also provide patients with a fast and accurate cataract diagnosis.[25–29] These advantages may translate into rural eye camps and rural settings.

4.1.1 Objective

The objective of this study was to conduct a cost-effectiveness analysis (CEA) of the theoretical implementation of a ML-based cataract screening eye camp in rural Nepal in order to assess if this new technology was superior to the traditional eye camps. Our primary interest was to evaluate the additional cost per correctly detected case from a healthcare perspective to assess if the program could improve the detection of cataract and non-cataract patients.

4.2 Methods

This study followed the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist **(Appendix I)**.[30]

4.2.1 Study Setting and Population

This cost-effectiveness analysis was developed to fit the Chitwan and Nawalparasi Districts of South-central Nepal which are located within the Lumbini and Narayani Zones, respectively. The total population of the Chitwan and Nawalparasi Districts was approximately 1,223,492 according to the 2011 National Census, and this population represented 20.9% of the total population in the Lumbini and Narayani Zones.[10,16] The Bharatpur Eye Hospital was used as the base hospital for this study, and it is in the border of the two districts. The base hospital functions as a central location where the eye camps stem from, and most of the employees at the eye camps are also employed from the base hospital. Often, the ophthalmologist at the base hospital will be involved with referral triage, treatment, and follow-up of the patients from the eye camps.

The adult cataract population was the study's main interest (aged 20 or older), and only participants who were screened by an eye camp in these regions were included (n=22,805).[10]

4.2.2 Model Design

A decision tree was constructed using TreeAgePro Suite 2022 R1.0 (TreeAge Software, Inc, Williamstown, Massachusetts) to compare the traditional diagnostic and screening eye camps and machine learning-based eye camps (**Figure 4.1**). Variable inputs and definitions are listed in **Appendix J**. In the decision tree framework, the implementation of a machine learning assessment was assumed to be a fully automated system that replaces the diagnostic assessment of an OA. There was a strength to assuming a fully automated model in order to assess the full potential of an ML-based screening program.

The model's health outcome of interest was the detection of any type of cataract in at least one eye (<6/18 to 6/60 in worse eye and 6/60 in the better eye) in adults referrable to an ophthalmologist in Nepal.[18] These classifications are distance vision impairment scales, and it refers to the severity of visual acuity in each eye. For example, 6/18 refers to the patient's ability to see clearly at 6m what should normally be seen at 18m distance. This diagnostic outcome was chosen because it is consistent with both the diagnostic criteria set by

Venkataswamy et al., Cheung et al., and the Nepal Blindness Survey Guideline.[9,19,21] Additionally, this CEA was conducted from a third-party healthcare perspective, and it was tailored to have a health policy and program implementation focus. Current eye camps that exist in Nepal and surrounding regions are often funded by non-governmental organizations which run the camps for non-profit. Therefore, to mimic reality, a narrower viewpoint was utilized to include costs incurred by third-party organizations that would implement machine learning into existing eye camps. The effectiveness measure used for the CEA was the cost of an accurately 'detected case' of cataract and non-cataract.[31]

4.2.3 Interventions

Traditional eye camp (Arm 1)

We designed this economic model for patients in the rural areas and villages of the Chitwan and Nawalparasi Districts in Nepal where eye hospitals and or clinics were not easily accessible to patients. In a typical eye camp, pamphlets, radio announcements, and verbal support from village leaders are implemented to promote and publicize the operation of the eye camps on the designated day of visit. The typical eye camp uses the Aravind model where a team of OA, OT, and nurses are hired for the duration of the camp. In the camps, the OA conducts a simplified eye examination and a referral to an ophthalmologist at the base hospital was given if any cataract was detected.

Cataract patients who were referred to an ophthalmologist receive transportation to Bharatpur Eye Hospital if they consented to further assessment and treatment.[1013,15] Patients who received a true-negative or false-positive diagnosis by the OA are considered to be free of cataracts. There are substantive health related implications for patients who are given a false negative result because they are then living with an undiagnosed cataract which can result in a lower quality of life due to lowered visual acuity.

ML-based eye camp (Arm 2)

The ML-based eye camp model resembled the traditional eye camp model in terms of logistics and operations, location, and publicity. The ML-based eye camp required a slit-lamp microscope, and ML software to process the slit-lamp images. In this model, we followed the feasibility of eye camps from Kandel et al. and assumed 75 eye camps operated in one year, which employed one OA and one nurse to take slit-lamp photos of the patient's eyes.[10] The ML algorithm instantaneously provided the OA and patient a result of whether they were positive or negative for cataracts. Patients who received a positive diagnosis would receive a referral to an ophthalmologist at the base hospital. It was assumed that all images taken by the OA were readable by the algorithm.[32]

4.2.4 Model Probabilities and Cost Data

The base-case model probabilities are shown in **Table 4.1**. The prevalence of any cataract (8.5%) in the Nepali population was retrieved from the 2011 Nepal Blindness Survey.[9] Based on a study by Soellener & Koenigstorfer, the authors found that patient compliance with a machine learning diagnostic program is higher than human assessment.[33] Therefore, for the purposes of this model, we assumed that there was full patient compliance to ML assessment – this followed our assumption for a fully automated model. We also assumed full patient compliance to the traditional eye camp arm. The population at the root node of the decision tree were patients who signed up and consented to be screened at an eye camp.

Estimates for the diagnostic accuracy of ML classifiers were obtained by our meta-analysis (Chapter 03) which reported the pooled sensitivity and specificity values from 9 studies.[21] The study investigated the diagnostic performance of ML classifiers for cataracts, and the authors computed the estimated pooled sensitivity and specificity values. Another study by Venkataswamy et al. was used to retrieve the diagnostic performance of OA for diagnosing cataracts under the Aravind model.[19]

Data sources for estimates of cost included published literature and official government reports. Direct costs were incorporated into our analysis and the costs were adjusted to 2021 USD. The total costs of 75 eye camps are estimated in **Table 4.2** and the ranges used for the sensitivity analysis. Due to the implementation of ML, we assumed that there would be a reduction in the need of workers at the ML-based eye camps compared to the traditional eye camp labour and therefore a reduction in the labour costs ($683.42 vs $1847.37).[10] These costs were retrieved and estimated from the costs reported by Kandel et al. when there was a reduction from four staff working at the eye camp to two staff based on varying eye camp models.[10] In past Aravind camp models, only two workers were employed at the eye camps, therefore it is feasible to assume that one OA and one nurse were hired per ML-based eye camp.[10] Additionally to the salary for OA and nurses, a wage supplement was included to the total labour costs in the analysis. The purpose of the wage supplement is to incentivize the OA and nurses who regularly work at the base hospital to take part in working at eye camps that may be in a more rural and distant location than the hospital.

Based on literature, we did not include start-up costs with the development of the algorithm in the model – these assumptions were fair and had been utilized in other studies. [32] Further, it was assumed that within the equipment and other consumable costs, a slit lamp microscope, and computer with a compatible software for ML was included. Logistical costs in both types of camps were fixed which includes both publicity and transportation costs. The eye camps in the villages in Nepal are often temporary, one-day camps which requires publicity and promotion to the villages in advance. Health promotion is crucial to rural populations for them to understand the importance of eye care management and to utilize these eye services.[10,19] Transportation costs include the vehicle and cost of gas to transport the equipment and staff between villages and to the base hospital.[10] The cost of running an eye camp was the total cost of labour, capital, and logistical expenses.

4.2.5 Effectiveness Measures

There were two effectiveness measures that were of interest to this study: (1) the probability of a true positive (tp) cataract case correctly detected and (2) the probability of the of a tp cataract cases or true negative (tn) non-cataract cases correctly detected by the OA and ML classifier. These proportions were calculated by taking the probability of the screened patient population being truly positive or negative for cataracts and receiving a true or false diagnosis. An incremental cost-effectiveness ratio (ICER) was calculated and generated to assess the cost associated with an additional correctly detected (1) case of cataract or (2) case of cataract and non-cataract after the implementation of the ML assessment.

4.2.6 Deterministic Sensitivity Analysis

Variable parameters included in the model that were considered as drivers were included in the sensitivity analyses. Each variable had an effect measure, and ranges were applied to the variables either based on their 95% confidence intervals, or an upper and lower 25% limit was applied. The sensitivity analyses only reported the cost per correctly detected case (true cataract and non-cataract case) per year. One-way sensitivity analyses were conducted in order to assess each variable's uncertainty to the model outputs**.** Additionally, multiway sensitivity analyses with combined model variables were also analyzed to generate extreme cases.

4.3 Results

4.3.1 Base-case Analysis

The base-case analysis considered a total population of 22,805 patients who were screened at one of the 75 eye camps over the period of one year in Nepal. The ML-based eye camp could correctly detect an additional 31 cases of cataract (tp), and 2546 additional cases of non-cataract diagnosis (tn) **(Table 4.3)**. In total, the ML-based eye camp can identify 2577 additional correct cases. The average cost-effectiveness ratios per cataract case detected were $23.87 with the ML-based eye camps and $45.89 with the traditional eye camps; the cost per correctly detected case was $0.24 and $0.51 respectively **(Table 4.4)**. In both cost-effectiveness analyses, the traditional eye camp was absolutely (strongly) dominated, this demonstrated that the ML-based eye camp was the more cost-effective method than the traditional eye camp **(Figure 4.2)**. The traditional eye camp is said to be dominated because it is more costly and identifies less correctly detected cases compared to the ML eye camps.

4.3.2 Sensitivity Analysis

One-way sensitivity analysis

The sensitivity analysis evaluated a range of model parameters including the prevalence of cataracts in Nepal, the diagnostic accuracy of ML classifiers and OA, and the labour and capital costs of eye camps. The multiple one-way deterministic sensitivity analysis results are outlined in **Table 4.5** and presented as a tornado diagram (**Figure 4.3**). A one-way sensitivity analysis was conducted for every input in the decision tree and 17 intervals were used to assess the change in ICER and dominance from the lower to upper bound of each parameter's uncertainty. The tornado diagram was plotted to visually display the results from Table 4.5 – the ICER tornado reports the range of ICERs generated for each parameter's uncertain range. Overall, the one-way sensitivity analysis demonstrated that all parameters were stable in all variations, and no variables changed the outcome that ML-based eye camps were the most cost-effective program option.

We observed that the parameter of greatest uncertainty for the model was the cost of labour in the traditional eye camps ($2,444.94 to $4,074.89) which ranged from $0.44 to $0.57 per correctly detected case in traditional eye camp (**Table 4.5)**. The next greatest uncertainty for the one-way analysis was the specificity of ML classifiers, and then the cost of labour in ML eye camps. However, it was observed that all one-way sensitivity scenarios continued to favour the ML-based eye camps as it was the dominant approach in all intervals of the analysis. To test the parameters of diagnostic accuracy of the OAs and ML classifiers, we used the lower and upper 95% confidence intervals into the analysis which continued to demonstrate ML as undominated.

Another important parameter that may influence the analysis is the prevalence of cataracts in Nepal. Based on the 2011 Nepal Blindness Survey, the prevalence of cataracts in the entire population is around 8.5%.[9] However, it is acknowledged that there may be varying distributions and prevalence of cataracts in different regions within a country such as the Chitwan and Nawalparasi districts. The one-way sensitivity analysis evaluated the lower and upper bounds of the cataract prevalence rate (6.63% to 10.63%) and found that the ML eye camps continued to be the cost-effective intervention at $0.23 to $0.25 per correctly detected case, respectively.

Two-way and multi-way sensitivity analysis

A two-way sensitivity analysis was also conducted to estimate any joint influences that two parameters may have together on the cost-effectiveness analysis. We considered multiple joint parameters including the cost of traditional and ML eye camp labour, and the sensitivity of OA and ML classifiers. Similar to the one-way sensitivity analysis, all two-way scenarios demonstrated that ML-based eye camps were dominant (**Appendix K**).

We also conducted a multi-way analysis to produce an extreme scenario case analysis where we varied the multiple parameters at once. The parameters for the variables used for each scenario are outlined in **Appendix L.** In the best-case scenario for ML eye camps, the diagnostic accuracy of ML classifiers was set to the upper 95% CI (sensitivity = 0.987, specificity = 0.980), and the diagnostic accuracy of OAs was set to their lower limits (sensitivity = 0.909, specificity = 0.830). The labour and capital costs of ML-based eye camps were set to their lower limits, while the labour and capital costs of traditional eye camps were set to their higher bound. In the best-case scenario, ML-based camps remained t dominant with the ICER being -$0.89 per additional case detected.

Alternatively, in the worst-case scenario for ML eye camps, the diagnostic accuracy of ML classifiers was set to the lower limits (sensitivity = 0.924, specificity = 0.815), and the diagnostic accuracy of OAs was set to their upper limits (sensitivity = 0.942, specificity = 0.858). Labour and capital costs of the ML-based eye camps were also set to their upper limits, and traditional eye camps was set to their lower limits. In the worst-case scenario, the ICER was $1.21 and there was no dominance in either strategy. Only in this worst-case scenario did we find that the ML eye camps were less effective, but less costly.

4.4 Discussion and Conclusion

To the best of our knowledge, this study is the first to economically evaluate the use of ML classifiers for cataract diagnosis.[32,34] The analysis showed that the ML-based eye camps could have lower personnel costs (and in total costs), while detecting more correctly diagnosed cases. While the ML-based eye camp was able to identify 31 additional cases of cataract in the Nepali population, its power and effectiveness came from its ability to correctly identify an additional 2546 of patients without cataracts that would have been over-diagnosed in a traditional eye camp.

Further, the one-way, two-way, and best-case multiway analysis demonstrated that despite the influences of branch probabilities and costs, ML-based eye camps remain to be the cost-effective strategy. In these specific instances, the negative ICER value depicts that ML-based eye camp was more effective, and less costly; the ICER lies in the southeast quadrant within the cost-effectiveness plane because the ML eye camps were less costly and more effective. Whenever an alternative lies in the southeast quadrant, that intervention is always accepted.[35,36] Only in the worst-case multiway sensitivity analysis did we find that the ML-based eye camps and traditional camps to have no dominance given by an ICER of $1.21. In this worst-case scenario, the ML eye camps were less effective than the traditional camps due to its lowered sensitivity and specificity, however it still demonstrated to be less costly. Both the sensitivity and specificity of the ML eye camps were lower than the traditional camps which suggest that overall, the screening program would detect fewer tn and tp cases. The worst-case scenario for the ML eye camps would be in the southwest quadrant of the cost-effectiveness plane which implied there may be a trade-off between the eye camps.[35]

Eye care services in rural Southeast Asia is scarce, and blindness from cataracts continues to be a pervasive issue due lack of access to eye care services in villages and mountainous regions.[4,6,37,38] A more timely and quick diagnosis of cataracts in rural Nepal can provide patients with better management and treatment of cataracts, and thus lead to an improved quality of life with vision. Although large efforts in Nepal have been made through the implementation of eye camps in the past few decades, the additional implementation of ML may further reduce labour costs, reduce healthcare worker burnout at the camps, and could provide diagnostically accurate results for patients.[25,39,40] Additionally, evaluating the reduced wait times, workload, and patient satisfaction of rural eye camps may be warranted in future studies to assess the full benefits of ML.

Our study should be interpreted within the context of certain limitations. First, the probabilities and assumptions used in the decision tree is based on the current availability of the literature and some uncertainties may exist due to the scarce information on health economics of the use of artificial intelligence.[34] Due to low literacy rates and low levels of education among the population in Nepal, patients may not fully understand or be compliant to the use of ML and there may be hesitancy or low adherence to the ML assessment.[16] Therefore, it is important for patients in the ML eye camp arm to be educated on ML derived medical decisions in order to gain trust and patient compliance.[33] The uncertainties for the model inputs and variables were attempted to be remedied by the multiple sensitivity analyses conducted.

Further limitations related to the uncertainties were the choice of sensitivity analyses conducted for this study. A one-way, two-way, and multi-way sensitivity analysis were included in this paper; however, a probabilistic sensitivity analysis (PSA) was not used. PSA is a method for accounting parameter uncertainty where samples are repeatedly drawn from each distribution to be used as the decision inputs.[41,42] The benefit of using PSA is that it explores joint uncertainty (ie. uncertainty resulting from all parameters simultaneously). In a future cost-effectiveness analysis for ML implementation, PSA can be used as the preferred analysis for parameter uncertainties.[41] Also, another limitation, for the sensitivity analysis in this paper, we applied an upper and lower 25% bound for the probability and cost inputs.

The sensitivity and specificity values of the ML classifier that were retrieved from Chapter 03 was also subject to limitations in this study. The ML diagnostic accuracy values from the meta-analysis were an aggregate of all imaging modalities available from the literature which included slit-lamp, fundus, and visible wavelength images. For the ML-based eye camp model in this study, we assumed that only slit-lamp microscopy was performed for each patient to take an image of the anterior segment of the patients' eyes. Therefore, there may be limitations that exist in those results due to the heterogeneous nature of the diagnostic performance of ML. In future analysis, we hope that there are more diagnostic accuracy studies of ML classifiers that uses slit-lamp images in order for a subgroup meta-analysis with sufficient power to be conducted.

Additionally, this CEA took on a health policy and program implementation perspective, therefore the effectiveness measure was the ability to identifying a correctly detected case (tp and tn). In future studies, it would be beneficial to also assess the visual deterioration in these patients by considering their quality-adjusted life year measure in order to assess the full effectiveness of the ML-based eye camps.

The results of this study may only be able to capture rural locations where access to eye care services is extremely limited, and there is no public insurance or health coverage for these types of ophthalmic services. Additionally, the population and regions that this study considers was the Chitwan and Nawalparasi Districts of South-central Nepal, so the results may not be generalization. However, this methodology could be applied for the provincial level if respective data is gathered. Therefore, further studies can assess the ability to implement ML classifiers for cataracts outside of the eye camp framework, and into other clinics and locations in semiurban to urban regions.

In conclusion, the results of this study demonstrated the practical and economic feasibility for a ML-based screening program to be implemented in existing eye camps in rural Nepal over the existing eye camp models. The implementation of ML in the Aravind eye camps may be utilized in other rural regions and LMICs. Both the patients and healthcare workers at the eye camps could benefit from the implementation of this program, and healthcare organizations in Nepal could consider investing in this type of program because of its cost savings.

4.5 References

1. Liu YC, Wilkins M, Kim T, Malyugin B, Mehta JS. Cataracts. The Lancet [Internet]. 2017;390(10094):600–12. Available from: http://www.sciencedirect.com/science/article/pii/S0140673617305445

2. Allen D, Vasavada A. Cataract and surgery for cataract. The British Medical Journal [Internet]. 2006 Jul 13;333(7559):128 LP – 132. Available from: http://www.bmj.com/content/333/7559/128.abstract

3. Asbell PA, Dualan I, Mindel J, Brocks D, Ahmad M, Epstein S. Age-related cataract. The Lancet [Internet]. 2005;365(9459):599–609. Available from: http://www.sciencedirect.com/science/article/pii/S0140673605179112

4. Murthy GVS, John N, Shamanna BR, Pant HB. Elimination of avoidable blindness due to cataract: where do we prioritize and how should we monitor this decade? Indian Journal of Ophthalmology [Internet]. 2012;60(5):438.

5. Evans J, Mwangi N, Burn H, Ramke J. Equity was rarely considered in Cochrane Eyes and Vision systematic reviews and primary studies on cataract. Journal of Clinical Epidemiology. 2020;125:57–63.

6. Thulasiraj RD, Nirmalan PK, Ramakrishnan R, Krishnadas R, Manimekalai TK, Baburajan NP, et al. Blindness and vision impairment in a rural south Indian population: the Aravind Comprehensive Eye Survey. Ophthalmology [Internet]. 2003 Aug 1;110(8):1491–8. Available from: https://doi.org/10.1016/S0161-6420(03)00565-7

7. He M, Wang W, Huang W. Variations and Trends in Health Burden of Visual Impairment Due to Cataract: A Global Analysis. Investigative Ophthalmology & Visual Science [Internet]. 2017 Aug 28;58(10):4299–306. Available from: https://doi.org/10.1167/iovs.17-21459

8. Bourne R, Steinmetz JD, Flaxman S, Briant PS, Taylor HR, Resnikoff S, et al. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study. The Lancet Global Health [Internet]. 2021;9(2):e130–43. Available from: https://www.sciencedirect.com/science/article/pii/S2214109X20304253

9. Nepal Netra Jyoti Sangh. Epidemiology of Blindness in Nepal [Internet]. 2011. Available from: https://www.iapb.org/learn/resources/epidemiology-of-blindness-in-nepal/

10. Kandel RP, Rajashekaran SR, Gautam M, Bassett KL. Evaluation of alternate outreach models for cataract services in rural Nepal. BMC Ophthalmology [Internet]. 2010;10(1):9. Available from: https://doi.org/10.1186/1471-2415-10-9

11. Collon S, Chang D, Tabin G, Hong K, Myung D, Thapa S. Utility and Feasibility of Teleophthalmology Using a Smartphone-Based Ophthalmic Camera in Screening Camps in Nepal. Asia Pacific Journal of Ophthalmology [Internet]. 2020;9(1):54–8. Available from: https://pubmed.ncbi.nlm.nih.gov/31990747

12. Adhikari B, Mishra SR, Schwarz R. Transforming Nepal's primary health care delivery system in global health era: addressing historical and current implementation challenges. Globalization and Health [Internet]. 2022;18(1):8. Available from: https://doi.org/10.1186/s12992-022-00798-5

13. Thapa R, Bam K, Tiwari P, Sinha TK, Dahal S. Implementing Federalism in the Health System of Nepal: Opportunities and Challenges. International Journal Health Policy Management [Internet]. 2019 Apr 1;8(4):195–8. Available from: https://pubmed.ncbi.nlm.nih.gov/31050964

14. Mishra SR, Khanal P, Karki DK, Kallestrup P, Enemark U. National health insurance policy in Nepal: challenges for implementation. Glob Health Action [Internet]. 2015 Aug 21;8:28763. Available from: https://pubmed.ncbi.nlm.nih.gov/26300556

15. WHO. Health Expenditure Profile: Nepal [Internet]. Global Health Expenditure Database. 2022 [cited 2022 Apr 12]. Available from: https://apps.who.int/nha/database/country_profile/Index/en

16. Government of Nepal. National Population and Housing Census 2011 (National Report). 2012.

17. Aravind Eye Care System. About Aravind [Internet]. Aravind Eye Care System. 2022 [cited 2022 Mar 10]. Available from: https://aravind.org/our-story/

18. Kharel Sitaula R, Joshi SN, Khanal S. Surgical eye camp in rural area of Nepal and its role in Vision 2020. Journal of Chitwan Medical College [Internet]. 2017 Feb 20;6(2 SE-Original Research Articles):1–5. Available from: https://www.nepjol.info/index.php/JCMC/article/view/16677

19. Venkataswamy G, Lepkowski JM, Ravilla T, Brilliant GE, Shanmugham CA, Vaidyanathan K, et al. Rapid epidemiologic assessment of cataract blindness. The Aravind Rapid Epidemiologic Assessment Staff. Int J Epidemiol. 1989;18(4 Suppl 2):S60-7.

20. Brilliant GE, Brilliant LB. Using social epidemiology to understand who stays blind and who gets operated for cataract in a rural setting. 1985.

21. Cheung R, So S, Malvankar-Mehta MS. Diagnostic accuracy of machine learning classifiers for cataracts: a systematic review and meta-analysis. University of Western Ontario. 2021 (*unpublished*)

22. Zhang H, Niu K, Xiong Y, Yang W, He Z, Song H. Automatic cataract grading methods based on deep learning. Computer Methods Programs Biomed [Internet]. 2019;182(doh, 8506513 PG-104978):104978.

23. Xu X, Zhang L, Li J, Guan Y, Zhang L. A Hybrid Global-Local Representation CNN Model for Automatic Cataract Grading. IEEE Journal of Biomedical and Health Informatics. 2020;24(2):556–67.

24. Shimizu E, YAZU H, Aketa N, Tanji M, Sakasegawa A, Nakayama S, et al. Innovative artificial intelligence-based cataract diagnostic method uses a slit-lamp video recording device and multiple machine-learning. Investigative Ophthalmology & Visual Science. 2021 Jun 21;62(8):1031.

25. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nature Biomedical Engineering [Internet]. 2018;2(10):719–31. Available from: https://doi.org/10.1038/s41551-018-0305-z

26. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J. 2019;6(2):94.

27. Datta S, Barua R, Das J. Application of artificial intelligence in modern healthcare system. Alginates—recent uses of this natural polymer. 2020.

28. Wu X, Huang Y, Liu Z, Lai W, Long E, Zhang K, et al. Universal artificial intelligence platform for collaborative management of cataracts. British Journal of Ophthalmology [Internet]. 2019;103(11 PG-1553–1560):1553–60.

29. Cheung R, Yu B, Iordanous Y, Malvankar-Mehta MS. The Prevalence of Occupational Burnout Among Ophthalmologists: A Systematic Review and Meta-Analysis. Psychological Reports [Internet]. 2020 Aug 31;0033294120954135. Available from: https://doi.org/10.1177/0033294120954135

30. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. BMC Medicine [Internet]. 2013;11(1):80. Available from: https://doi.org/10.1186/1741-7015-11-80

31. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford university press; 2015.

32. Ruamviboonsuk P, Chantra S, Seresirikachorn K, Ruamviboonsuk V, Sangroongruangsri S. Economic Evaluations of Artificial Intelligence in Ophthalmology. The Asia-Pacific Journal of Ophthalmology [Internet]. 2021;10(3). Available from: https://journals.lww.com/apjoo/Fulltext/2021/06000/Economic_Evaluations_of_Artificial_Intelligence_in.10.aspx

33. Soellner M, Koenigstorfer J. Compliance with medical recommendations depending on the use of artificial intelligence as a diagnostic method. BMC Medical Informatics and Decision Making [Internet]. 2021;21(1):236. Available from: https://doi.org/10.1186/s12911-021-01596-6

34. Wolff J, Pauling J, Keck A, Baumbach J. The Economic Impact of Artificial Intelligence in Health Care: Systematic Review. J Med Internet Res [Internet]. 2020 Feb 20;22(2):e16866–e16866. Available from: https://pubmed.ncbi.nlm.nih.gov/32130134

35. Klok RM, Postma MJ. Four quadrants of the cost-effectiveness plane: some considerations on the south-west quadrant. Vol. 4, Expert review of pharmacoeconomics & outcomes research. Taylor & Francis; 2004. P. 599–601.

36. Bambha K, Kim WR. Cost-effectiveness analysis and incremental cost-effectiveness ratios: uses and pitfalls. Eur J Gastroenterol Hepatol. 2004;16(6):519–26.

37. WHO. Blindness and vision impairment [Internet]. 2021 [cited 2021 Nov 17]. Available from: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

38. Khairallah M, Kahloun R, Bourne R, Limburg H, Flaxman SR, Jonas JB, et al. Number of People Blind or Visually Impaired by Cataract Worldwide and in World Regions, 1990 to 2010. Investigative Ophthalmology & Visual Science [Internet]. 2015 Oct 20;56(11):6762–9. Available from: https://doi.org/10.1167/iovs.15-17201

39. Tseng RMWW, Gunasekeran DV, Tan SSH, Rim TH, Lum E, Tan GSW, et al. Considerations for Artificial Intelligence Real-World Implementation in Ophthalmology: Providers' and Patients' Perspectives. The Asia-Pacific Journal of Ophthalmology [Internet]. 2021;10(3). Available from: https://journals.lww.com/apjoo/Fulltext/2021/06000/Considerations_for_Artificial_Intelligence.9.aspx

40. Kuziemsky C, Maeder AJ, John O, Gogia SB, Basu A, Meher S, et al. Role of Artificial Intelligence within the Telehealth Domain. Yearbook of Medical Informatics [Internet]. 2019/04/25. 2019 Aug;28(1):35–40. Available from: https://pubmed.ncbi.nlm.nih.gov/31022750

41. Jain, R, Grabner, M, Onukwugha, E. Sensitivity Analysis in Cost-Effectiveness Studies. PharmacoEconomics [Internet].2011; 29, 297–314. Available from: https://doi.org/10.2165/11584630-000000000-00000

42. McCabe, C, Paulden, M, Awotwe, I, Sutton A, Hall P. One-Way Sensitivity Analysis for Probabilistic Cost-Effectiveness Analysis: Conditional Expected Incremental Net Benefit. PharmacoEconomics [Internet]. 2020; 38, 135–141. Available from: https://doi.org/10.1007/s40273-019-00869-3

43. Brown GC. Vision and quality-of-life. Transactions of the American Ophthalmological Society. 1999;97:473-511.

4.6 Tables and Figures

Table 4.1 Base case model parameters and parameter ranges

Parameter	Value	Range	
Fixed Data elements			
Patients screened in study setting[10]	22,805 patients	-	-
Patient compliance to OA-based decision[30]	100%	-	-
Patient compliance to AI-based decision[30]	100%	-	-
Variable data elements			
Prevalence of any cataract in Nepal[9]	8.50%	6.38%	10.63%
ML assessment parameters[17]			
Sensitivity	0.948	0.815	0.987
Specificity	0.96	0.924	0.98
OA parameters[15]			
Sensitivity	0.932	0.909	0.942
Specificity	0.838	0.83	0.858

Table 4.2 Estimated costs for traditional eye camps and ML-based eye camps

Item	Cost of Traditional DST Camp	Range	Cost of AI-based DST Camp	Range
Labour Costs				
Salary for OA and Nurses	$1,847.37	$1,385.53 - $2,309.21	$683.42	$512.57 - $854.28
Wage supplement for OA and nurses for rural initiative	$1,412.54	$1,059.40 - $1,765.68	$711.30	$533.47 - $889.12
Capital Costs				
Equipment and other consumables	$116.56	$87.42 - $ 145.70	$116.56	$87.42 - $145.70
Maintenance, repairs, insurance	$629.52	$472.14 - $786.90	$629.52	$472.14 - $786.90
Logistical Costs				
Publicity	$262.97	-	$262.97	-
Transportation	$1,895.72	-	$1,895.72	-
Total Costs	$6,164.69		$4,299.49	

Table 4.3 Diagnostic test outcomes of traditional eye camps and ML-based eye camps per 22,805 patients in the study model

Measure	Traditional eye camp (OA Assessment)	ML-based eye camp (ML Assessment)
Total patients screened	22805	22805
True-positive (tp) result	1807	1838
True-negative (tn) result	17486	20032
False-positive (fp) result	132	101
False-negative (fn) result	3380	835

Table 4.4 Incremental cost-effectiveness results for traditional eye camps vs ML-based eye camp

Eye camp	Cost per patient ($)	Incremental cost	Effectiveness (probability of a case detected)	Incremental effectiveness	Cost effectiveness	ICER	Dominance
Cost per cataract case detected (true-positive cases)							
ML-based eye camp	0.189		0.006493		23.87		Undominated
Traditional eye camp	0.288	0.099	0.006276	-0.000217	45.89	-457.76	Absolutely dominated
Cost per case correctly detected (true-positive and true-negative cases)							
ML-based eye camp	0.189		0.777		0.243		Undominated
Traditional eye camp	0.288	0.099	0.567	-0.210	0.510	-0.475	Absolutely dominated

Table 4.5 One-way deterministic sensitivity analysis results

Parameter	Base-case Value	Range	ICER ($/case correctly detected)
Prevalence of any cataract	0.085	0.0638 to 0.1063	-$0.40 to -$0.51
Diagnostic accuracy of ophthalmic assistants			
Sensitivity	0.932	0.909 to 0.942	-$0.47 to -$0.48
Specificity	0.838	0.830 to 0.858	-$0.45 to -$0.57
Diagnostic accuracy of machine learning classifiers			
Sensitivity	0.948	0.815 to 0.987	-$0.48 to -$0.47
Specificity	0.96	0.924 to 0.980	-$0.67 to -$0.41
Labour Costs			
Traditional Eye Camp	$3,259.91	$2,444.94 to $4,074.89	-$0.29 to -$0.66
ML-based Eye Camp	$1,394.72	$1,046.04 to $1743.40	-$0.55 to -$0.40
Capital Costs			
Traditional Eye Camp	$746.08	$559.56 to $ 932.60	-$0.43 to -$0.52
ML-based Eye Camp	$746.08	$559.56 to $ 932.60	-$0.51 to -$0.44

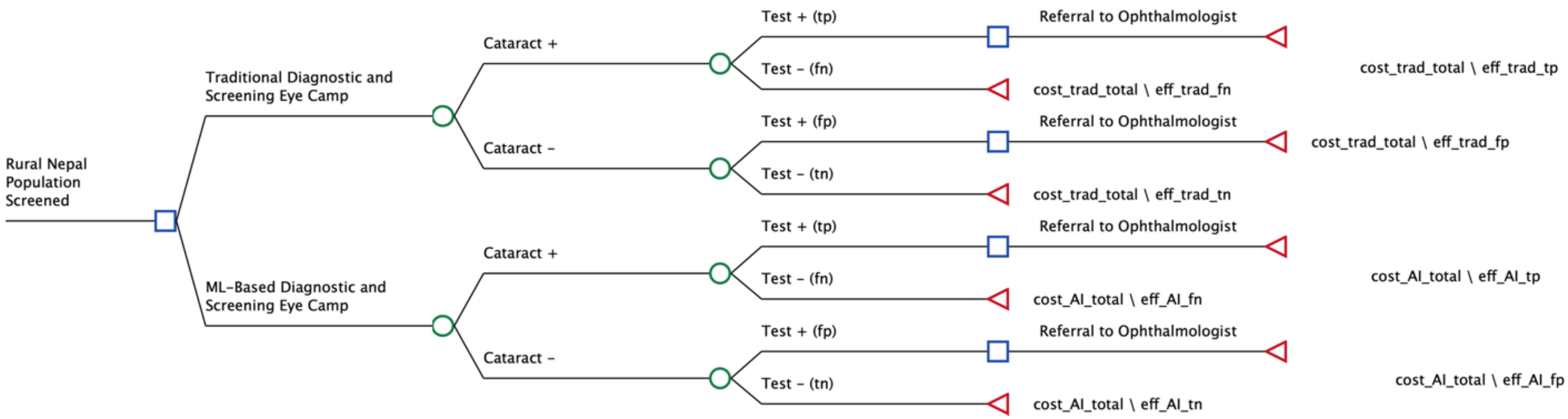


Figure 4.1 Decision tree showing the competing alternatives for cataract diagnosis and screening camps. Arm 1 illustrates the traditional eye camps; Arm 2 illustrates to the ML-based eye camps evaluated in the model. A square represents a decision node, a circle is a chance node, and the triangle represents the terminal node.

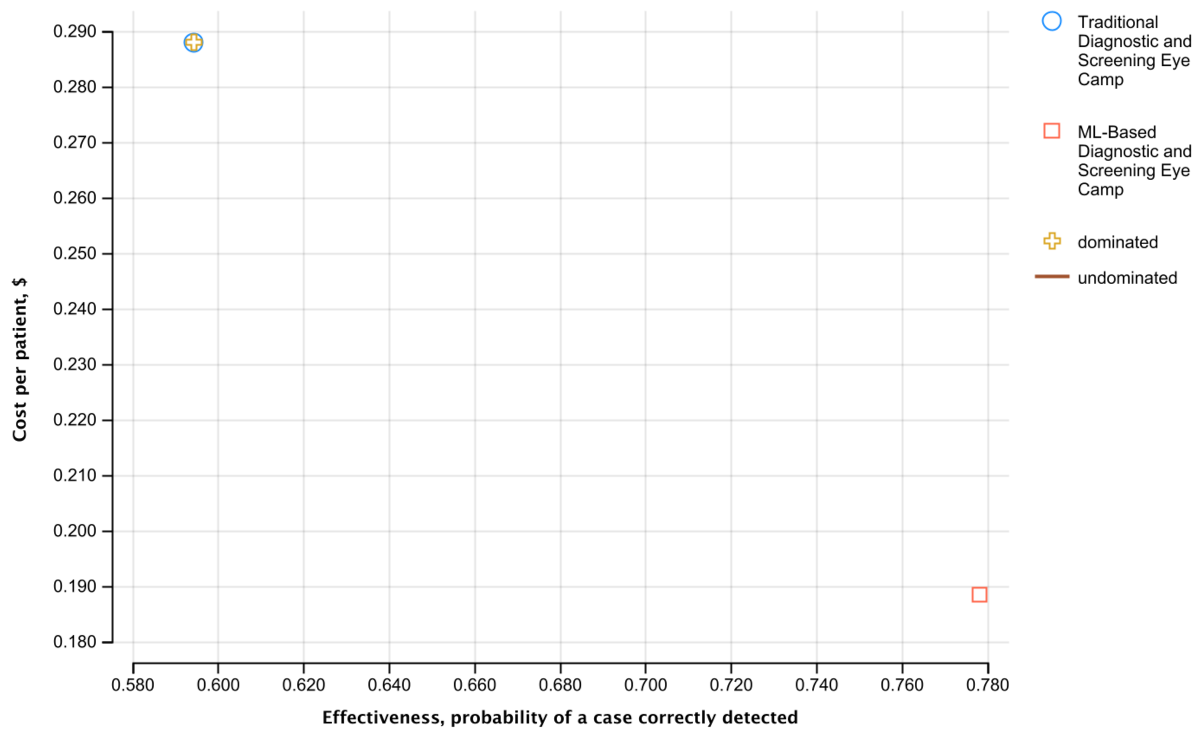


Figure 4.2 Cost-effectiveness plane of traditional eye camps vs introduction of ML-based eye camps. The traditional eye camp program is absolutely dominated by the ML-based eye camps. The blue circle represents the traditional eye camps which is dominated, and the red square represents the ML-based eye camps.

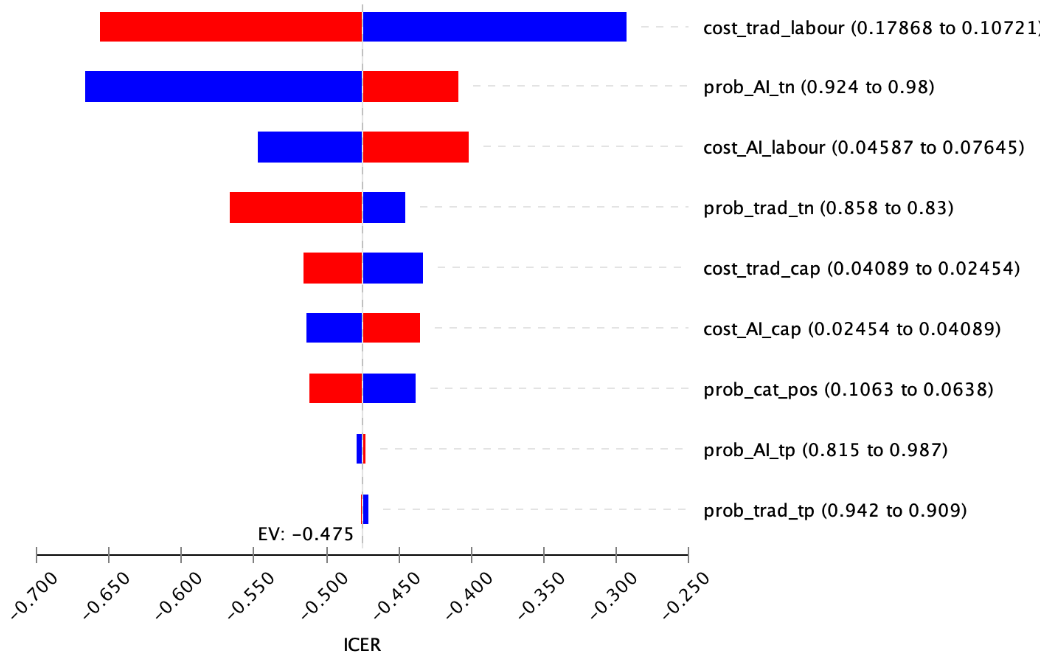


Figure 4.3 Tornado diagram of the one-way deterministic sensitivity analysis with critical variables. The vertical line represents the final ICER (-0.475). All the variables crossed the final ICER value which demonstrates that none of the parameters disturbs the final result from the base-case. The blue bar represents the change in direction from the baseline ICER of the lower bound, and the red bar represents the change of the upper bound of each parameter.

CHAPTER 5

5 Integrated Discussion

5.1 Overview

This chapter will discuss the interpretations and implications of the results from the meta-analysis of diagnostic studies, and the health economic analysis of the implementation of ML classifiers for cataract diagnosis. The objectives of this thesis were to (1) assess the diagnostic accuracy of ML classifiers, and (2) use the findings from objective 1 to explore the cost-effectiveness of ML assessment versus a traditional assessment in diagnostic and screening eye camps in rural Nepal.

5.2 Integrated Discussion of Results

Chapter 03 was the first study that investigated the diagnostic accuracy of machine learning classifiers for cataracts based on the current literature available. The study aimed to systematically review the current body of knowledge for the types of machine learning classifiers that researchers are using, and to meta-analyze the pooled sensitivity and specificity of these classifiers. The most recent database search conducted by the review was on September 12, 2021, and no limits were placed on the publication date of the studies.

Based on literature from both published and unpublished sources, 21 studies met the inclusion criteria, and 11 of those studies were used for the meta-analysis. The primary analysis of the study was investigating the adult cataract population, and a secondary analysis (sub-group analysis) was conducted for the pediatric cataract population. In both analyses through a hierarchical logistic regression, we observed high diagnostic accuracy in both the sensitivity and specificity of machine learning classifiers. The summary estimate for sensitivity was 0.948 [95% CI: 0.815-0.987] and specificity was 0.960 [95% CI: 0.924-0.980] for adult cataracts screening using a ML classifier. Additionally, similar results were found for pediatric cataracts. Compared to the diagnostic accuracy of OA in eye camps (sensitivity = 0.932, specificity = 0.838), the ML classifier outperforms the OA with higher sensitivity and specificity.[1] This suggests the potential for machine learning that could be used to screening patients with cataracts in rural settings.[2]

Using the results obtained from Chapter 03, the cost-effectiveness analysis in Chapter 04 utilizes the pooled sensitivity and specificity of machine learning classifiers to be implemented in eye camps in rural Nepal. These estimates and probabilities inserted into the decision tree was suitable for the model because all studies were identifying for "any cataract" (<6/18 to 6/60 in the worse eye and 6/60 in better eye).[3] Specifically, in the Chitwan and Nawalparasi Districts of Nepal, existing eye camps have been implemented in the past few decades In order to provide rural Nepalis with eye care services.[4] The CEA in Chapter 04 demonstrated that ML provides a cheaper and effective outcome over the traditional eye camp Aravind model. Sensitivity analyses were conducted to consider the variability of certain parameters included in the economic model.

We observed that the use of ML in eye camps in rural Nepal can identify 2577 additional correct cases. The cost per correctly detected case was \$0.24 and \$0.51 for the ML-based and traditional eye camp, respectively. The sensitivity analyses showed that in all scenarios except the worse-case scenario, ML-based eye camps were dominant. The assumption that a fully automated ML assessment in the current population could be likely because OAs are burning out in their current working conditions and eye care specialists are almost non-existent in these rural and mountainous regions.[5,6] The ICER was negative in all cases which suggests that we could accept this form of eye care delivery.[7] This study shows the potential and the feasibility for this type of technology to be implemented in rural Nepal.

5.3 Thesis Limitations

Brief limitations of each study were discussed in Chapter 03 and Chapter 04 respectively, but this following section provides additional limitations to consider in both studies.

5.3.1 Limitations in Chapter 03

Firstly, the limitation of a meta-analysis exists in its inclusion criteria and there are inherent biases in observational studies that could exist as stated by Egger et al.[8] In the body of literature in ML diagnosis for cataracts, there is a limited number of studies published, thus we included all studies regardless of risk of bias domains being scored as medium or high in the QUADAS-2 tool. Fortunately, all 21 included studies generally presented with low risk of bias, with only a few studies receiving a medium or high risk of bias in the 'flow and timing' and 'patient selection' domains.[9]

Additionally, as mentioned in Chapter 03, a random-effects model was utilized for the meta-analysis due to the high heterogeneity found across studies. Heterogeneity across the studies could exist due to location of the study (geographical locations), varying effect sizes across studies, precision of each effect size, methodology of each study, and study design.[8,10] It is noted, however, that regardless of the heterogeneity found from the analysis, a random-effects model would have been utilized.

In our statistical analysis, all types of ML classifiers were included and aggerated. Each type of algorithm has different methodological considerations. Additional subgroup analysis could be conducted if there were more studies found and included in the meta-analysis for each type of classifier. Future research could focus on sub-group analysis by type of ML classifier. We included both slit-lamp imaging and fundus imaging in our analysis, and we recognize that the types of images used to train an algorithm could be different.

5.3.2 Limitations in Chapter 04

As mentioned in Chapter 04, this study was the first CEA of any diagnostic ML algorithm for cataracts and there is limited availability of literature, thus assumptions were made to develop the economic model. We assumed that all images that were taken by the slit-lamp microscope were readable by the algorithm. An error can occur when the algorithm cannot produce an output because the images are blurry and cannot detect an image properly. The assumption was that the ophthalmic nurses or OA were properly trained, which would be the case, to take clear images to input them into the algorithm and would retake an image if done improperly. Due to the high demand for eye care services in villages, we assumed that patients would be compliant with a machine learning assessment as an OA assessment.[11]

In addition, we took a health policy and program implementation approach for this analysis, thus the effectiveness measured the number of cases correctly detected. This clinical outcome describes an intermediate end point for economic studies. Typically, life-years gained, and quality of life measures are used as final end point outcomes, to assess the patient's benefits or burdens by the program or intervention.[12] The critique in having an intermediate end point in a CEA is the difficulty of establishing a relationship between the intermediate and final end points. However, as stated by Drummond et al., there may be value in an intermediate end point for diagnostic tests when a long-term cost-effectiveness may be achieved from the intermediate outcome.[12]

Further, this study did not consider the patient-specific benefits of implementing an ML-based eye camp such as decreased patient wait times, decreased patient anxiety, stress and burden, increased patient productivity and economic opportunity, and quick and accurate diagnostic outcome.[13–15]

For this study, we only considered rural regions for the implementation of ML classifiers due to the existing nature of temporary eye camps. It may not be feasible to implement this program in locations such as Ontario where the public health insurance covers annual eye examinations for certain groups of people.

5.3.3 Limitations in the Applications of Artificial Intelligence

Although there is great potential for AI, there are also pitfalls in the use of these algorithms in the healthcare setting including algorithm bias. Algorithm bias and data privacy are examples are the increasing concerns with the full implementation of AI and ML in healthcare settings.[16,17]

Algorithm bias refers to the unwarranted skewing of the output results of an algorithm due to problems from the initial algorithm development and design.[18,19] The type of data that the algorithm is being taught and validated can make a large impact on the output that the algorithm produces. [17,18,20] Algorithm bias presents as a critical consideration when implementing AI technologies into the healthcare setting because there are evident examples of discrimination and bias of race, sex, and socio-economic status in the algorithm's outputs.[21] These biased (and perhaps discriminatory) outputs may exclude certain intersections of patients from receiving a necessary treatment or intervention and can prevent certain groups of people to have reasonable access to insured hospital and physician services, simply due to the inadequate dataset that is initially inputted into the algorithm.[21] If these people are discriminated against by the algorithms, then there are barriers to their access to care and health services.[22]

Commonly, a working algorithm can be sold or shared globally without considering the source of the training inputs.[23] Therefore, an AI algorithm that is developed and trained in one country can be used in another country with different patient characteristics and demographics. This presents a critical challenge when implementing an algorithm in the healthcare setting because the algorithm can provide an inaccurate diagnosis of a certain disease for a population of a different race, ethnicity, or body type.[18,23]

However, there is added strength in the sensitivity and specificity estimates that were obtained from the meta-analysis because the meta-analysis included studies that contained databases of populations that many come from the neighbouring countries of Nepal such as China and India. The included studies come from these Asian countries, which may increase the validity of our results to the Nepali population.

Additionally, this study tries to remedy the issues of algorithm bias through the distribution of benefits and burdens of the diagnostic outcomes and probabilities. A term called "distributive justice" is common in the realm of algorithm bias when referring to having action towards fairness — this can in the form of modification, adjustment, and redress of the algorithm.[21] In order to clearly outline the benefits or burdens of a diagnostic algorithm, a single accuracy proportion does not suffice. Often in many diagnostic accuracy studies of ML classifiers, only the F-score or F1-score is reported as the accuracy value for the classifier.[21] The F-score is a function of the positive predictive value (PPV) and the sensitivity which are derived from the tp, fp, and fn values. Therefore, the tn value is often disregarded in many accuracy studies of ML classifiers and it can bring a very important insight on the effectiveness of the screening program as a whole. Reporting all the true negative, true positive, false negative, and false positive values is very beneficial for policy driven data because it considers all aspects of the diagnostic accuracy.[19] This allows for the data to be transparent, and values have been clearly outlined throughout the thesis in both Chapter 03 and Chapter 04.

5.4 Conclusions and Future Directions

In conclusion, this thesis found that ML classifiers are more diagnostically accurate for cataracts than ophthalmic assistants and the implementation of a ML-based eye camp is more cost-effective than traditional eye camps in rural Nepal. Due to the high heterogeneity of the meta-analysis, and the assumptions made for the CEA, we should be cautious of the results. It may be feasible for existing eye camps in rural regions to implement ML in their eye camps if there is a safe and approved ML algorithm publicly available for cataracts. Health policymakers and healthcare organizations could continue to consider the benefits of digital medicine, and the positive impact that it can have on both the healthcare and patient perspective. In addition to the current guidance and implementation of ML for cataracts in the standard clinical setting, this thesis emphasizes the potential, feasibility, and need for ML classifiers to be expanded to other ocular diseases.

Although this thesis has identified gaps in the literature, and it has summarized the current body of literature on all ML classifiers for cataracts, there is still research and further investigations warranted. When there exists more research on ML classifiers and their diagnostic performance, further studies with multiple sub-group analyses based on ML classifier type and imaging modality can evaluate which presents the highest accuracy. It is also encouraged for future authors to report their tn, tp, fp, and fn values for the results to be properly aggregated by the meta-analysis if possible. Additionally, the CEA is applicable towards rural eye camps. Therefore, future economic analysis may be interested in testing the feasibility of ML being implemented in more urban areas in order to evaluate whether the benefits seen in villages are equivalent.

5.5 References

1. Venkataswamy G, Lepkowski JM, Ravilla T, Brilliant GE, Shanmugham CA, Vaidyanathan K, et al. Rapid epidemiologic assessment of cataract blindness. The Aravind Rapid Epidemiologic Assessment Staff. Int J Epidemiol. 1989;18(4 Suppl 2):S60-7.

2. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nature Biomedical Engineering [Internet]. 2018;2(10):719–31. Available from: https://doi.org/10.1038/s41551-018-0305-z

3. Kharel Sitaula R, Joshi SN, Khanal S. Surgical eye camp in rural area of Nepal and its role in Vision 2020. Journal of Chitwan Medical College [Internet]. 2017 Feb 20;6(2

SE-Original Research Articles):1–5. Available from: https://www.nepjol.info/index.php/JCMC/article/view/16677

4. Kandel RP, Rajashekaran SR, Gautam M, Bassett KL. Evaluation of alternate outreach models for cataract services in rural Nepal. BMC Ophthalmology [Internet]. 2010;10(1):9. Available from: https://doi.org/10.1186/1471-2415-10-9

5. Collon S, Chang D, Tabin G, Hong K, Myung D, Thapa S. Utility and Feasibility of Teleophthalmology Using a Smartphone-Based Ophthalmic Camera in Screening Camps in Nepal. Asia Pac J Ophthalmol (Phila) [Internet]. 2020;9(1):54–8. Available from: https://pubmed.ncbi.nlm.nih.gov/31990747

6. Nepal Netra Jyoti Sangh. Epidemiology of Blindness in Nepal [Internet]. 2011. Available from: https://www.iapb.org/learn/resources/epidemiology-of-blindness-in-nepal/

7. Klok RM, Postma MJ. Four quadrants of the cost-effectiveness plane: some considerations on the south-west quadrant. Vol. 4, Expert review of pharmacoeconomics & outcomes research. Taylor & Francis; 2004. p. 599–601.

8. Egger M, Smith GD, Altman D. Systematic reviews in health care: meta-analysis in context. John Wiley & Sons; 2008.

9. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct;155(8):529–36.

10. Borenstein M, Hedges L v, Higgins JPT, Rothstein HR. Introduction to meta-analysis. John Wiley & Sons; 2021.

11. Soellner M, Koenigstorfer J. Compliance with medical recommendations depending on the use of artificial intelligence as a diagnostic method. BMC Medical Informatics and Decision Making [Internet]. 2021;21(1):236. Available from: https://doi.org/10.1186/s12911-021-01596-6

12. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford university press; 2015.

13. Canadian Medical Association. The economic cost of wait times in Canada. Centre for Spatial Economics. 2008.

14. Globerman S, Esmail N, Day B, Henderson D. Reducing wait times for health care: what Canada can learn from theory and international experience. Fraser Institute, October. 2013;

15. Cheung R, Yu B, Iordanous Y, Malvankar-Mehta MS. The Prevalence of Occupational Burnout Among Ophthalmologists: A Systematic Review and Meta-Analysis. Psychological Reports [Internet]. 2020 Aug 31;0033294120954135. Available from: https://doi.org/10.1177/0033294120954135

16. Martin-Bariteau F, Scassa T. Artificial Intelligence and the Law in Canada. Artificial Intelligence and the Law in Canada (Toronto: LexisNexis Canada, 2021). 2020.

17. Zuiderveen Borgesius F. Discrimination, artificial intelligence, and algorithmic decision-making. 2018.

18. Henderson B, Flood CM, Scassa T. Artificial Intelligence in Canadian Healthcare: Will the Law Protect Us from Algorithmic Bias Resulting in Discrimination? Bradley Henderson, Colleen M Flood, and Teresa Scassa,“Artificial Intelligence in Canadian Healthcare: Will the Law Protect Us from Algorithmic Bias Resulting in Discrimination. 2021;19:2.

19. Baker JE. Reducing bias and inefficiency in the selection algorithm. In: Proceedings of the second international conference on genetic algorithms. 1987. p. 14–21.

20. Goldenberg A, Flood C. AI in Healthcare: What’s Law Got to Do With It? [Internet]. CDTS/CLTS uOttawa. 2020 [cited 2021 Dec 8]. Available from: https://www.youtube.com/watch?v=T76TFctCGVA

21. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science (1979). 2019;366(6464):447–53.

22. Deshaies B, Hall D. Responsible use of automated decision systems in the federal government [Internet]. Statistics Canada. 2021 [cited 2021 Dec 9]. Available from: https://www.statcan.gc.ca/en/data-science/network/automated-systems

23. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance [Internet]. 2021 [cited 2021 Dec 9]. Available from: https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200-eng.pdf

24. Government of Nepal. National Population and Housing Census 2011 (National Report). 2012.

APPENDICES

Appendix A: PRISMA Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	50
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	n/a
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	51
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	52
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	52
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	52
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	53
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	54
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	54
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	54

Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	56
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	55
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	56
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	56

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	56
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	56
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	57
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	57
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	58
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	59
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	59
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	58
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	60
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their	61

		relevance to key groups (e.g., healthcare providers, users, and policy makers).	
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	62
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	63
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	n/a

Appendix B: Database searches and keywords

Concept	**EMBASE**	**MEDLINE**	**CINAHL**	**Keywords**
Artificial Intelligence	Exp artificial intelligence/ or exp machine learning/ or exp deep learning/ or artificial neural network/	Exp artificial intelligence/ or exp machine learning/ or exp deep learning/ or exp Neural Networks, Computer	(MH "Artificial Intelligence+") or (MH "Deep Learning") or (MH "Neural Networks (Computer)") or (MH "Machine Learning+")	Artificial intelligence.mp. or machine learning.mp. or deep learning.mp. or neural network.mp.
Diagnosis	Exp diagnosis/ or exp prediction/	Exp diagnosis/	(MH "Diagnosis")	Diagnos*.mp. or prediction.mp.
Cataract	Exp cataract/	Exp cataract/	(MH "Cataract")	cataract.mp.
No Limits	110	40	19	
Limit to Humans and English Language	91	31	19	

Appendix C. Search strategy and results for grey literature

ProQuest Dissertations & Theses

Searches	**Results**
noft(artificial intelligence) AND noft(diagnosis) AND noft(ophthalmology)	5

Conference Searches

Conference	Link	Years searched	Search terms	Results/Comments
ARVO	https://arvojournals.org/solr/searchresults.aspx?q=meeting%20abstract%20AND%20artificial%20intelligence%20AND%20cataracts&restypeid=1	All years	“meeting abstract” AND “artificial intelligence” AND “cataracts”	Searched through meeting abstracts 23 results
AAO	https://secure.aao.org/aao/meeting-archive	All years	“artificial intelligence” Event type: Paper	Searched through all meetings and scientific posters 5 results
COS	https://www.cos-sco.ca/cpd/annual-meeting/	All years available	“artificial intelligence” “machine learning” “diagnosis”	Searched through abstracts and presentations 2 results

Appendix D. Kappa statistics calculation

Title and Abstract Screening

Review Authors	**RC**				
		Include	**Exclude**	**Unsure**	**Total**
SS	**Include**	22	2	0	**24**
	Exclude	4	100	2	**106**
	Unsure	0	5	0	**5**
	Total	**26**	**107**	**2**	**135**

P_0 0.903704

P_E 0.657119

kappa **0.719155**

Full Text Screening

Review Authors	RC				
		Include	Exclude	Unsure	Total
SS	Include	18	0	0	**18**
	Exclude	2	8	0	**10**
	Unsure	0	0	0	**0**
	Total	**20**	**8**	**0**	**28**

P_0 0.928571

P_E 0.561224

kappa **0.837209**

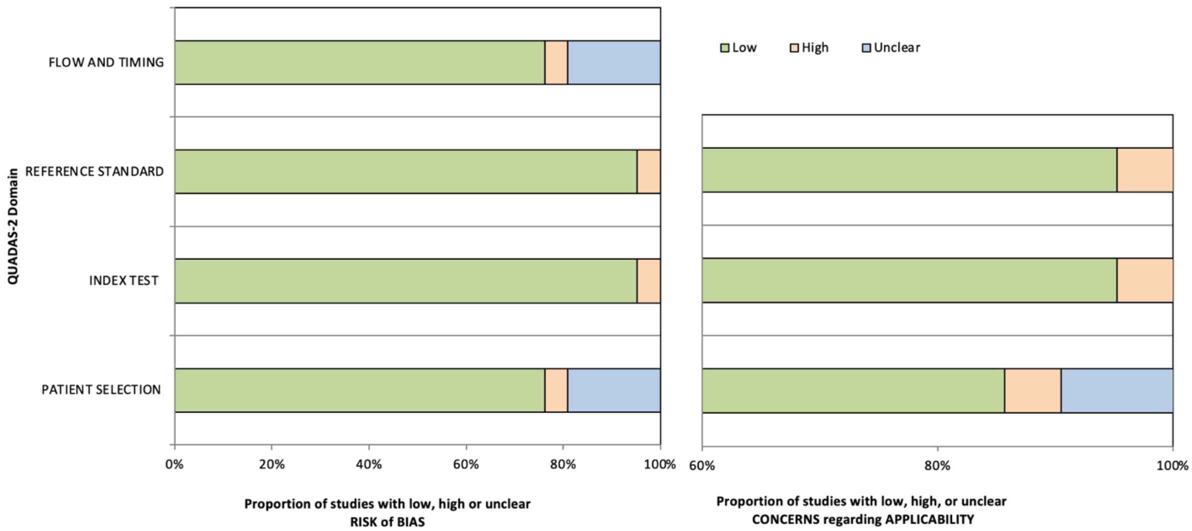
Appendix E. Simplified data extraction table

Study	Year	# Images used	# Images Labeled Normal	# Images Labeled non-normal	Accuracy	Sensitivity	Sensitivity SD	Specificity	Specificity SD
Acharya et al.	2009	2520	1080	1440	0.9	0.977		1	
Cao et al.	2020	1355	433	922	0.9483	0.9544		0.9353	
Dong et al.	2017	7851	4671	2176	0.847				
Guo et al.	2015	445	199	246	0.909				
Huang et al.	2009	1000			0.914				
Li et al.	2009	5820			0.95				
Li et al.	2010	5850			0.95				
Li et al.	2018	8030			0.972				
Lin et al.	2019	350	243	107	0.874	0.897		0.864	
Lin et al.	2020	2005	731	1274	0.81	0.79	0.02	0.82	0.04
		2005	731	1274	0.79	0.78	0.03	0.81	0.03
Liu et al.	2017	886	476	410	0.9707	0.9683	0.0002	0.9728	0.0001
Pratap & Kokil	2019	800	200	600	1				
Ran et al.	2018	5408	1948	3460	0.9704	0.9726		0.9692	

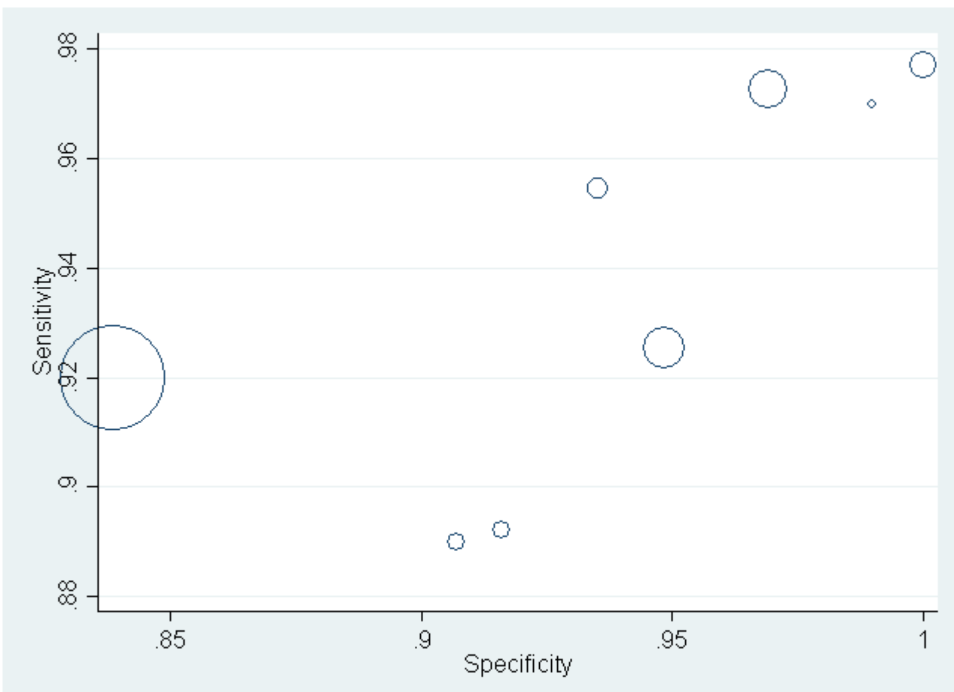
S V & R	2018	228	100	128	0.9696	0.97		0.99	
Wu et al.	2019	37638	4508	33130	0.8879	0.92	95% CI: 0.8733-0.9536	0.8385	95% CI: 0.7637-0.8971
Xu et al.	2021	8030	2212	5818	86.24	0.9010		0.8495	
Yang et al.	2016	1239	767	472	0.905	0.892		0.916	
		1239	767	472	0.899	0.89		0.907	
Zhang et al.	2017	5620	3269	2351	0.9352	0.9253		0.9484	
Zheng et al.	2014	460	158	302	0.9522				
Zhou et al.	2020	1355	433	922	0.94				

Appendix F. Results of the risk of bias assessment for studies included in the systematic review using QUADAS-2 Tool

	Domain 1		**Domain 2**		**Domain 3**		**Domain 4**
Study	Risk of Bias	Concerns about applicability	Risk of Bias	Concerns about applicability	Risk of Bias	Concerns about applicability	Risk of Bias
Acharya et al. (2009)	unclear	low	low	low	low	low	unclear
Cao et al. (2020)	low	low	low	low	low	low	low
Dong et al. (2017)	low	low	low	low	low	low	low
Guo et al. (2015)	unclear	low	low	low	low	low	unclear
Huang et al. (2009)	low	low	low	low	low	low	low
Li et al. (2009)	low	low	low	low	low	low	low
Li et al. (2010)	low	low	low	low	low	low	low
Li et al. (2018)	low	low	low	low	low	low	low
Lin et al. (2019)	low	low	low	low	low	low	low
Lin et al. (2020)	low	low	low	low	low	low	low
Liu et al. (2017)	unclear	unclear	low	low	low	low	unclear
Pratap & Kokil (2019)	unclear	unclear	low	low	low	low	unclear
Ran et al. (2018)	low	low	low	low	low	low	low
S V & R (2018)	low	low	low	low	low	low	low
Shimizu et al. (2021)	High	High	High	High	High	High	high
Wu et al. (2019)	low	low	low	low	low	low	low
Xu et al. (2021)	Low	Low	Low	Low	Low	Low	low
Yang et al. (2016)	low	low	low	low	low	low	low
Zhang et al. (2017)	low	low	low	low	low	low	low
Zheng et al. (2014)	low	low	low	low	low	low	low
Zhou et al. (2020)	low	low	low	low	low	low	low



Appendix G. SROC Curve



Appendix H. Hierarchical logistic regression results

Hierarchical Logistic Regression Results for Cataracts in Adults

Log likelihood = -104.8943 Number of ML classifiers = 9

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Bivariate						
E(logitSe)	2.908078	.7267127			1.483747	4.332409
E(logitSp)	3.183793	.3507146			2.496405	3.871181
Var(logitSe)	4.443394	2.541721			1.448133	13.63393
Var(logitSp)	1.059711	.5134012			.4100188	2.73887
Corr(logits)	.0321894	.3461407			-.5695972	.6115086
HSROC						
Lambda	6.588162	.8886644			4.846412	8.329912
Theta	-1.261845	.6828454			-2.600197	.0765079
beta	-.716711	.3742015	-1.92	0.055	-1.450132	.0167104
s2alpha	4.479614	2.383515			1.578821	12.71008
s2theta	1.050054	.5113627			.4042848	2.727318
Summary pt.						
Se	.9482443	.0356649			.8151379	.9870344
Sp	.9602198	.0133965			.9238894	.9795914
DOR	442.248	361.246			89.2011	2192.611
LR+	23.83709	8.105484			12.24091	46.4187

```
     LR- |   .0538998   .0371738                   .0139485   .2082796

   1/LR- |   18.55293   12.79565                   4.801237   71.69223

-------------------------------------------------------------------

Covariance between estimates of E(logitSe) & E(logitSp)    .008058
```

Hierarchical Logistic Regression Results for Pediatric Cataracts

```
Meta-analysis of diagnostic accuracy

Log likelihood   = -41.703961                 Number of ML Classifiers =       4

---------------------------------------------------------------------------

             |      Coef.          Std. Err.      z    P>|z|     [95% Conf. Interval]

-------------+-------------------------------------------------------------

Bivariate    |

  E(logitSe) |   2.008967   .6020843                   .8289034     3.18903

  E(logitSp) |   2.105182   .3449474                   1.429097    2.781266

Var(logitSe) |   1.407823    1.03089                   .3351572    5.913536

Var(logitSp) |   .4422734    .356941                   .0909328    2.151103

Corr(logits) |   .5943803   .3564424                  -.3763867     .943022

-------------+-------------------------------------------------------------

HSROC        |

      Lambda |   4.315963   .8600888                    2.63022    6.001706

       Theta |  -.6539444    .525611                  -1.684123    .3762343

        beta |  -.5789356   .4568128    -1.27   0.205    -1.474272    .3164009

     s2alpha |   2.516177   1.870604                   .5860463    10.80315

     s2theta |   .1600326   .1334146                    .031231     .820032

-------------+-------------------------------------------------------------

Summary pt.  |

          Se |   .8817353   .0627842                    .696123    .9604194
```

Sp |	.8914058	.0333914	.8067606	.9416551
DOR |	61.20009	51.78098	11.65618	321.3275
LR+ |	8.119543	2.862963	4.068147	16.20565
LR- |	.1326721	.0733519	.0448917	.3920966
1/LR- |	7.537381	4.167278	2.550392	22.27583

Covariance between estimates of E(logitSe) & E(logitSp)   .1171897

Appendix I: CHEERS Checklist

Topic	No.	Item	Location where item is reported
Title			
	1	Identify the study as an economic evaluation and specify the interventions being compared.	82
Abstract			
	2	Provide a structured summary that highlights context, key methods, results, and alternative analyses.	n/a
Introduction			
Background and objectives	3	Give the context for the study, the study question, and its practical relevance for decision making in policy or practice.	82
Methods			
Health economic analysis plan	4	Indicate whether a health economic analysis plan was developed and where available.	85
Study population	5	Describe characteristics of the study population (such as age range, demographics, socioeconomic, or clinical characteristics).	85

Topic	No.	Item	Location where item is reported
Setting and location	6	Provide relevant contextual information that may influence findings.	85
Comparators	7	Describe the interventions or strategies being compared and why chosen.	85
Perspective	8	State the perspective(s) adopted by the study and why chosen.	85
Time horizon	9	State the time horizon for the study and why appropriate.	85
Discount rate	10	Report the discount rate(s) and reason chosen.	84
Selection of outcomes	11	Describe what outcomes were used as the measure(s) of benefit(s) and harm(s).	85
Measurement of outcomes	12	Describe how outcomes used to capture benefit(s) and harm(s) were measured.	85
Valuation of outcomes	13	Describe the population and methods used to measure and value outcomes.	85
Measurement and valuation of resources and costs	14	Describe how costs were valued.	86
Currency, price date, and conversion	15	Report the dates of the estimated resource quantities and unit costs, plus the currency and year of conversion.	86
Rationale and description of model	16	If modelling is used, describe in detail and why used. Report if the model is publicly available and where it can be accessed.	87
Analytics and assumptions	17	Describe any methods for analysing or statistically transforming data, any extrapolation methods, and approaches for validating any model used.	85-90
Characterising heterogeneity	18	Describe any methods used for estimating how the results of the study vary for subgroups.	91
Characterising distributional effects	19	Describe how impacts are distributed across different individuals or adjustments made to reflect priority populations.	91

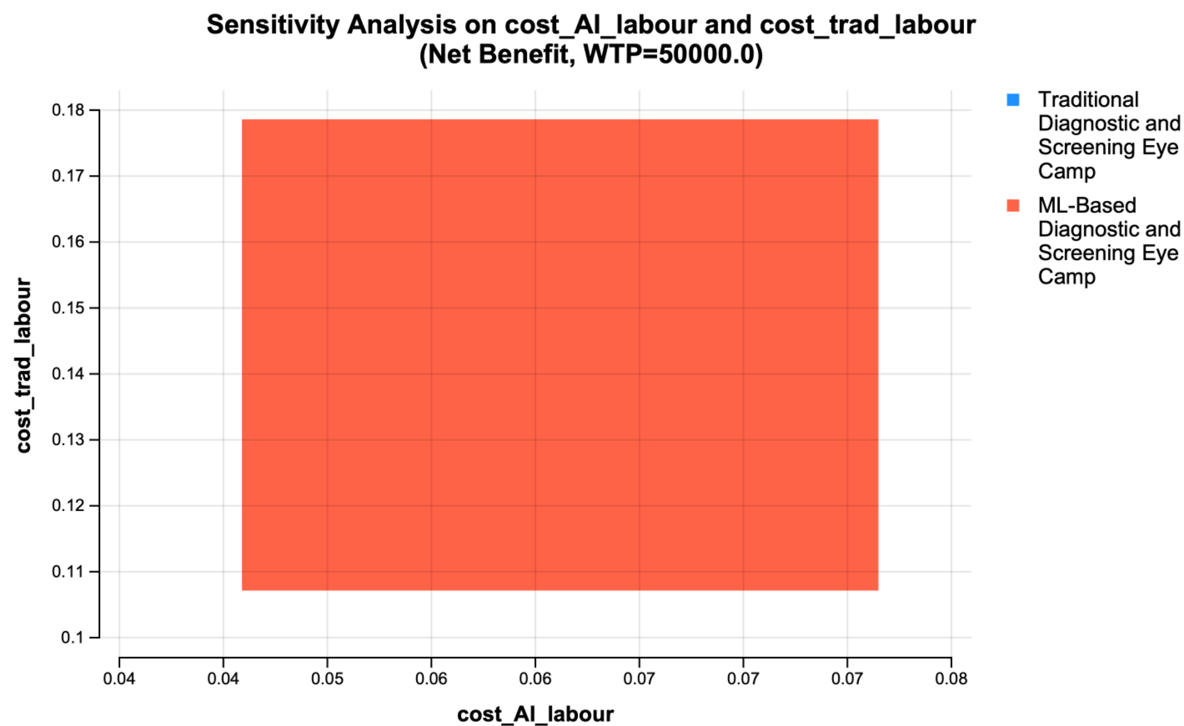
Topic	No.	Item	Location where item is reported
Characterising uncertainty	20	Describe methods to characterise any sources of uncertainty in the analysis.	91
Approach to engagement with patients and others affected by the study	21	Describe any approaches to engage patients or service recipients, the general public, communities, or stakeholders (such as clinicians or payers) in the design of the study.	90
Results			
Study parameters	22	Report all analytic inputs (such as values, ranges, references) including uncertainty or distributional assumptions.	91
Summary of main results	23	Report the mean values for the main categories of costs and outcomes of interest and summarise them in the most appropriate overall measure.	91
Effect of uncertainty	24	Describe how uncertainty about analytic judgments, inputs, or projections affect findings. Report the effect of choice of discount rate and time horizon, if applicable.	92
Effect of engagement with patients and others affected by the study	25	Report on any difference patient/service recipient, general public, community, or stakeholder involvement made to the approach or findings of the study	92
Discussion			
Study findings, limitations, generalisability, and current knowledge	26	Report key findings, limitations, ethical or equity considerations not captured, and how these could affect patients, policy, or practice.	96
Other relevant information			
Source of funding	27	Describe how the study was funded and any role of the funder in the identification, design, conduct, and reporting of the analysis	n/a

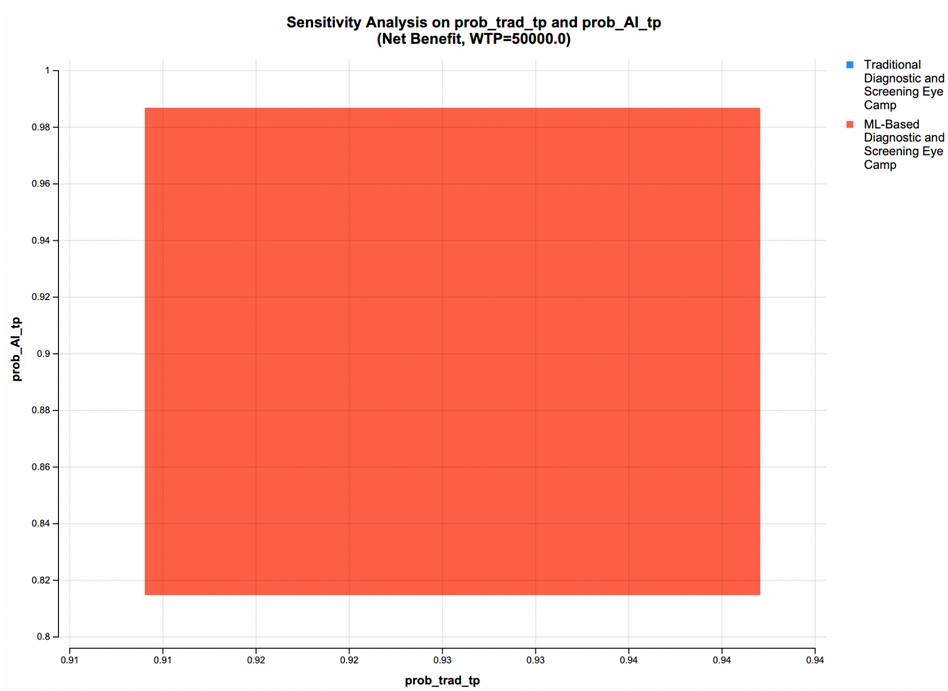
Topic	No.	Item	Location where item is reported
Conflicts of interest	28	Report authors conflicts of interest according to journal or International Committee of Medical Journal Editors requirements.	n/a

Appendix J: Base model variable inputs and definitions

Name	Root Definition
cost_AI_cap	0.03272
cost_AI_labour	0.06116
cost_AI_total	cost_AI_cap+cost_log+cost_AI_...
cost_log	0.09466
cost_trad_cap	0.03272
cost_trad_labour	0.14295
cost_trad_total	cost_trad_cap+cost_log+cost_tr...
eff_AI_fn	-prob_cat_pos*prob_AI_fn
eff_AI_fp	-prob_cat_neg*prob_AI_fp
eff_AI_tn	prob_cat_neg*prob_AI_tn
eff_AI_tp	prob_cat_pos*prob_AI_tp
eff_trad_fn	-prob_cat_pos*prob_trad_fn
eff_trad_fp	-prob_cat_neg*prob_trad_fp
eff_trad_tn	prob_cat_neg*prob_trad_tn
eff_trad_tp	prob_cat_pos*prob_trad_tp
prob_AI_fn	1-prob_AI_tp
prob_AI_fp	1-prob_AI_tn
prob_AI_tn	0.96
prob_AI_tp	0.948
prob_cat_neg	1-prob_cat_pos
prob_cat_pos	0.085
prob_trad_fn	1-prob_trad_fp
prob_trad_fp	1-prob_trad_tn
prob_trad_tn	0.838
prob_trad_tp	0.932

Appendix K: Two-way sensitivity analysis





Appendix L: Multi-way sensitivity analysis

Best case scenario:

Name	Root Definition
cost_AI_cap	667.89/22805
cost_AI_labour	624.27/22805
cost_AI_total	cost_AI_cap+cost_log+cost_AI_...
cost_log	2576.61/22805
cost_trad_cap	1113.14/22805
cost_trad_labour	4863.78/22805
cost_trad_total	cost_trad_cap+cost_log+cost_tr...
eff_AI_fn	-prob_cat_pos*prob_AI_fn
eff_AI_fp	-prob_cat_neg*prob_AI_fp
eff_AI_tn	prob_cat_neg*prob_AI_tn
eff_AI_tp	prob_cat_pos*prob_AI_tp
eff_trad_fn	-prob_cat_pos*prob_trad_fn
eff_trad_fp	-prob_cat_neg*prob_trad_fp
eff_trad_tn	prob_cat_neg*prob_trad_tn
eff_trad_tp	prob_cat_pos*prob_trad_tp
prob_AI_fn	1-prob_AI_tp
prob_AI_fp	1-prob_AI_tn
prob_AI_tn	0.98
prob_AI_tp	0.987
prob_cat_neg	1-prob_cat_pos
prob_cat_pos	0.085
prob_trad_fn	1-prob_trad_fp
prob_trad_fp	1-prob_trad_tn
prob_trad_tn	0.83
prob_trad_tp	0.909

Worst case scenario:

Name	Root Definition
cost_AI_cap	1113.14/22805
cost_AI_labour	1040.45/22805
cost_AI_total	cost_AI_cap+cost_log+cost_AI_...
cost_log	2576.61/22805
cost_trad_cap	667.89/22805
cost_trad_labour	2918.27/22805
cost_trad_total	cost_trad_cap+cost_log+cost_tr...
eff_AI_fn	-prob_cat_pos*prob_AI_fn
eff_AI_fp	-prob_cat_neg*prob_AI_fp
eff_AI_tn	prob_cat_neg*prob_AI_tn
eff_AI_tp	prob_cat_pos*prob_AI_tp
eff_trad_fn	-prob_cat_pos*prob_trad_fn
eff_trad_fp	-prob_cat_neg*prob_trad_fp
eff_trad_tn	prob_cat_neg*prob_trad_tn
eff_trad_tp	prob_cat_pos*prob_trad_tp
prob_AI_fn	1-prob_AI_tp
prob_AI_fp	1-prob_AI_tn
prob_AI_tn	0.815
prob_AI_tp	0.924
prob_cat_neg	1-prob_cat_pos
prob_cat_pos	0.085
prob_trad_fn	1-prob_trad_fp
prob_trad_fp	1-prob_trad_tn
prob_trad_tn	0.858
prob_trad_tp	0.942

CURRICULUM VITAE

Name:	Ronald Cheung
Post-secondary Education and Degrees:	Western University London, Ontario, Canada 2017-2021 BMSc
	Western University London, Ontario, Canada 2021-2022 MSc
Honours and Awards:	Lucille and Norton Wolf Trainee Publication Award London Health Research Day May 2022
	Proteus Innovation Competition Winner WORLDiscoveries, Western University April 2022
	Ontario Graduate Scholarship 2021-2022
	Western Graduate Research Scholarship 2021-2022
Related Work Experiences:	Graduate Teaching Assistant Western University 2022
	Research Assistant Princess Margaret Cancer Centre, Toronto Radiation Medicine Program 2021-Present
	Research Assistant St. Joseph's Health Care London Ivey Eye Institute 2019-Present

Publications:

Cheung R, Ito E, Lopez M, Rubinstein E, Keller H, Cheung F, Liu Z, Liu F-F, Wong P. Evaluating the Short-Term Environmental and Clinical Effects of a Radiation Oncology Department's Response to the COVID-19 Pandemic (STEER COVID-19). *International Journal of Radiation Oncology, Biology, Physics*. 2022. (*in press*)

Zhao B, **Cheung R**, Malvankar-Mehta MS. Risk of Parkinson's Disease in Glaucoma Patients: A Systematic Review and Meta-Analysis. *Current Medical Research and Opinion.* 2022. https://doi.org/10.1080/03007995.2022.2070377

Cheung R, Chun J, Sheidow T, Motolko M, Malvankar-Mehta MS. Diagnostic Accuracy of Current Machine Learning Classifiers for Age-Related Macular Degeneration: A Systematic Review and Meta-Analysis. *Eye (London, England)*. 2021. https://doi.org/10.1038/s41433-021-01540-y

Cheung R, Yu B, Iordanous Y, Malvankar-Mehta MS. The Prevalence of Occupational Burnout Among Ophthalmologists: A Systematic Review and Meta-Analysis. *Psychological Reports*. August 2020. doi:10.1177/0033294120954135

Yu B, **Cheung R**, Hutnik C, Malvankar-Mehta MS. The Prevalence of Obstructive Sleep Apnea in Glaucoma Patients: A Systematic Review and Meta-Analysis. *Journal of Current Glaucoma Practice.* 2021. doi:10.5005/jp-journals-10078-1301

Li W, **Cheung R,** Malvankar-Mehta MS. Comparing the Diagnostic Accuracy of Telemedicine Utilization Versus In-Person Clinical Examination for Retinopathy of Prematurity in Premature Infants: A Systematic Review. *Journal of American Association for Pediatric Ophthalmology and Strabismus*. 2021. https://doi.org/10.1016/j.jaapos.2021.12.006

Conferences:

London Health Research Day Conference **May 2022**
London, Ontario

- "Implementation of a machine learning-based cataract screening program in rural Nepal"

London Ophthalmology Research Conference **Nov 2021**
London, ON

- "Prediction Accuracy of Current Intraocular Lens Power Calculations Based on Artificial Intelligence Methods: A Systematic Review and Meta-Analysis."

University Health Network, Radiation Medicine Program Research Symposium **Aug 2021**
Toronto, Ontario

- "Evaluating the Short-Term Environmental and Clinical Effects of a Radiation Oncology Department's Response to the COVID-19 Pandemic (STEER COVID-19)"
- Awarded Best Presentation

Princess Margaret Cancer Education Summer Series **Aug 2021**
Toronto, Ontario
- "Evaluating the Short-Term Environmental and Clinical Effects of a Radiation Oncology Department's Response to the COVID-19 Pandemic (STEER COVID-19)"
- Awarded Best Presentation

World Glaucoma Congress **Jun 2021**
Tokyo, Japan
- "Diagnostic Accuracy of Current Machine Learning Classifiers for Cataracts: A Systematic Review and Meta-Analysis"

London Health Research Day Conference **May 2021**
London, Ontario
- "Diagnostic Accuracy of Machine Learning Classifiers for Cataracts: A Systematic Review and Meta-Analysis"

Canadian Ophthalmological Society Conference **Aug 2019**
Vancouver, British Columbia
- "The Prevalence of Obstructive Sleep Apnea in Glaucoma Patients: A Systematic Review and Meta-Analysis"