APPLICATION OF MACHINE LEARNING FOR GLAUCOMA SURGICAL OUTCOMES USING EHR DATA

By

Wei-Chun Lin

A DISSERTATION

Presented to the Department of Medical Informatics and Clinical Epidemiology and the Oregon Health & Science University
School of Medicine
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
in
Health and Clinical Informatics Research

AUGUST 2022

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the PhD dissertation of

Wei-Chun Lin

has been approved.

Michelle R. Hribar, PhD, MS – Mentor/Advisor

Xubo Song, PhD, MS – Chair

Michael F. Chiang, MD, MS – Member

Aiyin Chen, MD – Member

Nicole G. Weiskopf, PhD – Member

Table of Contents

Dedication

To my beloved parents, Song-Sun and Hsiou-Hsia

and my lovely wife, Yi-Jou

Acknowledgements

I could not have achieved as much as I have without the support of a lot of people. I would like to express my gratitude to my advisers, Drs. Michelle Hribar and Michael Chiang, for their guidance throughout my years of study at DMICE OHSU in many aspects (both research and life).

I want to give my sincere thanks to my committee members, Drs. Xubo Sung, Aiyin Chen, Nicole Weiskopf, for their insightful suggestions and encouragement in my research. I always feel so refreshing after the DAC meeting discussion with all of you.

I would also like to thank the members of the Casey Eye Institute Ophthalmic Informatics Lab – Jimmy Chen, Ryan Swan, Adam Rule, Isaac Goldstein, and Aaron Coyner – for their valuable inputs and help for my projects.

Of course, a huge thank you goes to National Library of Medicine for providing me the training grant to my research work.

Thank you for the wonderful Department of Medical Informatics and Clinical Epidemiology faculty and staff, Exam cohort, and attendees of RAD for constant support!

Finally, I am extremely grateful to all the support from my family, especially my parents and my wife, during my pursuit of my doctorate degree. Without their encouragement, it would be hard for me to complete my study.

Chapter 1: Introduction

Glaucoma is a group of chronic and progressive eye diseases caused by damage to the optic nerve, which is usually related to increased intraocular pressure. In 2020, more than 80 million[1] people were diagnosed with glaucoma, and the number is projected to increase to 110 million by 2040 worldwide.[2] Glaucoma is currently the second leading cause of irreversible blindness worldwide and often results in long-term life quality impairment.[3] In the United States, more than 3 million people are affected by glaucoma and the average direct cost of glaucoma treatment is close to $2,500 per year for patients with advanced glaucoma.[4] As glaucoma progresses to blindness, the indirect and direct costs of the disease become even more expensive, often requiring skilled home nursing, adaptive devices, and resulting in the loss of social status and self-esteem.

Medication therapies are crucial for glaucoma treatment - the main goal is to slow disease progression and preserve the quality of life. Multiple types of eye drops are used to reduce intraocular pressure via either reducing aqueous fluid production or increasing drainage. However, medications' adverse effects and slow disease progression without clear symptoms affect patient compliance.[5, 6] Moreover, tracking prescribed medications for glaucoma patients is important for patient care and clinical research; yet, extracting accurate medication information from electronic health records (EHR) is not an easy task. The accuracy of the medication list is questionable and often accurate medication data is available only in free text notes.[7, 8] Thus, an automated medication extraction tool is needed for glaucoma clinical care and research.

Surgical intervention may be needed if the maximum dose of glaucoma medications can not halt disease progression. Trabeculectomy remains one of the most common surgical procedures for glaucoma, especially in the developing world. However, the long-term surgical failure rates of trabeculectomy have been reported as 22% to 40% in different studies.[9-12] Proper early postoperative management is crucial for long-term surgical outcomes and varies according to different complications. The most common causes of long-term surgical failures include elevated IOP and hypotony (low IOP), and the early postoperative treatment plan for each may be very different. Researchers have investigated the risk factors of surgical failures and used pre or postoperative IOP to predict long-term surgical failure. However, the fluctuation of IOP during the early postoperative period and the complexity of surgical recoveries make predicting long-term surgical outcomes more challenging. Therefore, a trustworthy quantitative model for identifying if a patient has a high risk of long-term surgical failure and the cause of this failure (low or high IOP) is needed.

Artificial intelligence (AI) can be a possible strategy to address these challenges. With widespread EHR adoption and a large volume of clinical data available, AI techniques have been broadly applied in the medical field, including ophthalmology. In ophthalmology, studies have demonstrated that AI applications can be productive in many fields, such as diagnosis improvement and disease screening using imaging data. However, there are still many existing challenges, especially in applying AI techniques with real EHR data. Moreover, there is lacking sufficient attention to surgical outcome predictions. To address these challenges, I have proposed the following specific aims:

Aim 1: Identify the possible issues of the secondary use of EHR data in AI applications in ophthalmology – Conduct a systematic literature review to investigate current AI applications using EHR data for prediction and management of ocular diseases. The objective of this literature review was to gain better understanding of the AI techniques used and their performance, explore the potential problems of secondary use of EHR data, and provide future directions to clinical practice and research. I drew upon PRISMA[13] (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2009 flow diagram to perform the literature review using an exhaustive search in the PubMed database.

Aim 2: Approach the challenges of secondary use of EHR data, especially exploring the accuracy of medication lists for glaucoma patients – Perform EHR data quality assessment and explore the accuracy of medication lists for glaucoma patients. Develop a name entity recognition model to extract current medication information and adherence from clinical progress notes. Also, demonstrate a prototype automatic tool to help with medication reconciliation for glaucoma patients.

Aim 3: Predict multiclass long-term surgical outcomes for patients who underwent trabeculectomy and identify possible risk factors – Develop a prediction model to classify which patients have a high risk of long-term surgical failure due to high or low IOP. Predict the specific cause for primary trabeculectomy that addresses the need for effective post-operative management. This work was accomplished by using a multimodal neural network with structured EHR data and free-text operation notes. The third aim also explored the possible risk factors related to specific surgical outcomes using SHAP model interpretation tool.

Overall, our goal is to explore AI applications in ophthalmology and help with clinical care for glaucoma patients. The first study described the AI applications for ocular diseases and explored the knowledge gaps and problems. The second aim shows the data quality issues of glaucoma patients, such as the inaccuracy of medication list. Also, an NLP-based model was developed to assist in medication reconciliation and medication adherence extraction. This tool may be useful for patients' safety of medication usage and compliance improvement. Finally, the objective of the third study is to assist in postoperative clinical management for glaucoma patients. Early postoperative care is crucial for the long-term surgical outcomes of primary trabeculectomy. A reliable long-term surgical outcome prediction model will be very helpful for arranging appropriate treatment plans and allocating medical resources. Several possible risk factors for surgical failures were identified in the prediction model. These findings can bring novel insights into clinical care and the direction of future clinical studies.

Chapter 2: Background

GLAUCOMA

Symptoms

Glaucoma is a group of eye diseases characterized by progressive degeneration of optic nerve and visual field loss that is irreversible.[14] Although the pathogenesis of glaucoma is not fully discovered, intraocular pressure (IOP) control is the most important factor in influencing the progression of glaucoma.[15, 16] The imbalance of secretion of aqueous humor and drainage may increase intraocular pressure and damage retinal ganglion cells. **Figure 1** shows the illustration of open-angle glaucoma and closed-angle glaucoma.[16] Typically, glaucoma symptoms start from a gradual loss of peripheral vision then following by progressive loss of central vision. Eventually, glaucoma can progress to complete blindness.

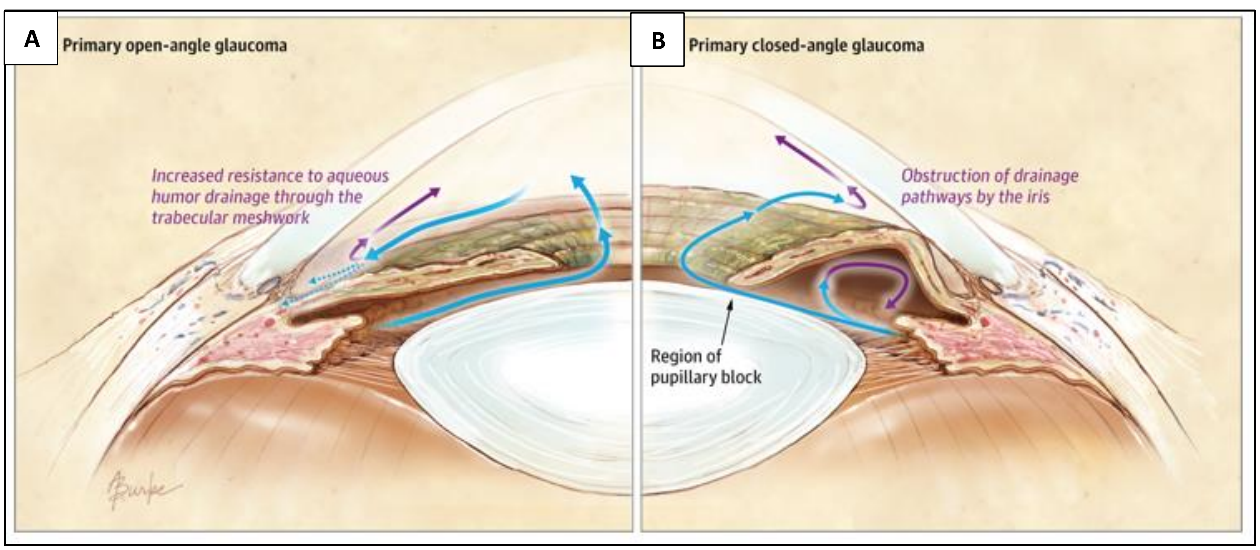

Figure 1: Illustration of glaucomatous eyes. Figure 1A shows the mechanism of open-angle glaucoma, which has increased resistance to aqueous outflow through the trabecular meshwork.

Figure 1B show the structure of primary closed-angle glaucoma with the obstructed drainage pathways. Figure adapted from "The Pathophysiology and Treatment of Glaucoma."[16]

Epidemiology

To date, glaucoma is the second leading cause of irreversible blindness worldwide.[3] Because the damage to the eye is slow and painless, only half of the patients are aware of the disease and irreversible nerve damage often happens long before the diagnosis.[17] The proportion of blindness attributable to glaucoma varies considerably from the lowest values in South Asia to a high prevalence in sub-Saharan Africa,[3] where open-angle glaucoma is the most common form (90%).[2] In 2020, there are 80 million people[1] worldwide with open-angle glaucoma, and the number was projected to increase to 111 million in 2040,[2] with an estimated 3 million afflicted in the United States alone. In the same year, the prevalence of glaucoma in the United States affects about 1.9% of individuals aged over 40.[18] The economic burden (medical costs, assistance programs in the US estimated at ~$3 billion/year) doubles when considering indirect costs associated with productivity loss, physical consequences (increase in hip fractures; increase in family care), and decreased quality of life. In the United States, the average direct cost of glaucoma treatment is close to $2,500 per year for patients with late-stage glaucoma.[4] Given that early detection of glaucoma, initiation and adherence to treatment strongly correlate with socioeconomic status.[19]

Diagnosis and Exams

The current clinical diagnosis of glaucoma is based on the measurements of intraocular pressure (IOP > 21 mm Hg), visual function (visual field test) and retinal structure. Eye exam devices such as stereoscopic ophthalmoscopy and optical coherence tomography (OCT) are used to examine the optic nerve head and degeneration of the retinal nerve fiber layer.[20, 21] Perimetry records the functional damage in vision by testing for arcuate scotoma (regional loss of peripheral vision), and tonometry is used to measure IOP.[20, 21] Patients at risk for the primary angle-closure type of the disease are additionally examined with gonioscopy, which bounces light into the eye using a specialized lens. In the glaucomatous eye, the optic nerve appeared to be cupped with thinning of the neuro-retinal rim neuroretina rim is narrower, and the center is cupped.

Visual Field

The measurement of visual field change shows the trend of disease progression. **Figure 2** shows an example of the glaucomatous optic nerve head and visual field test result. Summary statistic of visual field changes, such as Mean Deviation (MD) and Visual Field Index (VFI) is often used to perform the analyses. Rates of visual field loss are typically expressed as linear rates of change of dB per year in practice. However, the rates of visual field worsening are not necessarily constant over time in reality. For example, patients' compliance with treatment and treatment intensity can change the worsening rate. Visual filed progressive rates for glaucoma patients in treated clinical practice are varied among studies with a range from -0.05dB per year to -0.62dB per year.[22-24] On the other hand, in Heijl's study, they reported that the median rate of MD loss in untreated eyes of 118 glaucoma patients was -0.4 dB per year.[25] For VFI change, studies have reported average rates of VFI loss in glaucoma patients from -1.1% to -1.5% per year.[26, 27] Moreover, visual fields decline due to aging has estimated to be -0.06dB per year.[28] Previous studies suggest that most glaucoma patients do not progress very quickly, but a sizeable part of patients shows higher progress rates. In these cases, the mean rates of visual field loss are worse than average rates.[23-25] The reported proportion of glaucomatous eyes progressing at faster than -1.5 dB per year have varied from 3 to

17% in some previous studies.[23, 25, 26, 29, 30] Also, some studies found that 15-20% of glaucomatous eyes seem to progress at rates of VFI loss higher than 5% per year.[27, 31] The variety of rates of glaucoma progression makes it difficult for medication treatment and leading the need for surgery.

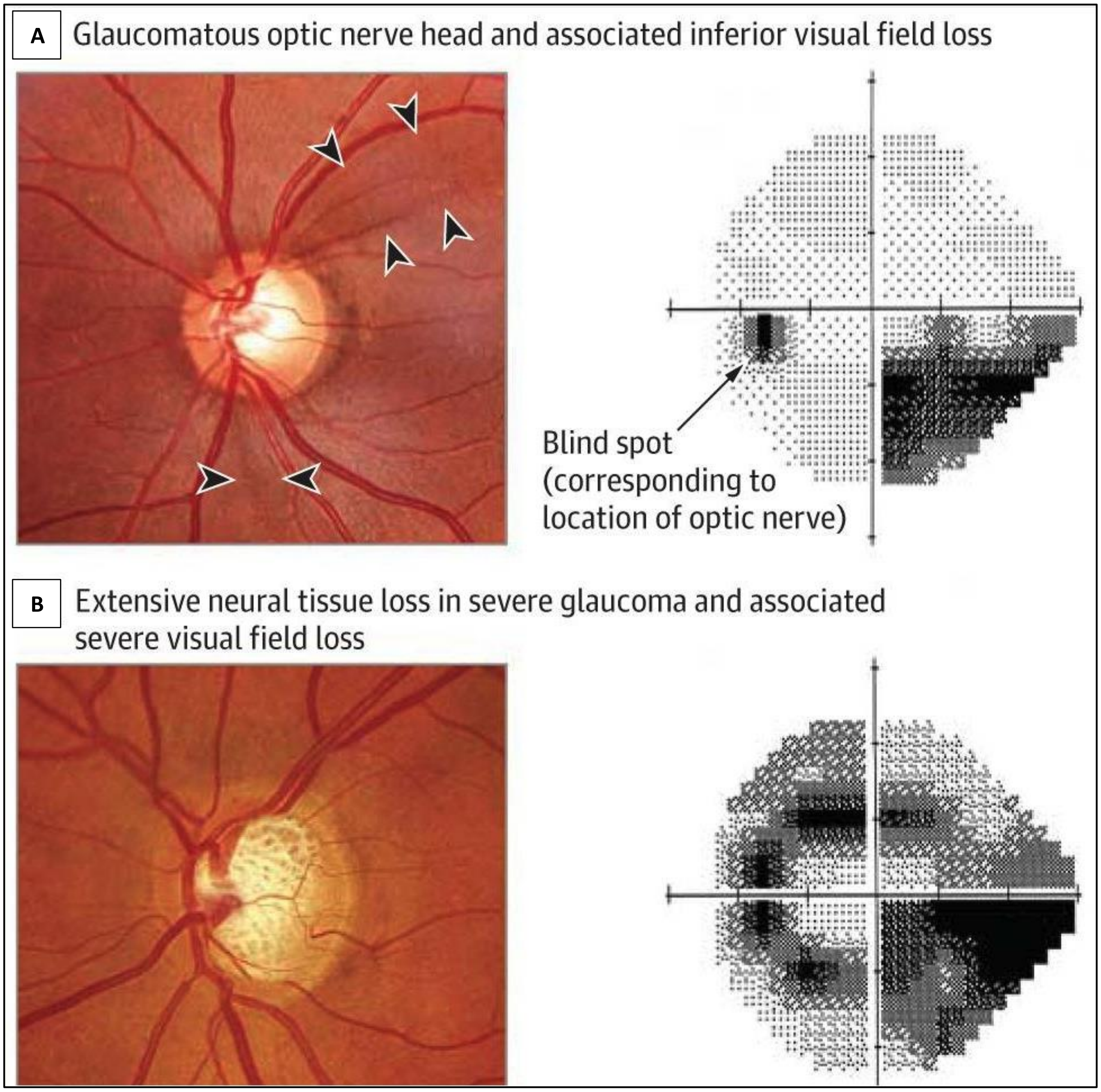

Figure 2: Glaucomatous and severe glaucomatous optic nerve heads and corresponding visual field test results. Figure A, glaucomatous optic nerve presents a thin neural retinal rim and enlarged optic cup. The visual field loss (inferior defect) corresponded to the superior neural losses.

Figure B, severe glaucomatous neural fiber loss and enlarged optic cup. The visual field shows defect in both the superior and inferior hemifield. Figure adapted from “The Pathophysiology and Treatment of Glaucoma.”[16]

Optical Coherence Tomography

Retinal nerve fiber layer (RNFL) measurements measured by Optical Coherence Tomography (OCT) have been the most commonly used metrics to measure structural change. OCT employs the interferometric technique with a low-coherent light source to construct the cross-sectional images. OCT is analogous to ultrasound except that it uses light as opposed to sound and does not need physical contact with the tissue. OCT can visually differentiate the tissue layers of the retina because of the differences in the tissue light scattering properties. Thus, OCT is capable of characterizing the morphological changes in RNFL prior to the potential of visual loss. The estimates of mean RNFL rates of loss in healthy eyes are reported in the range of -0.33μm/year (cross-sectional analysis) and -0.52 μm/year (longitudinal analysis).[32] The progression in glaucoma patients measured using spectral-domain OCT-based RNFL thinning has been found to range between -0.76 μm/year and -1.5 μm/year.[33-35] Therefore, OCT measured RNFL thickness as the structural changes can be another indicator for glaucoma progression.

Treatment of Glaucoma

The main goal of glaucoma treatment is to slow disease progression and preserve the quality of life. Multiple studies have shown that intraocular pressure reduction is the primary and the only proven way to treat the disease.[36-39] The treatment of glaucoma relies on medication eyedrops or surgeries, which either reduce aqueous fluid production or increase drainage. Physicians should use the fewest medications and minimize adverse effects to control patients' intraocular pressure. Several different groups of pressure-lowering medications, including eye drops and oral medication are available. The first line of medical therapy is prostaglandin analogues, which reduce IOP by reducing outflow resistance that increases aqueous humor flow.[40] However, prostaglandin analogues may cause some adverse effects, such as conjunctival hyperemia, loss of orbital fat, and periocular skin pigmentation. Therefore, second-line glaucoma medications are used when the adverse effects, contraindications of prostaglandin analogues are found, or IOP is inadequately controlled with the first line medication. Second-line agents include carbonic anhydrase inhibitors, beta-adrenergic blockers, alpha-adrenergic agonists, and pilocarpine.[41] Some glaucoma eye drops, such as beta-adrenergic blockers may cause significant systemic adverse effects that make medication control for glaucoma more challenging. Enhancing patients' compliance can increase the success of medication therapy.

Medication Treatment

Medication therapies are crucial for glaucoma treatment; however, extracting accurate medication information from the EHR remains a challenging task. A medication list is generated through clinical care; this discrete format data provides an opportunity for medication data extraction. Before using the medication list to track medicine usage for glaucoma patients, it is important to understand the accuracy and completeness of the information. Several issues of accuracy of medication information were reported, such as accuracy of medication list, medication discrepancies in an integrated EHR, and meaningful information of medication may be recorded in the clinical notes.[7, 8] Characterizing the accuracy of the medication list can help people understand the quality of automated data to inform the clinical care team and use it for medical research. In addition, prior studies show that physicians direct very little attention to EHR medication lists, and instead spend most time reviewing the impression and plan section.[42, 43] It seems reasonable to expect that medications recorded in narrative notes can be a reliable source of medication records. Describing the agreement between a patient's medication list and the clinical narrative progress note can be used to evaluate the accuracy of the medication list. However, manually reviewing narrative progress notes and medication lists is a time-consuming task.

As previously mentioned, enhancing patients' compliance can help with disease progression. However, chronic and initially asymptomatic diseases, such as glaucoma, are susceptible to medication non-adherence and affect the disease progression. Since the majority of patients with glaucoma are managed initially with medical therapy, sustained and consistent patient adherence is critical for preventing the progression of the disease.[44, 45] Poor compliance might decrease the treatment effect and lead to disease progression. Non-adherence in patients with glaucoma has been reported to vary from 24 to 59%.[46-48] Several barriers to glaucoma medication adherence were reported, such as difficulties with self-instillation of eye drops, lack of motivation, and forgetfulness.[5, 6 20, 21] Medication adherences can be measured by direct and indirect methods, but the cost of the direct method and accuracy of the indirect method need to be concerned.[49] On the other hand, physicians often record patients' compliance and adherence in the clinical notes. A previous study shows the extraction of patients' noncompliance from clinical notes can identify a larger proportion of patients who self-reported poor medication adherence compared to an automated EHR pull alone.[50] Therefore, a reliable method to get active medication information and adherence for glaucoma patients from clinical notes will be needed.

Surgical Treatment

Surgical intervention may be needed if the maximum dose of glaucoma medications cannot achieve adequate intraocular pressure reduction or halt disease progression. For patients with poorly adherent or rapid progression, surgery may be considered first-line therapy. General glaucoma surgeries include laser and incisional surgeries. The estimated number of incisional glaucoma surgeries is 274 per million people per year.[51] Laser trabeculoplasty uses short pulses of low-energy light to affect biological changes in the trabecular meshwork. These changes can improve drainage and lower intraocular pressure. Laser procedures are very safe and can be performed in the clinic office room. However, the effects of the procedure decrease gradually over time with an approximate 10% failure rate every year.[52, 53]

Trabeculectomy

Incisional glaucoma surgeries include drainage device implantation and trabeculectomy. Drainage device implantation is also called minimally invasive glaucoma surgery. The drainage device can reduce IOP immediately through a valving mechanism.[54] The most common postoperative complications include fibrosis and tenon capsule formation around the implant plate decreased drainage of the aqueous humor, leading to increased IOP and eventual device failure in some patients.[54] Therefore, many anti-inflammatory agents, such as steroids, anti-VEGF, and antifibrotic medications, have been used to prevent fibrosis with variable success.[55]

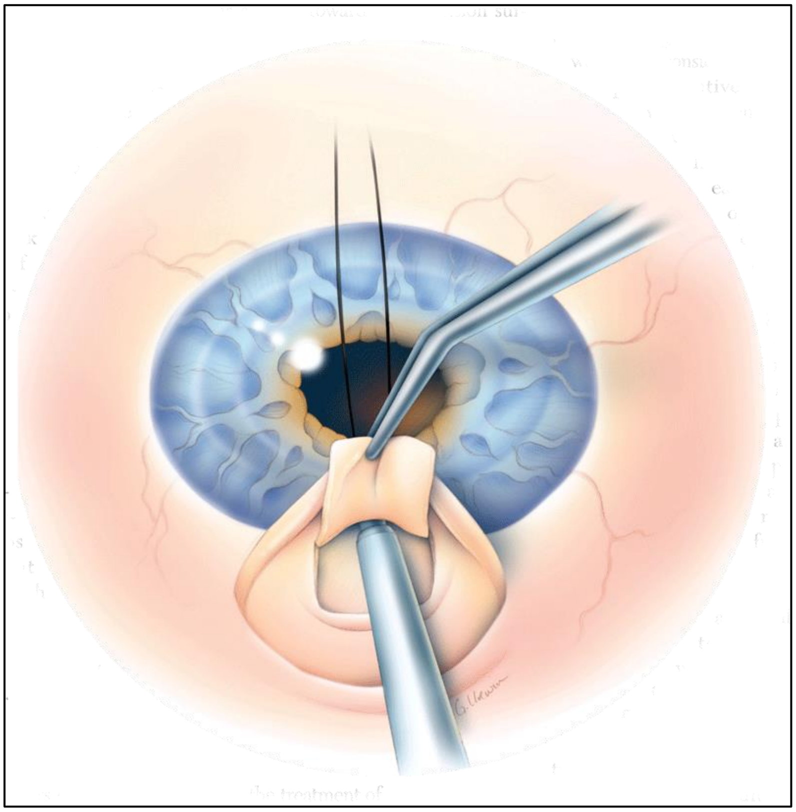

Figure 3. Illustration of trabeculectomy. Surgeons cut out the trabecular meshwork or the tissue anterior to the trabecular meshwork to provide a drainage way for the outflow of aqueous humor from the inside of the eye. Figure adapted from “Is It Time to Retire the Trabeculectomy?”[56]

Although the popularity of drainage device implantation has steadily increased, trabeculectomy is still one of the most common glaucoma incisional surgeries, especially in developing countries.[57] Trabeculectomy is a type of filtering surgery to improve the fluid drain out of the eye to decrease intraocular pressure. Surgeons will remove a small portion of the trabecular meshwork and or adjacent corneoscleral tissue to provide a drainage way for the outflow of aqueous humor from inside of the eye. **Figure 3** displayed an example of trabeculectomy.[56, 58] Releasable or adjustable sutures are commonly used to reduce the chance of postoperative complications (**Figure 4**).[59] In addition, anti-scarring agents are frequently used in surgical sites to mitigate fibrosis-proliferative response and increases the surgical success rate, but this may increase the rate of complications such as infection and hypotony.

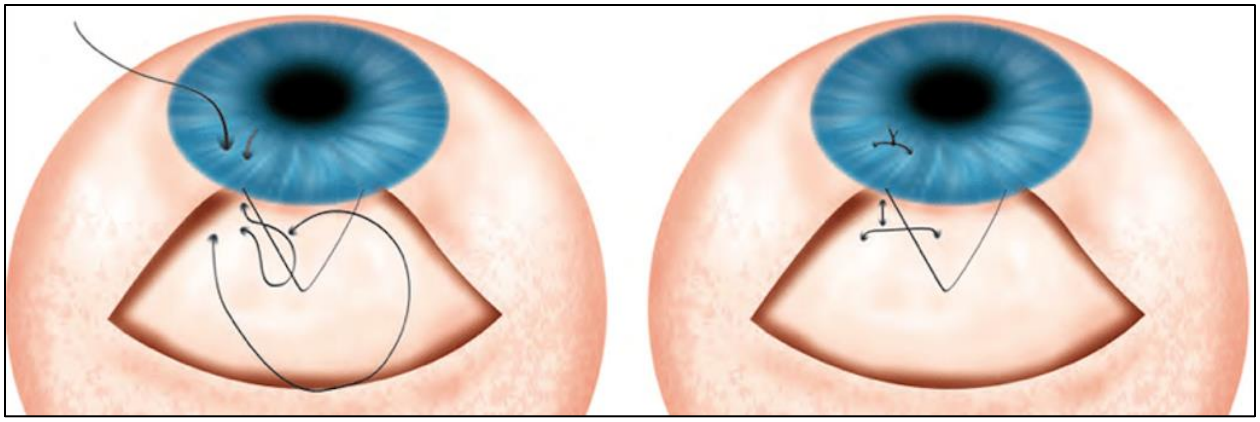

Figure 4. Example of releasable sutures. Figure adapted from "Releasable Sutures in Trabeculectomy."[59]

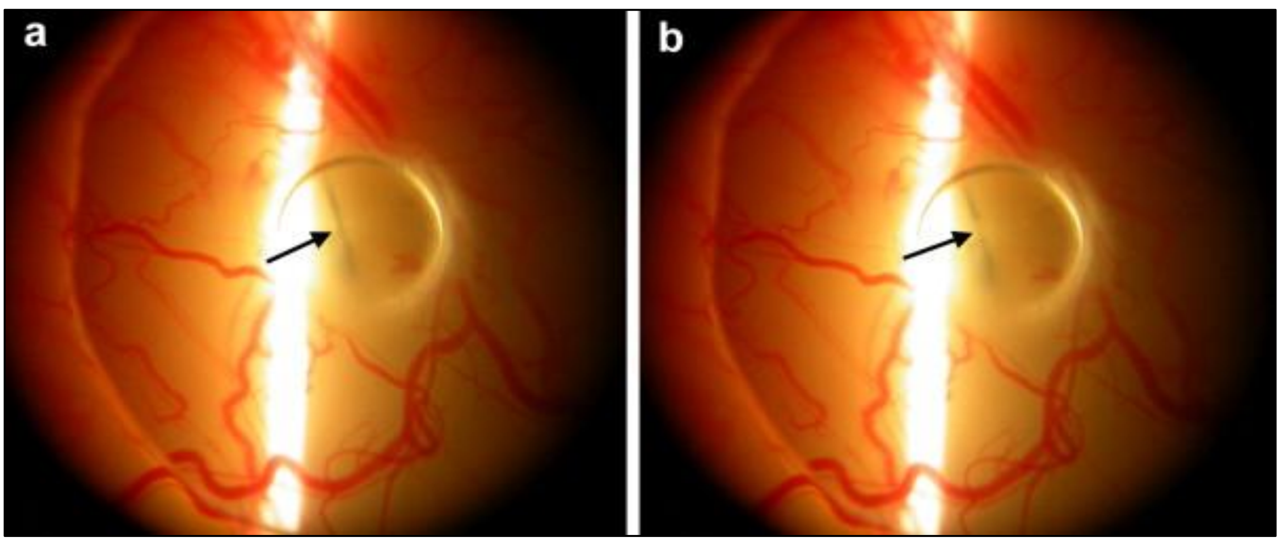

Figure 5. Laser suture lysis image. The suture was lysed by Argon laser in the black arrow. Figure adapted from "Argon laser suture lysis following glaucoma filtering surgery–A short introduction to the procedure."[60]

Surgical Failures of Trabeculectomy

However, the long-term surgical failure rates of trabeculectomy were around 30% reported in several studies.[9-11] In the Tube versus Trabeculectomy Study, the failure rate of trabeculectomy was 28.2% in the second year and 30.7% in the third year.[9] Also, Garris et al. reported the overall surgical failure rate at 2 years was 32% for trabeculectomy.[10] In a multicenter study by Kirwan et al., the surgical failure rate was from 22% to 35%, according to different criteria.[11] The definition of surgical failure may be variable among different studies. In general, surgical failure of trabeculectomy was defined as patients needing re-operation, loss of light perception vision, showing consistently elevated IOP (> 21 mmHg or less than 20% reduction below baseline) or hypotony (<=5 mmHg) after 3 months of primary surgery.[9] The success of trabeculectomy highly depends on post-operative management within the first 3 months following surgery.[61] After the trabeculectomy procedure, the patient is seen on the first day postoperatively. Topical corticosteroids are often used to alleviate inflammation or scarring tissue proliferation. Depending on the patient's condition and intraocular pressure control, the follow-up interval can vary. A common routine is scheduling a follow-up visit one week post-operatively and then every one to two weeks within the first two to three months post-operatively.[62]

Postoperative Management of Trabeculectomy

Proper IOP control in the early postoperative period is critical for long-term surgical outcomes and is affected by different surgical complications.[63] The most common early postoperative complications involved either elevated intraocular pressure or hypotony. Scarring is the most common and challenging complication related to post-operative elevated IOP.[64] Physicians may lower the intraocular pressure by releasing the scleral flap sutures using laser suture lysis or removing the releasable sutures altogether. Also, the post-operative antifibrotic agent 5-fluorouracil (5-FU) can be used to modulate wound healing and limit scar tissue formation.[65] On the other hand, excessively low IOP post-surgery is caused by conjunctival wound leaks and over-filtration.[61, 66] Physicians may postpone laser suture lysis or decrease the dose and frequency of topical steroid use to reduce the risk of hypotony. If the hypotony persists and leads to severe complications like choroidal, operations may be needed.

Several studies have investigated the possible risk factors for failure of trabeculectomy. For example, in Broadway et al.'s review study, they summarized the risk factors for failure of trabeculectomy, including previous ocular surgery, uveitic or neovascular glaucoma, black race, multiple topical glaucoma medications used for a long time, and young age.[67] The previous ocular surgeries include failed previous trabeculectomy, cataract extraction, and conjunctival incisional procedures, while the possible causes to explain this are the breakdown of the blood-aqueous barrier and anatomic disturbance that affect the wound healing.[67] Also, Landers et al. reported similar findings that the risk of trabeculectomy failure was younger or had uveitic glaucoma and patients with pseudoexfoliation or aphakia were more likely to progress to blindness.[68] These risk factors may be helpful with the surgical outcome prediction for trabeculectomy. In addition, other studies focused on the prediction power of intraocular pressure for surgical failure, including pre-operative IOP or early postoperative IOP. Some studies suggested that early postoperative IOP at 1 month had higher predictive power for long-term surgical outcomes[69, 70] but the predictive power of pre-operative IOP was controversial.[71]

In conclusion, despite the fact that several risk factors of trabeculectomy surgical failure were identified and early postoperative IOP seems to help with surgical outcome prediction, the surgical outcomes still remain hard to predict. The fluctuation of intraoperative pressure during the early postoperative period and the complexity of surgical recoveries make identifying which patient has a higher risk of long-term surgical failure more difficult. Classic statistical regression methods based on intraocular pressure alone or with limited features cannot provide sufficient information to predict long-term surgical outcomes accurately. Furthermore, since the treatment plans for elevated IOP and hypotony might be contrary, it is more clinically useful to predict a patient's surgical failure risk due to the specific cause. Therefore, there is a strong need for a reliable quantitative model for identifying a patient's risk of surgical failures due to different complications, which could aid the decision-making of post-operative management.

ARTIFICIAL INTELLIGENCE

The rapid adoption of electronic health records (EHRs) has generated huge volumes of clinical data that can help with secondary use in research. EHR data analysis can be used to support clinical decision-making, medical concept extraction, disease screening, diagnosis, and risk assessment. However, there are many different types of EHR data, such as patient demographic information, diagnosis codes, laboratory tests, medication prescriptions, imaging, free-text clinical documents, and billing codes. The complexity and heterogeneity of EHR data make it difficult to reuse. One of the promising strategies to handle those heterogeneous data is the artificial intelligence (AI) techniques. Artificial intelligence is a broad field where the computer simulates human intelligence that is programmed to mimic human action or thinking. Over the past decades, AI techniques have been applied within and outside of the scientific community. The primary characteristic of AI is its ability to rationalize and take actions that achieve specific goals with the best possibility.

Machine Learning

Machine learning (ML) is a subset of artificial intelligence where computer programs can automatically "learn" from the data without being assisted by humans. Machine learning algorithms use computational methods to learn patterns and information directly from large amounts of data without relying on a predetermined equation or being explicitly programmed. In other words, machine learning algorithms find natural patterns within data to get novel insights and make accurate predictions. Overall, the goal of machine learning generally is to understand the structure of the data and fit that data into models that can be understood and utilized by people. In ML, there are three main types include supervised learning, unsupervised learning, and reinforcement learning (**Figure 6**).[72, 73] In supervised learning, a model learns from "ground truth" data in a training data set that contains labeled output data and then can predict the output for new cases. The algorithm is typically a classifier with categorical output or a regression algorithm with continuous output. In unsupervised learning, the model learns from a training data set without labeled output and identifies underlying patterns or structures within its input data, which is mainly used for clustering problems.[74] Reinforcement learning maximizes the cumulative reward to improve the model performance by interacting with the environment. There are many machine learning algorithms applied in medicine. Some of the most popular algorithms include random forest,[75] logistic regression,[76] support vector machine,[77] gradient boosting,[78] least absolute shrinkage and selection operator (LASSO),[79] and AdaBoost[80] were introduced in the chapter 3. In addition, aside from classical machine learning algorithms, some other techniques were popularly used for clinical studies and applications, including deep learning and natural language processing (NLP). In the next sections, we will describe the basic introduction of deep learning and NLP.

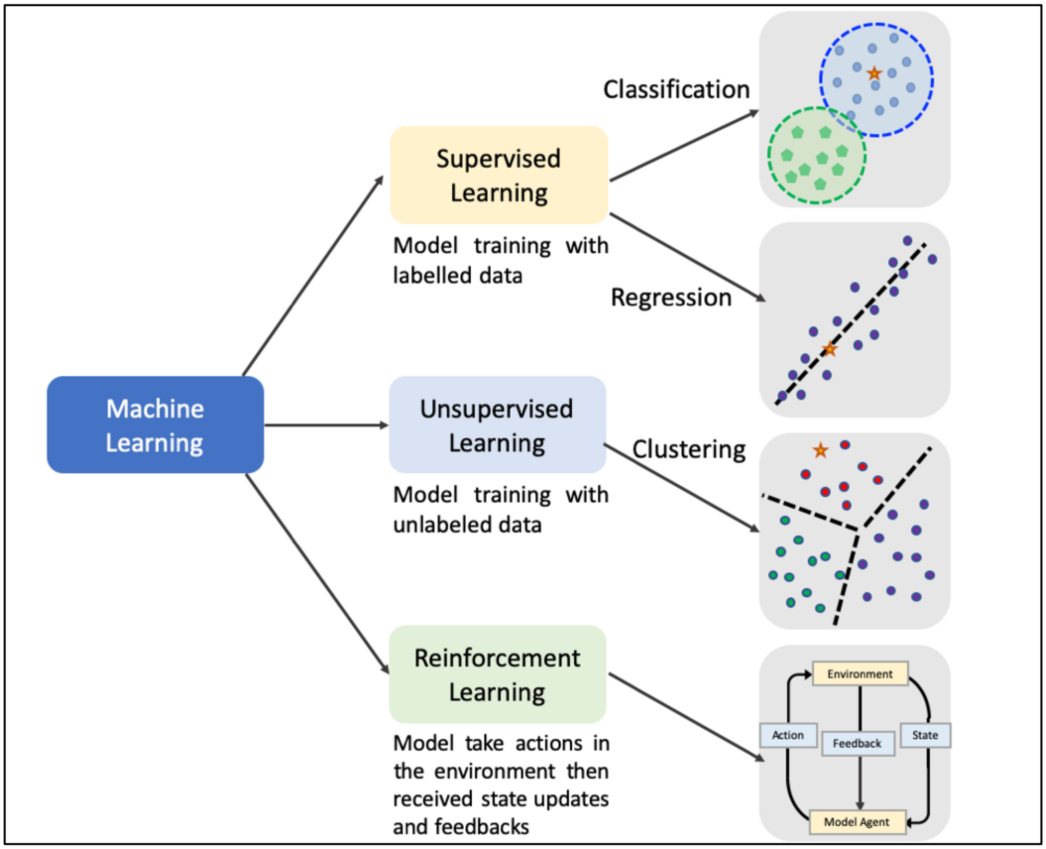

Figure 6. Illustration of three major types of machine learning. Supervised learning includes classification used to predict category and regression used to predict numerical value. The main difference between supervised and unsupervised learning based on whether labels are given.

Deep Learning

Artificial Neural Networks

Deep learning is a subset of machine learning techniques based on artificial neural networks (ANNs) that mimic human brain processing. As shown in **Figure 7**, multiple layers of computation are constructed in a deep learning model, and each layer is used to perform computations on data from the previous layer. The layers between the input layer and the output layer are called hidden layers. While the information may flow from the input to subsequent output layers (feedforward), information can also flow backward from hidden layers to input layers (backpropagation). The inputs and outputs of hidden layers are not reported; deep learning algorithms present only the final outcome of the output layer.[81] Deep learning does not always need to use structured features for input as machine learning does; hence, deep learning is useful for raw images because they do not have to be prefiltered as they do for machine learning algorithms. After processing raw input through multiple layers within deep neural networks, the algorithms find appropriate features for classifying output.

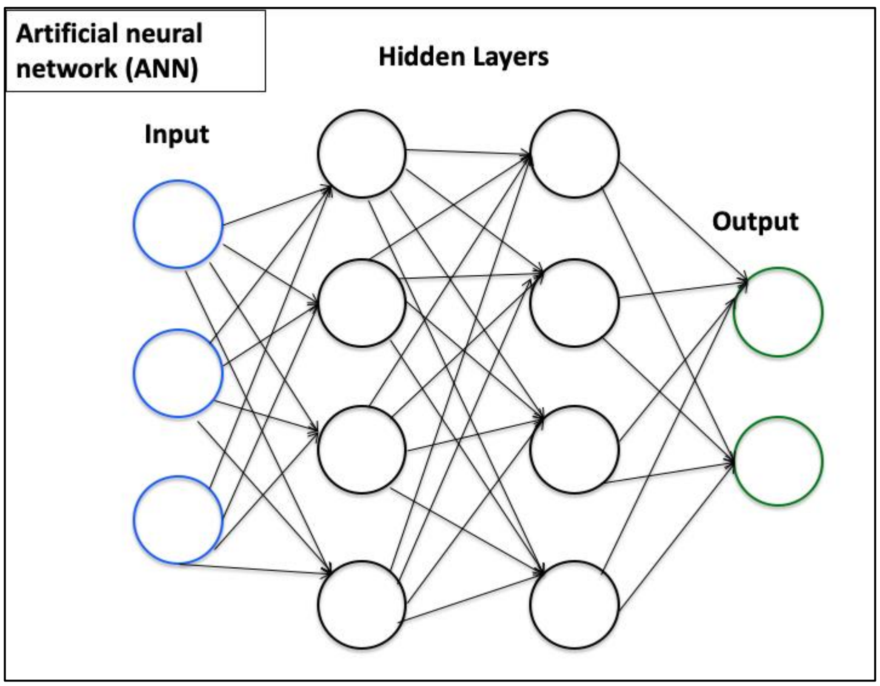

Figure 7. Illustration of artificial neural network (ANN). A classical ANN contains input layer, hidden layers, and output layer. A neural network can possess any number of hidden layers with one or more neurons. Each neuron is connected to each other and each layer could have different activation functions.

Recurrent Neural Network

A recurrent neural network (RNN) is a subtype of deep learning. In a traditional neural network, the model is able to learn non-linear relationships with large dimensions. However, a static neural network is not able to learn the temporal information, which is trained by itself. It is important for time-series data to learn the information from the previous prediction. That led to the development of RNN, which remembers the information from prior inputs throughout the time steps in the training process. More specifically, in an RNN model, each output of the hidden state from the previous time step is used as the input to the hidden state of the next time step. RNNs were used to predict disease progress and clinical event onset in several studies.[82-84] Also, the RNN model was used to build a visual field prediction algorithm with a series of visual field data.[84]

However, one of the most common problems of an RNN model was short-term memory. If the input sequence is long enough, an RNN model will be hard to carry information from earlier time steps - the later ones are much more important in this case. In detail, during the calculation of backpropagation, RNN models can easily suffer from the vanishing gradient problem. Gradients are the values that were used to update the neural network weights. The vanishing gradient problem might occur when the gradient shrinks as it backpropagates over time.[85] Therefore, if a gradient value becomes extremely small, it cannot help with the model learning. In RNN, the earlier layers easily get small gradient values and the model will forget early step information.

Long Short-Term Memory

To address this problem, two advanced RNN-based algorithms were created, including long short-term memory (LSTM)[86] and gated recurrent units (GRU).[87] **Figure 8** shows the basic structure of a LSTM cell. These algorithms contain internal mechanisms called gates to regulate the memory of information. The gates are used to learn which input sequence is more important to keep and skip the less important one. In this way, the model can pass relevant information over a long distance to make predictions. LSTM has popularly used in many fields such as speech recognition, sentiment analysis, text generation, time series analysis and text classification.

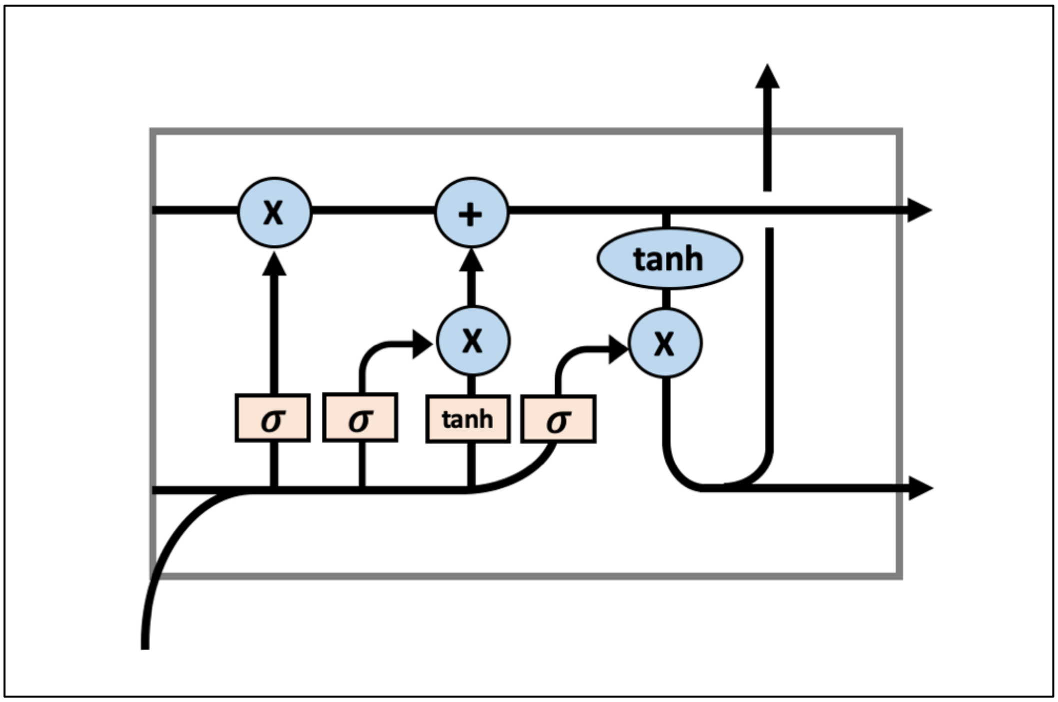

Figure 8. Structure of LSTM Cell. There are two major parts of an LSTM cell: (1) Cell state and (2) Cell gates. The cell state is the horizontal line shown on top of the figure, which passes the information between cells. The gates shown at the bottom of the figure, which regulated the information can be added or removed.

Transformer

A transformer model is a neural network with self-attention and positional embedding mechanisms, which is popularly used in many fields, such as NLP and computer vision. In the paper "Attention Is All You Need" Vaswani et al.[88] first introduced a novel architecture called Transformer. The transformer is a neural network with attention mechanisms and positional embeddings for transforming sequence to sequence. **Figure 9A** shows the basic structure of the transformer with encoder block and decoder block. Two key components were described in the encoder block: Multi-head attention and positional embeddings. Multi-head attention is the core mechanism of a transformer block, which contains n different attention layers to find the attention token/word from the sequence. Also, positional embeddings can enhance and maintain the position of each word along the sequence. Both of these mechanisms can provide information about the relationship between different words. Furthermore, a transformer model can train in parallel and manage the text data in a non-sequential way, which means sentences are processed as a whole rather than word by word. In this way, a transformer can void forgetting of past information over a long distance. Long-distance memory loss is the most critical issue for RNN-based neural networks, even for the LSTM and GRU models. For a classification problem, only encoder blocks were used to extract information.[89] The output of each encoder is passed to the next encoder, and the process goes on. Output from the final encoder layer can be passed to linear layers or concatenated with other input data (**Figure 9B**).

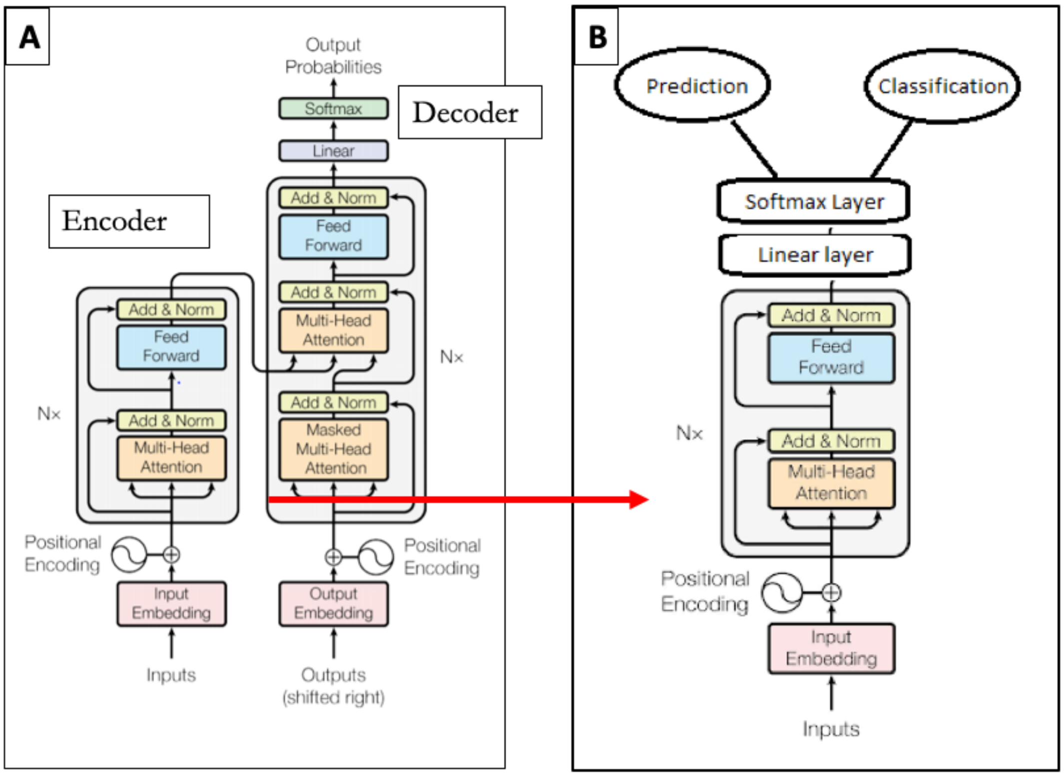

Figure 9. Architecture of the transformer. Figure 9A shows the structure of transformer including encoder block and decoder block with input embeddings and multi-head attention layers. Figure 9B demonstrates the encoder blocks connected to an output layer. Figure adapted from "Attention Is All You Need."[88]

Deep Learning Fusion Model

Clinical and EHR data are increasingly getting multimodal and heterogeneous, such as structured EHR data, image data, text data, time-series data, and audit logs. Integrating multimodal data may be able to capture the underlying complex relationship among different data sources of data to improve the model performance. Deep learning-based data fusion strategies are a popular approach for handling these heterogeneous data. Deep neural networks are flexible and capable of combining different functional blocks in a single model. Generally, the fusion strategies can be categorized as early fusion, intermediate fusion, and late fusion (**Figure 10**).[90] Input data are concatenated directly in early fusion; thus, the resulting vector is treated like unimodal input.[90] In an intermediate fusion, the input features can be processed through some neural networks (LSTM, transformer block, etc.) and then the learned features are concatenated with other input features.[90] Lastly, in a late fusion, the original input features or learned features are not concatenated, but combined decisions from each sub-model (with unimodal input features).[90] The decisions of each sub-model can be the predicted label or probability. The late fusion strategy is similar to ensemble learning or voting averaging methods in classical machine learning.

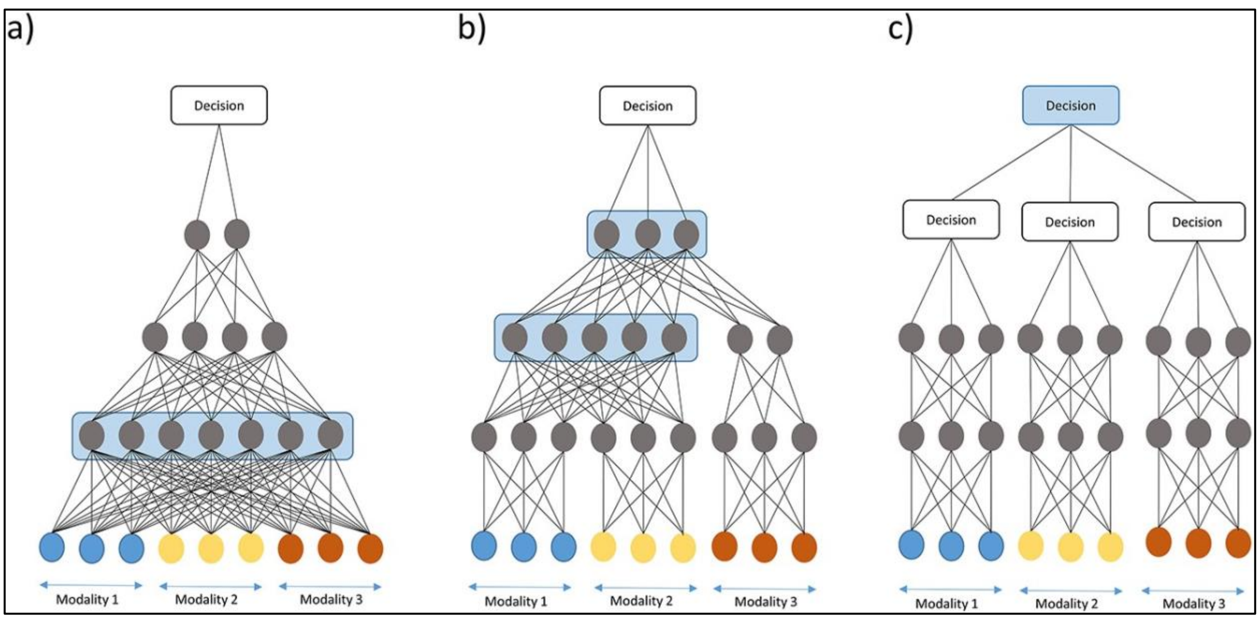

Figure 10. Deep learning-based fusion strategies. The blue layers contain the inputs from multiple modalities and learn the information simultaneously. (a) Early fusion strategies (b) Intermediate fusion strategies. (c) Late fusion strategies. Figure adapted from "Multimodal deep learning for biomedical data fusion: a review."[90]

Natural Language Processing

Natural language processing (NLP) is a branch of AI in which computers attempt to interpret human language in written or spoken form. NLP combines many fields, such as computational linguistics (rule-based modeling of human language), machine learning, and deep learning to perform specific tasks. Using these technologies, computers can process human language, understand its meaning, or find hidden rules or patterns. With advancements in computing power and machine learning development, NLP has been applied in many fields. In the clinical field, NLP has been successfully used to process free-text EHR data for deep contextualized word representations,[91] information extraction,[92] semantic analysis,[93] and Chatbot.[94] In the coming sections, the three most relevant techniques will be described, including named entity recognition (NER), word embeddings, and text classification.

Named Entity Recognition

Natural language processing algorithms for clinical information extraction have been actively researched over the past years. The methods have evolved from simple logic and rule-based systems to complex deep learning architectures.[95, 96] One of the common ways for information extraction is by transforming free-text data into a coded form, such as universal medical language system (UMLS). In addition, a rule-based system using semantic lexicons was used to handle more complex linguistic features with the advances in deep learning, which played an important role in developing more capable models for natural language processing. Several NLP tools were integrated with deep learning models like named-entity recognition (NER).

NER model is a sub-task of a natural language process tool, which seeks to classify words into predefined groups and assign labels to them.[97] The NER models built with deep learning techniques can extract entities from text corpus by not only identifying the keywords of entities but also by leveraging the context of the entity in the sentence. Furthermore, with language model pre-trained embeddings, the NER models leverage the proximity of other words that appear along with the entity in domain-specific literature.[98] NER models were used to extract medication information[98] or medical terms[99] in previous studies and can be a useful tool to extract other medical concepts such as clinical exam results or medication adherence from free-text clinical notes.

Word Embeddings and Text Classification

Word embeddings were developed as a numerical representation of textual data that allows similar words have a similar representation. A pre-trained word embeddings can be used as input layers to deep neural networks. Word embeddings are a kind of classifying technique where each word is given a numeric vector in a predefined vector space. Each word is mapped to a unique vector. Several methods are used to train the word embedding to learn the vector values, such as word2vector[100] and GloVe (global vectors for word representation)[101]. There are two methods commonly used for word2vector models include Skip-gram[100] and CBOW (continuous bag of words)[100]. **Figure 11** shows the architectures of a Skip-gram model and **Figure 12** shows the architecture of CBOW. We used CBOW to preprocess operation notes as input features for the surgical outcome prediction models. Overall, the goal of word embedding is using a densely distributed representation of each word instead of a huge sparse word representation, such as a one-hot encoding.

As mentioned above, pre-trained word embeddings can be used as input features to a predictive model. One of the most common approaches is text classification.[102] Text classification is a machine learning or deep learning technique that assigns a set of predefined categories to text. This technique is one of the fundamental parts of NLP and has been applied in many fields such as sentiment analysis, topic labeling, spam detection, and outcome prediction model.

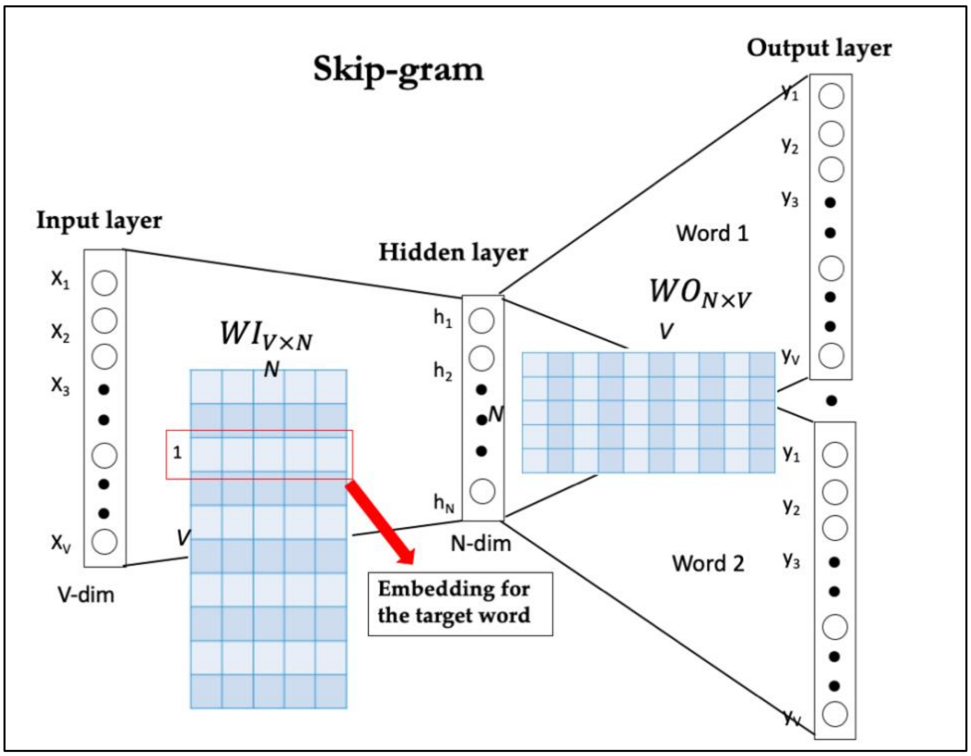

Figure 11. Architectures of Skip-gram. The figure shows the structure of the Skip-gram model, which used the target word to predict surrounding words. The ultimate goal of the word embedding models is not the outputs of the networks, and rather the goal is to learn the weights of the hidden layer that are used as "word embeddings."

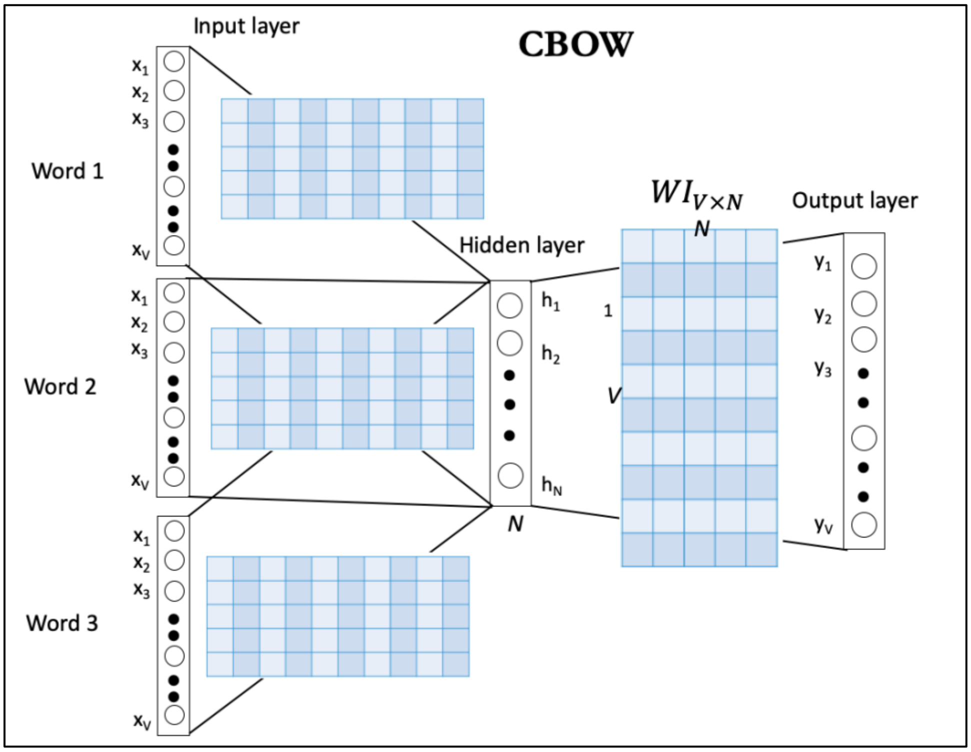

Figure 12. Architectures of CBOW. In the CBOW model, the surrounding words are combined to predict the target word. Similarly, the output of a CBOW model is the weights of the hidden layer. We used a CBOW model to convert the operation notes to pre-trained word embeddings.

Data Preprocessing

EHR data contains rich information, and secondary use of EHR data with machine learning/deep learning models can be a useful way for disease prediction. However, the raw EHR data may contain some errors and have plenty of missing data in real-world settings. For example, glaucoma patients might not have follow-up visits every six months or might not take eye exams for each visit. The missing data of the EHR dataset can cause problems for many machine learning algorithms. Specifically, missing data poses a major difficulty for modeling longitudinal data since most statistical models assume feature-complete data.[103] Therefore, it is important to manage missing data for secondary use of EHR data. Some studies handle this issue by removing subjects or time points with missing data, thus potentially losing a large quantity of data. Two major methods are used to manage missing data.[104] First, the missing data can be preprocessed in a separate step. This means we can impute the missing data with zero, variable's mean, or other machine learning strategies.[105, 106] Second, directly integrating the missing data issues into the machine learning models or training strategies. For example, marginalizing the missing data with Bayesian approaches.[107]

Several methods are used for preprocessing strategies to impute the missing data in a separate process. A basic and popular approach to data imputation is using statistical methods to estimate a value for the missing data and replace it. A common statistic calculated includes (1) mean value of the column, (2) median value of the column, (3) mode value of the column, and (4) zero or a constant value. For the time-series dataset, some other methods can be considered, such as forward filling and linear filling. Forward filling involved imputing the missing data with the last timepoint available data.[108] And the linear filling is similar to the forward filling but uses previous and future time points for imputation.[109] In addition, a sophisticated approach, "iterative imputation" is commonly used, which involves defining a model to predict each missing feature as a function of all other features and repeating this process of estimating feature values multiple times.[110] In the following study, linear imputation was used to handle the missing numeric data.

Model Evaluation Metrics

In general, the model performance is evaluated via some form of accuracy metrics on the training and test dataset for a supervised learning problem. In the training phase, a validation dataset or other validation methods, such as k-fold cross validation[111] were used to tune the algorithms to optimize the model settings that may work well for the new population. On the other hand, test dataset are samples from the original dataset that have not been seen by algorithms during the training phase. If the model shows much better performance on the training dataset rather than the test dataset, which refers to the model overfit to the training dataset.[112, 113] This indicates the model learned the detailed and noise pattern in the training data and this pattern does not apply to new data and has a negative effect on the model generalizability. Many techniques can be used to avoid or alleviate overfitting problems in deep learning or machine learning models, such as weight regularization, activity regularization, dropout, adding noise, and changing the model structure (e.g. reducing tree depth in a random forest model).[113] In contrast, underfitting refers to a model that performs poorly on both training and test datasets.[112] Ideally, a good fitted model can stand at the sweet spot between underfitting and overfitting with strong a performance on both datasets.

Aforementioned, a supervised learning problem is most commonly evaluated with some kind of prediction accuracy metrics. For a regression model, the average mean squared error (MSE) and R-Squared are popular evaluation metrics for the model performance. The MSE is calculated by the error between the predicted value and target value, which indicates how close a predicted value to the real one. R-Squared is the ratio of the sum of squares regression (SSR) and the sum of squares total (SST). R-Squared value is used to evaluate the goodness of its line, which indicates how good the regression model explains observed data.

For a classification model, the commonly used evaluation metrics include receiver operating characteristics (ROC) curve and area under ROC curve (AUROC), F1 score, precision, recall/sensitivity, and specificity. The outputs of most classification models are the probability of each class based on the threshold. Take binary classification, for example, the probability threshold is normally set at 0.5. With different threshold settings, some accuracy metrics can be changed.

Therefore, the ROC curve and AUROC are used to measure the classification ability at all threshold settings. The ROC curve is plotted with the true-positive rate (TPR) against the false-positive rate (FPR). The AUROC represents the model's ability to distinguish different classes. **Figure 13** illustrates the ROC curves with different AUROC. An excellent model has AUROC near 1, which means the model can perfectly separate two classes, while when AUROC is 0.5, it means the model has no class separation ability.

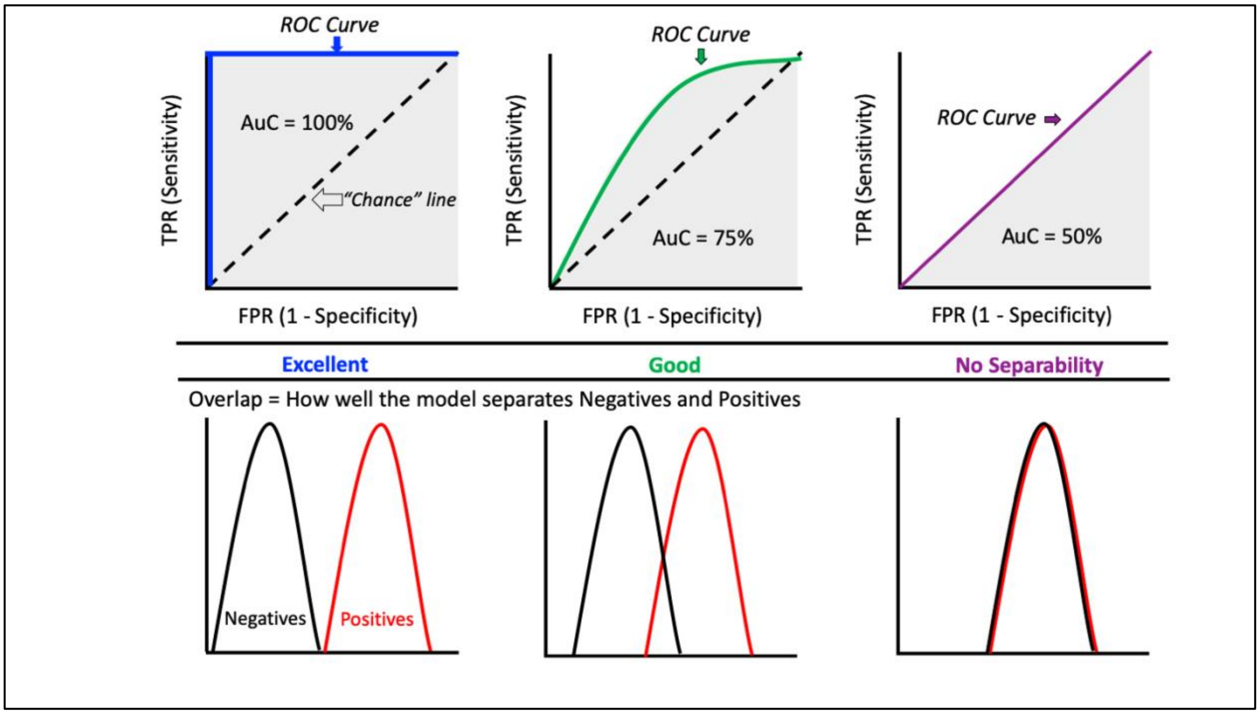

Figure 13. Receiver operating characteristics (ROC) curves illustration. The left figure shows an ideal situation. The two curves (on the bottom) do not overlap, which means the model has perfect separability. The middle figure shows a classification with good separability, but there is still part of the output that is mis-labeled. When AUC is 0.75, which means there is a 0.75 chance that the model can correctly distinguish between positive class and negative class. The right figure

is the worst situation. The model shows no capability to distinguish between positive class and negative class. And the AUC will be close to 0.5.

Besides, classification models are also commonly evaluated by using accuracy, F1 score, precision, recall, and specificity.[114] These metrics were computed by true positive (TP), false-positive (FP), false-negative (FN), and true negative (TN). The equations of these metrics are listed below.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

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

$$Recall/Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Where true positive or true negative were counted if the model correctly predicted an instance as positive or negative, respectively. On the other hand, if the model incorrectly labeled an instance as positive or negative, it is a false positive or false negative, respectively. The accuracy measure is pretty straightforward - it determines how the correct values are predicted. A recall score is the ratio of correctly predicted class to the all actual class, and a precision score is the ratio of correctly predicted class to the total predicted class. F1 -score uses a combination of precision and recall to calculate an average score for both of them.

For multiclass classification models, the evaluation metrics are more complicated. The evaluation metrics in binary classification need to be adapted to work for multiclass prediction. A standard method uses one vs. rest (ovr) and one vs. one (ovo) strategies. In the ovr metrics, the model compares each class against all the other classes, which means all other classes are treated as the same class. In this way, the selected class is considered as the "positive" class, while all the other classes are considered as the "negative" class. In the ovo metrics, all possible two-class combinations are compared and calculated. For a ROC curve or AUROC score, the model performance can be calculated by averaging all the ovr or ovo scores to get the final score. In general, one vs. rest (ovr) is more commonly used for adapting the evaluation metrics for a multiclass classification problem. In addition, macro averaging or micro averaging can be used to calculate the average F1 score, precision, and recall for all classes. Macro averaging is pretty straightforward - computing the arithmetic mean of all the per-class evaluation metrics. And the micro average aggregates the contributions of all classes to get the average metric.

AI Applications in Ophthalmology

In ophthalmology, although most AI techniques are used for image processing and computer-aided diagnosis (CAD) system development, there are still many classical ML models being used in this field. Clinical applications for disease diagnosis, monitoring, and risk assessment have recently become a focus of machine learning research. Examples of ophthalmologic diseases where machine learning has been applied to include glaucoma, age-related macular degeneration, and diabetic retinopathy. Random forest classifiers with EHR data have been used to predict the risks of cataract surgery complications.[115] Also, bootstrapped least absolute shrinkage and selection operator model were used to identify highly associated features.[115] Similarly, Chaganti et al. also utilized random forest classifiers to evaluate the predictive power of the different sources of data in identifying optic nerve disease.[116] In Fraccaro et al.'s study, several machine learning algorithms, such as support vector machines random forests and AdaBoost, were used to develop predictive models to diagnose age-related macular degeneration.[117] In addition, machine learning algorithms were used to predict glaucoma progression as well.[118]

Deep learning has been popularly applied in ophthalmology as well, especially for image-based diagnosis systems. Coyner et al. demonstrated that deep learning models could use for quality control for retinal fundus images,[119] generating synthetic retinal fundus images for data augmentation or physician training,[120] and developing the risk prediction model for retinopathy of prematurity patients.[121] Similarly, Brown et al. presented that the deep learning algorithm has comparable or better accuracy than human experts for plus disease diagnosis in retinopathy of prematurity.[122] Besides, combined deep learning models can be used to predict disease progression. In Dixit et al.'s study, the convolutional LSTM model can capture local and global trends in visual fields over time and is able to assess glaucoma progression. Furthermore, a combined architecture model using machine learning with output information from a deep learning model was used to identify high-risk patients with myopic regression after corneal refractive surgery. The researchers used ResNet50 (for image analysis) and XGBoost (for integration of all variables and fundus photography) to develop the prediction model.[123]

Although AI techniques have been popularly applied in ophthalmology fields, there are still many existing challenges. One of the major issues is data quality. The predictive model performance highly depends on the data quality. Structured EHR data collected from clinical practice may suffer from data quality problems, such as missing data, loss of follow-up visits, incorrect data entry, and incomplete information. While imaging data has less affected by the data quality issues among all other clinical data, there is no well-established standard for many imaging techniques. And the data access for image data in HER system is always complicated and frustrating. The data quantity is the most common issue for AI applications in many different fields, especially for deep learning algorithms. Comparing to classical machine learning algorithms, deep learning may need more training data to enhance the model performance. However, the EHR data from a single institute or research project are often limited and sharing EHR data across different institutes is extremely difficult. In addition, the model interpretation is another challenging point of AI application. People want to have some better understanding of model prediction rather than accept a "black box." Furthermore, most of the previously published studies in the ophthalmology field of AI application focused on diagnosis improvement or disease screening using imaging data, lacking sufficient attention to disease progression or surgical outcome predictions. We will discuss and address some of these challenges in the following chapters.

Chapter 3: Applications of Artificial Intelligence to Electronic Health Record Data in Ophthalmology

Note: The following was published in the Translational Vision Science & Technology journal. Citation: Lin, W. C., Chen, J. S., Chiang, M. F., & Hribar, M. R. (2020). Applications of artificial intelligence to electronic health record data in ophthalmology. Translational vision science & technology, 9(2), 13-13.

ABSTRACT

Widespread adoption of electronic health records (EHRs) has resulted in the collection of massive amounts of clinical data. In ophthalmology in particular, the volume range of data captured in EHR systems has been growing rapidly. Yet making effective secondary use of this EHR data for improving patient care and facilitating clinical decision-making has remained challenging due to the complexity and heterogeneity of this data. Artificial intelligence (AI) techniques present a promising way to analyze these multimodal datasets. Yet, while AI techniques have been extensively applied to imaging data, there are a limited number of studies employing AI techniques on EHR data. The objective of this review is to provide an overview of different AI methods applied to EHR data in the field of ophthalmology. This literature review highlights that the secondary use of EHR data has focused on glaucoma, diabetic retinopathy, age-related macular degeneration, and cataracts using artificial intelligence techniques. These techniques have been used to improve ocular disease diagnosis, risk assessment, and progression prediction. Techniques such as supervised machine learning, deep learning, and natural language processing were most commonly used in the articles reviewed.

INTRODUCTION

Rapid adoption of electronic health records (EHRs) in recent decades has generated large volumes of clinical data with potential to support secondary use in research.[124-126] Indeed, one recurring justification for EHR adoption has been to support collection and analysis of "big data" to gain meaningful insights.[127, 128] The clinical research community has expressed growing interest in developing effective techniques to reuse clinical data from EHRs, in part due to the benefits of secondary data reuse over primary data collection.[129, 130] Researchers reusing EHR data may not need to recruit patients or collect new data, potentially reducing cost compared to traditional clinical research. Moreover, EHR data often contain valuable longitudinal data regarding a patient's status, medical care, and disease progression which have been previously shown to support clinical decision support,[131] medical concept extraction,[132] diagnosis,[133] and risk assessment.[134]

However, there are challenges associated with reusing EHR data, particularly due to its complexity and heterogeneity. For example, in ophthalmology, patient data contained in EHRs may include fields as diverse as demographic information, diagnoses, laboratory tests, prescriptions, eye exams, imaging, and surgical records. Interpreting these heterogeneous data require strategies such as information extraction, dimension reduction, and predictive modeling typical of machine learning, and more broadly artificial intelligence (AI) techniques. Applying AI to EHR data has been productive in a variety of domains. For instance, studies in cardiology have broadly employed AI techniques with EHR data for early detection of heart failure,[135] predicting the onset of congestive heart failure,[136] and improving risk assessment in patients with suspected coronary artery disease.[137] Likewise in ophthalmology, machine learning models with EHR data have been used to predict risks of cataract surgery complications, improve diagnosis of glaucoma and age-related macular degeneration, and perform risk assessment of diabetic retinopathy.[138-141]

While the application of artificial intelligence to EHR data related to ocular diseases has increased over the past decade, there have been no published reviews of this literature. One literature review of machine learning techniques applied in ophthalmology was published in 2017,[142] however, the included studies mainly focused on application of machine learning techniques to imaging data, rather than EHR data. This manuscript addresses this knowledge gap by reviewing the literature applying artificial intelligence techniques to EHR data for ocular disease diagnosis and monitoring. With this review, we explore the type of AI techniques used, the performance of these techniques, and how AI has been applied to specific ocular diseases, providing future directions to clinical practice and research.

METHODS

An exhaustive search was performed in the PubMed database using search terms related to "Artificial intelligence", "Electronic health records", and "Eye" in any field of articles. See the Appendix for the full query. The results were then examined and narrowed according to the following criteria:

1. Duplicates were removed.
2. Studies were eliminated for lack of relevance after review of the title and abstract; studies that used only imaging data without any EHR data were excluded.
3. Studies without direct clinical application or not related to the topic were excluded.

The review process is summarized in **Figure 1**. One author (WL) identified articles for inclusion through manual title, abstract, and content review. Two authors (WL & JSC) extracted data about each study: the aim, disease, algorithm, specific techniques, performance assessment and conclusion of the articles that met the inclusion criteria, as summarized in **Table 1**.[138-141, 143-151]

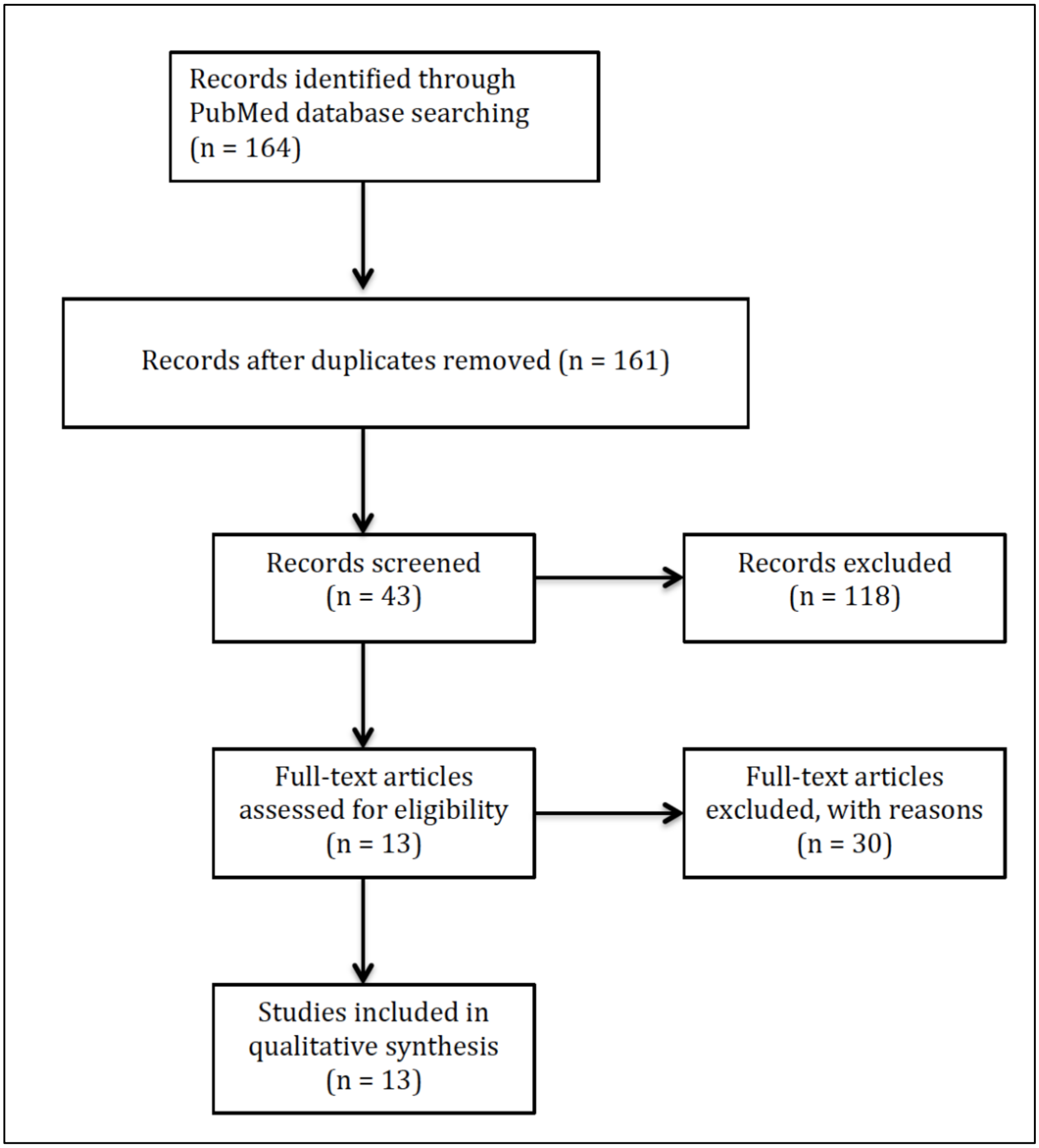

Figure 1: Flow diagram for the literatures selection.

Table1. Studies on ocular diseases using artificial intelligence techniques with EHR data

Author	Aim	Disease	Algorithm Type	Specific Techniques	Performance	Conclusions
Lin et al.[143]	Disease detection	Myopia	Supervised machine learning	Random forest	95% confidence interval (CI) for predicting onset of high myopia. 3 years onset prediction (AUC: 94%-98.5%), 5 years (85.6% - 90.1%), 8 years (80.1% - 83.7%)	Use machine learning with EHR data can accurately predict myopia onset.
Lee et al.[144]	Improve diagnostic accuracy	Age-Related Macular Degeneration (AMD)	Deep learning	Convolutional neural networks	For each patient, AUC (97.45%), accuracy (93.54%), sensitivity (92.64%), and specificity (93.69%)	Linked OCT images to EMR data can improve the accuracy of the deep learning model, which used to distinguish AMD from normal OCT images.
Baxter et al.[145]	Risk assessment	Open-angle glaucoma	Supervised machine learning Deep learning	Logistic regression, Random forests, ANNs	AUC of logistic model (67%), random forest (65%), ANNs (65%)	Existing systemic data in the EHR can identify POAG patients at risk of progression to surgical intervention.
Chaganti et al.[139]	Identify risk factors and improve diagnostic accuracy	Glaucoma, intrinsic optic nerve disease, optic nerve edema, orbital inflammation, and thyroid eye disease	Supervised machine learning	Random forest	AUC of classifiers: glaucoma (88%), intrinsic optic neuritis (76%), optic nerve edema (78%), orbital inflammation (77%), thyroid eye disease (85%)	EMR phenotype (from pyPheWAS) can improve the predictive performance of random forest classifier with imaging biomarkers.
Apostolova et al.[146]	Patient identification	Open globe injury	Supervised machine learning Text-mining	SVM NLP - Word embeddings	Text classification: precision (92.50%), recall (89.83%)	Free-form text with machine learning methods can used to identify open globe injury.
Saleh et al.[141]	Risk assessment	Diabetic retinopathy (DR)	Supervised machine learning	Fuzzy random forest (FRF), dominance-based rough set approach (DRSA)	Performance of FRF: Accuracy (80.29%), sensitivity (80.67%), specificity (80.18%) Performance of DRSA: Accuracy (77.32 %), sensitivity (76.89 %), specificity (77.43%) of DRSA.	Ensemble classifiers (RFR and DRSA) can be applied for diabetic retinopathy risk assessment. The 2-steps aggregation procedure is recommended.

Rohm et al.[147]	Predict progression	Age-related macular degeneration (AMD)	Supervised machine learning	AdaBoost, Gradient Boosting, Random Forests, Extremely Randomized Trees, Lasso	Accuracy of logMAR visual acuity (VA) prediction after VEGF injections. 3 month: MAE (0.14), RMSE (0.18) 12 month: MAE (0.16), RMSE (0.2)	EHR data of patients with neovascular AMD can be used to predict visual acuity by using machine learning models.
Yoo et al [148]	Risk assessment	Diabetic retinopathy	Supervised machine learning	Ridge, elastic net, and LASSO	In external validation, LASSO predicted DR: AUC (82%), accuracy (75.2%), sensitivity (72.1%), and specificity (76.0%).	LASSO with high-dimensional EHR can be used to predict DR risk among diabetic patients.
Fraccaro et al.[140]	Improve diagnostic accuracy	Age-related macular degeneration (AMD)	Supervised machine learning	Logistic regression, Decision trees, SVM, random forests, and AdaBoost.	AUC of random forest, logistic regression, and AdaBoost (92%); SVM, decision trees (90%)	Machine learning algorithms using clinical EHR data can be used to improve diagnostic accuracy of AMD.
Sramka et al.[149]	Improve surgical outcome	Cataracts	Supervised machine learning Deep learning	Support vector machine regression (SVM-RM) Multilayer neural network ensemble model (MLNN-EM)	Both SVM-RM and MLNN-EM achieved significantly better results than the Barrett Universal II formula in the ±0.50 D PE category.	SVM-RM and MLNN-EM with EHR data can be used to improve clinical IOL calculations and improve cataract surgery refractive outcomes.
Peissig et al.[150]	Patient identification	Cataracts	Text-mining	Natural language processing (NLP)	The multi-modal model shows results including sensitivity (84.6%), specificity (98.7%), PPV (95.6%), and NPV (95.1%)	A multi-modal strategy incorporating optical character recognition and natural language processing can accurately increase the number of cataracts cases identified.
Gaskin et al.[138]	Identify and predict risks of cataract surgery complications	Cataract	Supervised machine learning Text-mining	Bootstrapped LASSO, Random forest Natural language	Based on the LASSO model, younger age (<60 years old), prior anterior vitrectomy or refractive surgery, history of age-related macular degeneration (ARMD), and complex cataract surgery were risk	Bootstrapped LASSO can be used to identify risk factors of post-operative complications of cataract surgery. Random forest shows good reliability for predicting cataract surgery complications.

				processing (NLP)	factors associated with postoperative complications. The random forest model shows high NPV > 95% and moderate sensitivity (67%) and AUC (65%)	
Skevofilakas et al.[151]	Risk assessment	Diabetic retinopathy	Deep learning Supervised machine learning	Feed forward Neural Network (FNN) and improved Hybrid Wavelet Neural Network (iHWNN) Classification and Regression Tree (CART)	AUC of hybrid Decision Support System (DSS) (98%), iHWNN (97%), FNN (88%), and CART (86%).	Hybrid DSS trained on imaging and related EHR data can estimate the risk of a type I DM patient developing diabetic retinopathy.

RESULTS

The PubMed query returned 164 articles published through August 2019. In total, 161 papers were reviewed after removing 3 duplicates. 118 papers were excluded due to lack of relevance based on title and abstract. A total of 13 articles were considered which met inclusion criteria (**Figure 1**).

Artificial Intelligence Techniques

Three major techniques were used in these studies: 11 studies used supervised *machine learning*, of which 3 studies specifically employed a *deep learning* technique and 2 studies also used *natural language processing* to generate structured data suitable for analysis from unstructured text. Only one study used deep learning by itself and another study used natural language processing independent of other techniques (**Table 1**). **Figure 2** illustrates a simplified machine learning process, and the relationship among these three techniques. In short, natural language processing can be used to extract useful information from text-based data and process it into a format suitable for machine learning. Supervised machine learning techniques, some of which employ deep learning algorithms, can then be applied to these and other structured datasets to develop predictive models or classifiers.

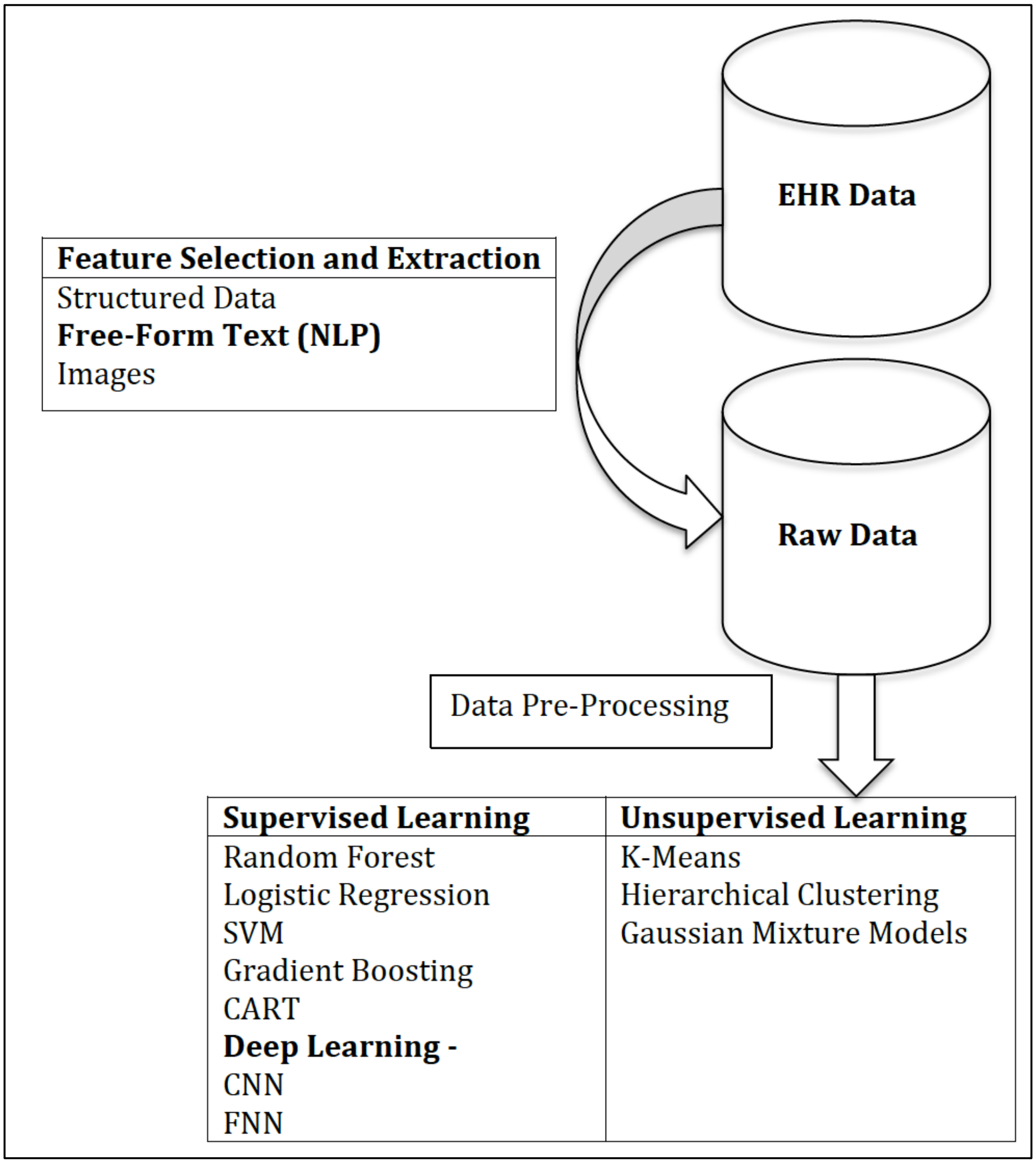

Figure 2: Schematic of the steps of machine learning application. NLP, natural language processing; SVM, support vector machine; CART, classification and regression tree; CNN, convolutional neural network; FNN, feed forward neural network.

Machine Learning

Machine learning techniques are computational methods that learn patterns or classifications within data without being explicitly programmed to do so.[152] Machine learning can be divided into two methods based on the use of "ground truth" data: supervised learning and unsupervised learning. In supervised learning, a model learns from "ground truth" data in a training dataset that contains labeled output data, and then can predict the output for new cases. The algorithm is typically a classifier with categorical output or a regression algorithm with continuous output. In unsupervised learning, the model learns from a training dataset without labeled output and identifies underlying patterns or structures within its input data. In medicine, machine learning has been widely used in several specialties such as radiology, cardiology, oncology, and ophthalmology for improving diagnostic accuracy and early diseases detection.[153] In this review, most studies used supervised machine learning techniques such as logistic regression,[76] support vector machines (SVMs),[154] Classification and Regression Tree (CART),[155] random forest,[75] AdaBoost,[80] and gradient boosting.[156]

As shown in **Figure 3B**, logistic regression is an extension of linear regression (**Figure 3A**). In linear regression, the data is modeled as a linear relationship that can be used to predict a value for a given input. In logistic regression, a non-linear function, called the logistic function, converts prediction values into binary categories based on a threshold. Some methods can be used to improve the prediction accuracy of logistic regression, such as least absolute shrinkage and selection operator (LASSO).[157] LASSO is a statistical method that selects a smaller subset of predictor variables most related to the outcome variable and shrinks regression coefficients to improve accuracy and generalizability. SVM is another popular machine learning model used for classification analysis. As shown in **Figure 3C**, a boundary is created to split input data into two distinct groups and can be used to classify new data into similar distinct categories.

A decision tree is an important supervised machine learning algorithm. **Figure 3D** illustrates a decision tree with a root node as a start followed by the branched nodes and terminal nodes. The root node is the first decision node representing the best predictor variable. Each branched node represents the output of a given input variable. As more input variables are added to subsequent branching nodes, the decision tree becomes more sophisticated in predicting the outcome variable at the terminal nodes.

Ensemble methods combine multiple machine learning models and are commonly used to improve the performance of prediction models. The two most common methods: bootstrapping aggregation (bagging) and boosting were shown in **Figure 3E**. In a bagging method, multiple subsets of data are randomly selected from the original dataset and each subset data are used to train a separate prediction model. The final predictions will be aggregated from all prediction models. Random forest algorithms are examples of an ensemble machine learning method that combine bagging and decision trees. Boosting is another technique that combines multiple models to create a more accurate one. Adaboost and gradient boosting are widely used boosting machine learning algorithms.

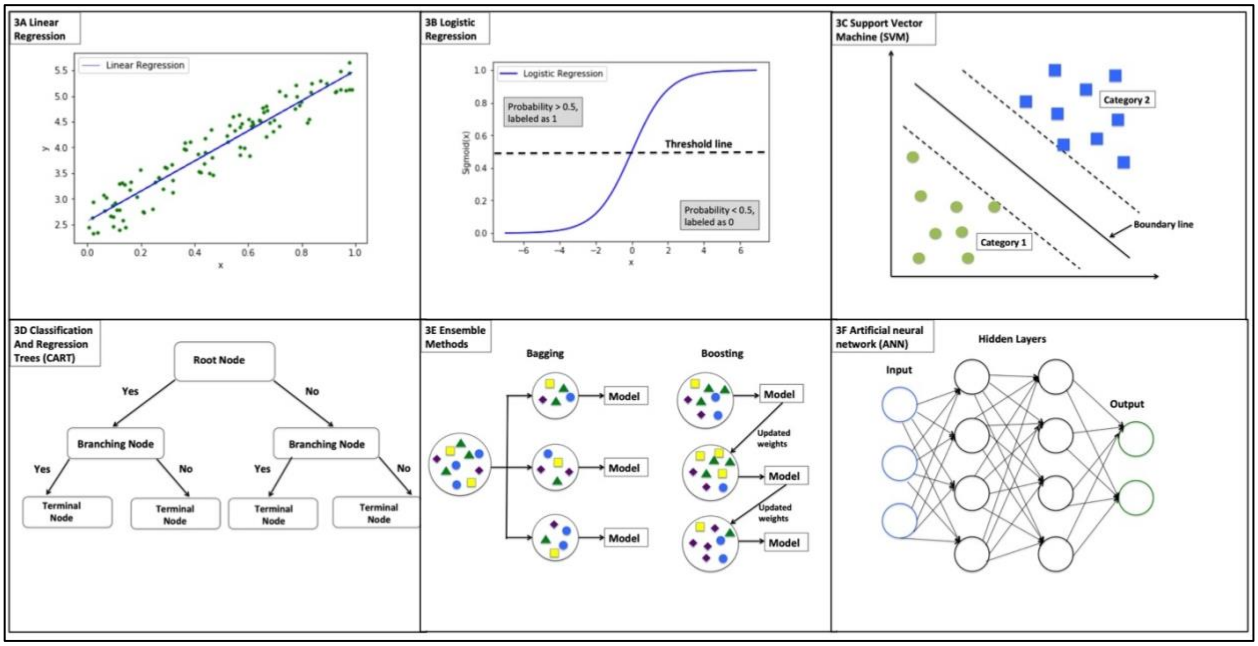

Figure 3: Illustrations of machine learning models. 3A. Linear regression; 3B. Logistic regression; 3C. Support vector machine; 3D. Classification and regression trees (CART); 3E. Ensemble methods; 3F Artificial neural network (ANN).

As shown in the **Table 1**, random forest was used by Lin et al.[143] to predict myopia onset and by Chaganti et al.[116] to improve the diagnostic accuracy of glaucoma. In addition, Baxter et al.[118] used random forest and logistic regression to identify patients with open-angle glaucoma who had a risk of progression to surgical intervention. Fraccaro et al.[117] used logistic regression, decision trees, SVMs, random forests, and AdaBoost to improve diagnostic accuracy of AMD. In addition, fuzzy random forest (FRF) and dominance-based rough set approach (DRSA) were used by Saleh et al.[141] for DR risk assessment. And Gaskin et al.[115] used random forest and bootstrapped LASSO to identify and predict risks of cataract surgery complications. Moreover, Yoo and Park[148] used elastic net and LASSO to predict DR risk among diabetic patients.

Deep Learning

Deep learning is a subset of machine learning techniques based on artificial neural networks (ANN) that mimic human brain processing. As shown in **Figure 3F**, multiple layers of computation are constructed in a deep learning model, and each layer is used to perform computations on data from the previous layer. The layers between the input layer and the output layer are called hidden layers. While the information may flow from the input to subsequent output layers (feedforward), information can also flow backward from hidden layers to input layers (backpropagation). The input and output of hidden layers are not reported; deep learning algorithms only present the final outcome of the output layer.[81] Deep learning does not use structured features for input as machine learning does; therefore, deep learning is useful for raw images since they do not have to be pre-filtered as they do for machine learning algorithms. After processing raw input through multiple layers within deep neural networks, the algorithms find appropriate features for classifying output. In this review, several articles used deep learning algorithms such as artificial neural networks (ANNs),[158] convolutional neural networks (CNNs),[159] multilayer neural network ensemble models (MLNN-EM),[160] and feed forward neural networks (FNNs).[161] CNN is a subtype of deep neural network commonly used in image classification. In a CNN model, special convolution and pooling layers are used to reduce a raw image to essential features necessary for the model to classify or label the image. In other words, these techniques use machine learning to determine model input features from the raw image data, rather than a human or a separate image processing program. MLNN-EM is a learning technique that integrates several neural networks to aggregated outcome. In addition, FNN is another subtype of neural network where the information moves forward in (one direction) from root nodes; information never moves backwards. The nodes between input and out layers do not form a cycle of information.

As shown in **Table 1**, Lee et al.[20] used CNNs to distinguish AMD from normal OCT images, Baxter et al.[21] used ANNs to identify open-angle glaucoma patients at risk of progression to surgery. Also, Sramka et al.[149] used MLNN-EM and support vector machine regression (SVM-RM) to improve clinical intra-ocular lens (IOL) calculations and Skevofilakas et al.[151] used feed forward Neural Network (FNN) and improved hybrid wavelet neural network (iHWNN) to develop hybrid decision support system for predicting DR risk among diabetic patients.

Natural language processing (NLP)

NLP is a branch of artificial intelligence where computers attempt to interpret human language, in written or spoken form. By utilizing NLP, researchers can extract information from text; some uses in medicine include separating progress notes into sections, determining diagnoses from notes, and identifying the documentation of adverse events.[162] As shown in **Table 1**, Apostolova et al.,[146] Peissig et al.,[150] and Gaskin et al.[138] describe the use of natural language processing in extracting cataract information from free-form text clinical notes.

Outcome Metrics for Evaluation of Performance of AI Techniques

The evaluation of the performance of different AI techniques depends on the chosen algorithm, the purpose of the study, and the input dataset. In supervised machine learning algorithms, classifiers are evaluated based on a comparison between the known categorical output and the predicted categorical output. For outputs with two categories, the accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) can be computed. Another important evaluation metric is the AUC-ROC (Area under the curve – Receiving operating characteristic), which is used to evaluate performance of classifiers based on different thresholds.

ROC is a probability curve that visualizes the true positive rate (sensitivity) change with respect to false positive rate (1 – specificity) for different threshold values used in the model. The AUC represents the ability of a model to distinguish between different outcome values.[163] An AUC equal to 1 is ideal and represents the model's ability to perfectly distinguish between two outcomes. On the other hand, an AUC of approximately 0.5 is the worst case because it means that the model is not better than chance for distinguishing between two outcomes.

As shown in **Table 1**, eight studies used AUC-ROC to evaluate the performance of classifiers.[138-140, 143-145, 148, 151] The range of AUC-ROC was from 65% to 98.5% and the median AUC in all included studies was 90%. In addition, precision and recall were used to evaluate the performance of text-mining algorithms.[164] Apostolova et al.[146] and Peissig et al.[150] used precision and recall to evaluate the performance of text classification.

For regression models, two evaluation metrics, mean absolute error (MAE) and root mean squared error (RMSE), are commonly used to measure accuracy for continuous variables. They measure the average difference between actual observations and predictions. MAE shows the absolute differences with equal weight for each difference. In contrast, RMSE penalized larger errors by taking the square of the difference before averaging. In Rohm et al.'s[147] study, MAE and RMSE were used to evaluate visual acuity prediction.

Application of AI to Clinical Ophthalmology

Artificial intelligence techniques have been applied clinically to improving ocular disease diagnosis, predicting disease progression, and risk assessment (**Table 1**). Several diseases were studied in papers included in this review including glaucoma, cataract, age-related macular degeneration (AMD), and diabetic retinopathy (DR). We will present the benefits of artificial intelligence techniques with EHR data in these diseases.

Glaucoma

Two studies in this review focused on the field of glaucoma and used supervised machine learning techniques to improve diagnosis and predict progression.[139, 145] In Chaganti et al.'s study,[16] a good performance was obtained (AUC of glaucoma diagnosis 88%) and results showed that addition of electronic medical record (EMR) phenotype could improve the classification accuracy of a random forest classifier with imaging biomarkers.[139] On the other hand, Baxter et al.[145] reported a moderate performance (AUC 67%) in a study that used EHR data to predict risk of progression to surgical intervention in open-angle glaucoma patients. In addition to model performance, it is also important to know which factors can be used to improve disease diagnosis. The work performed by Chaganti et al.[139] began to explore this problem by comparing performance of classifiers using EMR phenotype, visual disability scores (VDS), and imaging metrics.

Cataracts

Three studies applying different artificial intelligence techniques to cataract diagnosis and management were reviewed. In Peissig et al.'s study,[150] natural language processing was used to extract cataract information from free-text documents. An EHR-based cataract phenotyping algorithm, which consisted of structured database querying, information from free-text notes, and optical character recognition on scanned clinical images, was developed to identify cataract subjects. The result of the study showed good performance (PPV >95%).[150] Additionally, Gaskin et al.[15] used supervised machine learning algorithms to identify risk factors and to predict intra-operative and post-operative complications of cataract surgery. The investigators used data mining via NLP to extract cataract information from the EHR system.[138] The predictive model showed moderate performance (AUC 65%) and the risk factors associated with surgical complications included: younger patients, refractive surgery history, AMD history, and complex cataract surgery. These risk factors were associated with post-operative complications, and the predictive model showed moderate performance (AUC 65%). Supervised machine learning (support vector machine regression – SVM-RM) and deep learning (multilayer neural network ensemble model - MLNN-EM) algorithms were used to improve the intraocular lens (IOL) power calculation by Sramka et al.[149] Both SVM-RM and MLNN-EM model provided better IOL calculations than the Barrett Universal II formula.

Age-Related Macular Degeneration (AMD)

Three studies used AI in AMD were reviewed. Lee et al.[20] used deep learning techniques to improve the diagnosis of AMD. Optical coherence tomography (OCT) images of each patient were linked to EMR clinical end points extracted from EPIC (Verona, WI) for each patient to predict a diagnosis of AMD. The model had high accuracy with an AUC 97% in distinguishing AMD from normal OCT images.[144] Another study conducted by Rohm et al.,[147] used supervised regression models to accurately predict visual acuity in response to anti-vascular endothelial growth factor (VEGF) injections in patients with neovascular AMD. Models predicting treatment response may have implications in encouraging patients adhering to intravitreal therapy. Also, as demonstrated by Fraccaro et al.,[140] supervised machine learning techniques can be incorporated into EHR systems providing real-time support for AMD diagnosis.

Diabetic Retinopathy

DR is one of the most common comorbidities of diabetes and frequent screening exams for diabetic patients are resource consuming. Three studies explore this problem by using AI techniques with EHR data to determine patient risk for developing DR. Saleh et al.[141] used two kinds of ensemble classifiers: fuzzy random forest (FRF) and DRSA to predict DR risk using EHR. Good performance (accuracy 80%) of the FRF model was shown in this study. Similarly, Yoo et al.[148] proposed a comparison between the learning models: ridge, elastic net, and LASSO using the traditional indicators of DR. They showed that the performance of LASSO (AUC 81%) was significantly better than the traditional indicators (AUC of glycated hemoglobin 69%; AUC of fasting plasma glucose 54%) in diagnosing DR. In addition, a hybrid Decision Support System was developed by Skevofilakas et al.[151] to estimate the risk of a type I DM patient to develop DR. The hybrid DSS showed an excellent performance with an AUC of 98%. Overall, these studies show that integrating these techniques with an EHR system has promise in improving early detection of diabetic patients at risk of DR progression.

DISCUSSION

This article reviews the literature applying artificial intelligence techniques to EHR data to aid ocular disease diagnosis and risk assessment. We focus the discussion on three areas: AI techniques used to analyze EHR data, the performance of techniques, and the ocular diseases most commonly analyzed.

First, secondary use of EHR data via artificial intelligence techniques can be used to improve ocular disease diagnosis, risk assessment and disease progression. The predictive models across the eight classifiers showed good performance with a median AUC of 90%. We found one study focused on postoperative complications of cataract prediction reported moderate accuracy with 65%.[138] The reason may be due to insufficient predictors, such as lack of surgeon relevant information. Also, standard classification techniques may not be able to handle imbalanced data very well.[165, 166] For example, when a dataset contains a very few number of disease or complications cases, there is not enough data about these cases for the model to accurately learn how to predict these cases. On the other hand, excellent performance of classifiers trained on combined EHR and image data were reported by Skevofilakas et al.[151] and Lee et al.[144] For future studies, one feasible direction might be to develop the hybrid model that employs both the routine EHR data and image datasets to have a more complete picture of patient variables associated with the outcome of interest.

Second, supervised machine learning was the most common technique used with EHR data to analyze ocular diseases. Studies often focused on improving diagnosis, predicting progression, or risk assessment for early detection. The predictors defined were based on the risk factors of disease, demographic features found from literature review and clinical experiences. None of the studies reviewed used unsupervised machine learning techniques where the desired output and the relationship between the outcome variable and the predictors are unknown. These methods are used to identify clusters of data that are similar and can help discover the hidden factors that are useful for improving the diagnosis. However, unsupervised learning has been successfully applied to other fields. For example, Marlin et al.[167] demonstrated that the probabilistic clustering model for time series data from real-world EHR could be able to capture patterns of physiology and be used to construct mortality prediction models. For future studies, unsupervised machine learning techniques might be used to find hidden patterns from EHR data for improving clinical predictions of ocular diseases.

Finally, in this review, studies that analyzed EHR data with AI techniques mainly focused on four diseases: glaucoma, DR, AMD, and cataracts. The focus on these diseases (glaucoma, AMD, and DR) is likely due to their prevalence as the major causes of irreversible blindness in the world.[168] Early detection or treatment can delay or halt the progression of such diseases, reduce visual morbidity, and preserve a patient's quality of life.[169, 170] AI techniques can be used to achieve this goal. Furthermore, cataract surgery is the most common refractive surgical procedure and is one of the most common surgeries performed in ophthalmology.[171] Risk assessment of the postoperative complications and decreasing the risk of re-operation are crucial to patient outcomes, and AI techniques can help approach these issues.

This review presents the AI techniques used in vision sciences based on EHR data. However, several problems still need to be addressed for future studies. One of the major problems is data quality. EHR data used for research is essentially different from data collected during a traditional clinical research study. EHR data collected from clinical practice may have incomplete information due to incorrect data entry, non-answers, and recording errors. Consequently, the performance of machine learning models will be dependent on data quality and is an issue when using AI techniques with EHR data.[172-175] Additionally, except for the work reported by Lin et al.,[143] all reviewed studies were single-center studies. Thus, the results of studies may not be generalizable to other healthcare systems.

While imaging data do not suffer from the data quality issues of other clinical data, there is no well-established gold standard for many imaging techniques. For instance, Garvin et al.[176] presented an automated 3D intraretinal layer segmentation algorithm using OCT image data. The gold standard was determined by two retinal experts' recommendations. This requires more time and resources to analyze and cross-validate the outcomes. Also, different pre- and post-processing algorithms, hardware configurations, and image processing steps are intended to improve image quality for easier automated diagnosis. However, these factors often make models difficult to replicate. In addition, using imaging analysis without other prior information, such as medical history information, may also affect the model performance and lead to biased results. Therefore, integration of imaging data and routine EHR data allow us to obtain prior information to input to the predictive model.

CONCLUSION

Artificial intelligence techniques are rapidly being adopted in ophthalmology, and have potential to improve the quality and delivery of ophthalmic care. Moreover, secondary use of EHR data is an emerging approach for clinical research involving artificial intelligence, particularly given the availability of large-scale data sets and analytic methods.[177-179] In this review, we describe applications of artificial intelligence methods to ocular diseases and problems such as diagnostic accuracy, disease progression, and risk assessment, and find that the number of published studies in this area has been relatively limited due to challenges with the current quality of EHR data. In the future, we expect that artificial intelligence using EHR data will be applied more widely in ophthalmic care, particularly as techniques improve and EHR data quality issues are resolved.

APPENDIX

PubMed Search Query

(
(("electronic"[All Fields] AND "health"[All Fields] AND record[All Fields])
OR (“electronic"[All Fields] AND "medical"[All Fields] AND record[All Fields]) OR
(("computerised"[All Fields] OR "computerized"[All Fields]) AND "medical"[All Fields] AND record[All Fields])
OR ("electronic health records"[MeSH Terms])
OR (“medical records systems, computerized"[MeSH Terms])
OR (“electronic health data” [All Fields])
OR (“personal health data” [All Fields])
OR (“personal health record” [All Fields])
OR (“personal health records” [All Fields])
OR (“Health Record” [All Fields])
OR (“computerized patient medical records” [All Fields])
OR (“computerized medical record” [All Fields])
OR (“computerized medical records” [All Fields])
OR (“computerized patient records” [All Fields])
OR (“computerized patient record” [All Fields])
OR (“computerized patient medical record” [All Fields])
OR (“electronic health records” [All Fields])
OR (“electronic patient record” [All Fields])
OR (“electronic healthcare record” [All Fields])
OR (“ehr” [All Fields])
OR (“ehrs” [All Fields])
OR (“emr” [All Fields])
OR (“emrs” [All Fields])
OR (“phr” [All Fields])
OR (“phrs” [All Fields])
OR (“patient record” [All Fields])
OR (“patient health record” [All Fields])
OR (“healthcare record” [All Fields])
)

AND

(
("Machine"[All Fields] AND "Learning"[All Fields])
OR (“Artificial"[All Fields] AND "intelligence"[All Fields])
OR ("deep learning"[All Fields])
OR (“Machine intelligence"[All Fields])
OR (“Natural language processing"[All Fields])
OR (“NLP"[All Fields])
OR (“ML"[All Fields])

OR (“DL"[All Fields])
OR (“vector machine "[All Fields])
OR (“random forest"[All Fields])
OR (“neural network "[All Fields])
)

AND
English[lang]

AND

(
("Ophthalmology"[All Fields])
OR (“Eye"[All Fields])
OR (“Vision"[All Fields])
OR (“visual"[All Fields])
OR (“Diabetic retinopathy"[All Fields])
OR (“Cataract"[All Fields])
OR (“Glaucoma” [All Fields])
OR (“Cornea” [All Fields])
OR (“Pediatric Ophthalmology and Strabismus” [All Fields])
OR (“Retina” [All Fields])
OR (“Retinal disease” [All Fields])
OR (“Uveitis” [All Fields])
OR (“Neuro-ophthalmology” [All Fields])
OR (“Ophthalmic genetics” [All Fields])
OR (“Inherited retina diseases” [All Fields])
OR (“Oculoplastics” [All Fields])
OR (“Ocular Oncology” [All Fields])
OR (“Cataract surgery” [All Fields])
OR (“Comprehensive eye care” [All Fields])
OR (“Oculofacial plastics and reconstructive surgery” [All Fields])
OR (“Vision rehabilitation” [All Fields])
OR (“Contact Lenses” [All Fields])
OR (“Myopia” [All Fields])
OR (“Age-related macular degeneration” [All Fields])
OR (“ROP” [All Fields])
OR (“Congenital cataract” [All Fields])
OR (“Low vision” [All Fields])
OR (“Pediatric ophthalmology” [All Fields])
OR (“Strabismus” [All Fields])
OR (“Oculoplastic surgery” [All Fields])
OR (“Comprehensive ophthalmology” [All Fields])
OR (“Refractive error” [All Fields])
OR (“Refractive surgery” [All Fields])
)

Chapter 4: Extraction of Active Medications and Adherence Using Natural Language Processing for Glaucoma Patients

Note: The following was published in the AMIA Annual Symposium Proceedings 2021. Citation: Lin, W. C., Chen, J. S., Kaluzny, J., Chen, A., Chiang, M. F., & Hribar, M. R. (2021). Extraction of Active Medications and Adherence Using Natural Language Processing for Glaucoma Patients. In AMIA Annual Symposium Proceedings (Vol. 2021, p. 773). American Medical Informatics Association.

ABSTRACT

Accuracy of medication data in electronic health records (EHRs) is crucial for patient care and research, but many studies have shown that medication lists frequently contain errors. In contrast, physicians often pay more attention to the clinical notes and record medication information in them. The medication information in notes may be used for medication reconciliation to improve the medication lists' accuracy. However, accurately extracting patient's current medications from free-text narratives is challenging. In this study, we first explored the discrepancies between medication documentation in medication lists and progress notes for glaucoma patients by manually reviewing patients' charts. Next, we developed and validated a named entity recognition model to identify current medication and adherence from progress notes. Lastly, a prototype tool for medication reconciliation using the developed model was demonstrated. In the future, the model has the potential to be incorporated into the EHR system to help with real-time medication reconciliation.

INTRODUCTION

The rapid adoption of electronic health records (EHRs) has generated large-scale clinical data that has been re-used for many purposes, including patient phenotyping,[180] pharmacovigilance,[181, 182] comparative effectiveness research,[183] clinical decision support,[184, 185] and quality improvement and research.[186] Although secondary use of EHR shows many benefits such as improved healthcare quality, reduced healthcare costs, and effective clinical research,[130, 187] there are many challenges that still need to be addressed. One of the biggest challenges is the accuracy and completeness of EHR data, specifically medication information.[188]

The accuracy of medication data is crucial for patient safety, quality of care, and clinical research. Inaccurate or incomplete medication records can lead to polypharmacy, adverse medication interactions, and decreased data reliability in research.[189] The medication list is a structured record of a patient's medication data which is populated automatically by electronically prescribed medications or manually through medication reconciliation.[190] However, the EHR system may not always capture medication data correctly or prevent errors in the medication list.[191] Previous studies have shown that medication lists frequently contain errors, including duplicated documentation of medications, outdated discontinued prescriptions in the medication list, and missing medications prescribed elsewhere.[190, 192-196] In addition, prior studies show that physicians direct very little attention to EHR medication lists, and instead spend most time reviewing the impression and plan section.[197, 198] It seems reasonable to expect that medications recorded in narrative notes are more reliable and can be helpful with medication reconciliation. Medication reconciliation is a process to create and maintain patients' most current and accurate list of medications.[199, 200]

However, manual reviewing progress notes for medication data extraction in EHR is time-consuming and labor-intensive. Natural language processing (NLP) is a promising strategy for capturing medication information from the free-text progress note. With advancements in machine learning and the large text corpora available in EHR, NLP has been successfully used to process free-text EHR data, for deep contextualized word representations,[201] information extraction,[202] and semantic analysis.[203] Named entity recognition (NER) is a sub-task of information extraction, which seeks to identify words or phrases into pre-defined categories with specific labels.[204] Over the past years, the NER technique has been applied to extract medication information, such as drug names, frequency, dosage, adverse drug events, adherence, etc., from free-text documents.[99, 205-208] For example, a conditional random field (CRF) model was used to develop a NER model to detect medication attributes and adverse drug events.[206] Also, bidirectional long short-term memory (LSTM) model was used for named entity recognizing for medication information.[205, 207] More recently, pre-trained deep learning models were widely used for biomedical information extraction.[208, 209] However, to our knowledge, there has not been a well-developed NLP tool to identify a list of *current medications* for a specific disease such as glaucoma and help with medication reconciliation.

The purpose of this study is to develop a NER model for extracting patients' current ophthalmologic medication and adherence from free-text notes for glaucoma patients. Glaucoma is characterized by progressive degeneration of the optic nerve and irreversible visual field loss, and it is the leading cause of irreversible blindness worldwide.[210] The majority of glaucoma patients are treated using medical therapy, and the accuracy of medication documentation is crucial in glaucoma management [211] However, the accuracy of glaucoma medication documentation is unclear. In addition, glaucoma patients' medication non-adherence rate has been reported to vary from 24% to 59%.[212, 213] Therefore, a reliable method to assess glaucoma patients' current ophthalmologic medication and adherence is needed.[214] Finally, the reliability of medication data is important for glaucoma research, such as prediction models for disease progression. In this study, we first manually reviewed patient charts for discrepancies in medication documentation between medication lists and progress notes. Next, we trained and tested a NER model for extracting current medication from progress notes and evaluated its accuracy. Finally, we demonstrated an approach for medication reconciliation using the NER model on small sample progress notes.

METHODS

This study was approved by the Institutional Review Board at Oregon Health and Science University (OHSU). OHSU is a large academic medical center in Portland, Oregon. This study was conducted at Casey Eye Institute, OHSU's ophthalmology department serving all major ophthalmology subspecialties. The department performs over 130,000 outpatient examinations annually and is a major referral center in the Pacific Northwest and nationally. In 2006, OHSU implemented an institution-wide EHR (EpicCare; Epic Systems, Verona, WI) to handle all ambulatory practice management, clinical documentation, order entry, medication prescribing, and billing.

The study contains three phases (1) Explore medication discrepancies between the medication list and the progress note for glaucoma by manually reviewing charts; (2) Develop a NER model to extract patients' current ophthalmologic medication and medication adherence from progress notes for glaucoma patients and (3) Apply the NER model to perform medication reconciliation.

1. Manual Chart Review of Medication Lists and Progress Notes

Progress notes and medication list data from EHR were extracted for 150 randomly selected Casey Eye Institute patients with encounter ICD10 diagnosis codes related to glaucoma from January 23, 2019, to September 28, 2020. The patient's most recent office visit notes were manually reviewed by three independent reviewers. The medications recorded in the narrative notes were abstracted and compared to the EHR medication list at the time of visit. All ophthalmologic medications and over-the-counter (OTC) medications (e.g., artificial tears) were collected. All medications listed in the notes but not on the medication list or vice versa were labeled. Cross-validation among the three reviewers was conducted by using a subset of 20 encounter notes (96.4% agreement).

2. NER Model for Extracting All Ophthalmic Medications

We sampled a dataset with 507 progress notes from office visits at the Casey Eye Institute from January 01, 2019, to December 31, 2019, with encounter ICD10 codes associated with glaucoma. The dataset was constructed by random stratified sampling from all ophthalmology visits according to the department and primary provider name. The documents were manually annotated for nine categories: Drug Name, Route, Frequency, Dosage, Strength, Duration, Adverse Drug Event (ADE), Adherence, and Current Medication Use. All medication names, including generic names, brand names, and abbreviations, were sourced from publicly available online resources and glaucoma specialists. An open-source tool (Doccano; Open source: Doccano; 2018) was used to annotate the documents.[215] Due to the limited number of ADE entities, we discarded this category and kept the other eight entities. **Figure 1** displays an example of the annotation. The annotated dataset was randomly split into 75% for training and 25% for testing. A 10% randomly sampled subset of documents from the training data was used as a validation set for turning the hyperparameters. **Table 1** presents the description of the datasets and annotation statistics.

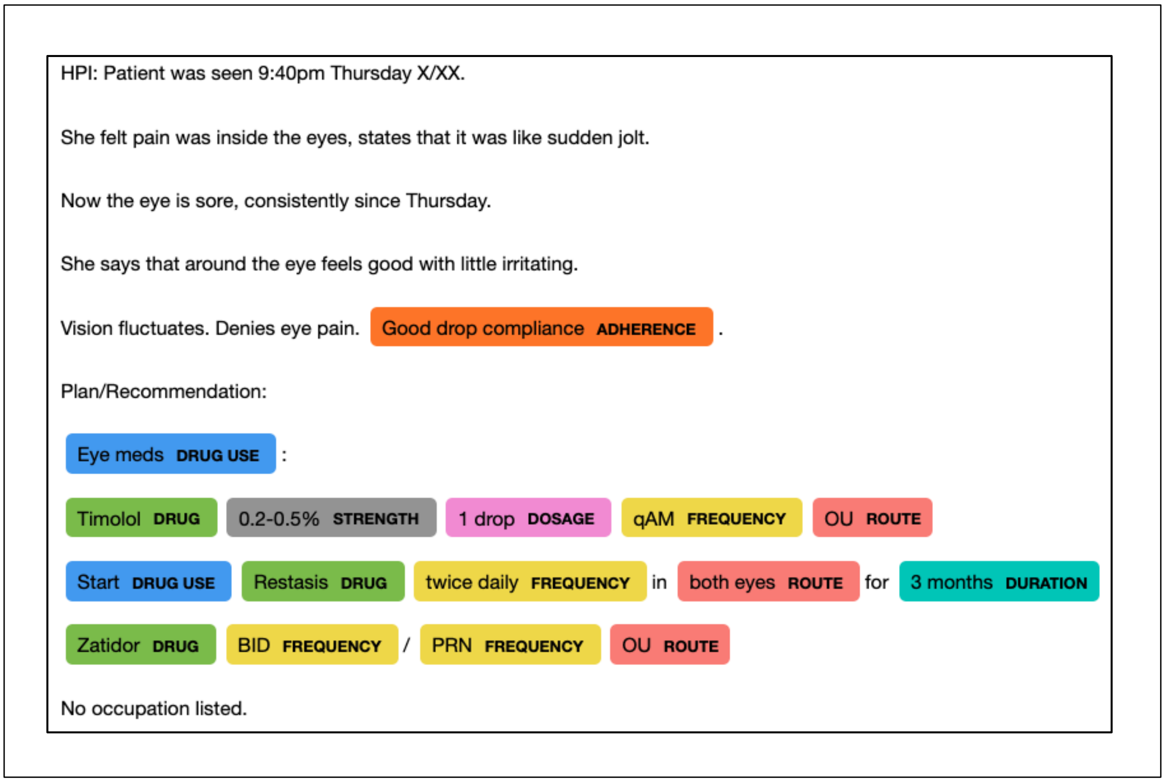

Figure 1. Example of note annotation by an open-source tool. Medication drug name, strength, dosage, frequency, route, duration, current use, and adherence are identified.

We used named entity recognition, a sub-type information extraction technique, to extract medication information and adherence from clinical notes. The NER model was developed in Python 3.7.6 using the spaCy library.[216, 217] The spaCy library is a free open-source library for NLP. The architecture of spaCy's NER model is based on convolutional neural networks which uses a word embedding strategy using sub-word features and "Bloom" embeddings.[218, 219] In this study, the training task contains 200 epochs with experiments with multiple hyperparameter settings. Different learning rates (initial at 1e−2, 1e−3, 2e−3, 1e−4, 2e−4) were tested and adjusted by two optimizers: Adaptive Moment Estimation (Adam) and stochastic gradient descent. We use a decaying dropout rate (0.5 - 0.35; 1e-3) to avoid overfitting. Also, we experimented with different batch compounding sizes and regularization schemes to optimize the model. The results of the

NER model's extraction for the test set were determined by comparing the manually annotated and the NER model's extracted entities. The model performance was evaluated by using F1 score, precision, recall, and the micro-averaged score, which aggregates the contributions of all categories to calculate the average metrics.[114]

Table 1. Distribution of annotated entities and number of progress notes in training and testing datasets.

Named Entities	Train	Test	Total
Drug	2029	505	2534
Frequency	1722	411	2133
Route	1666	371	2037
Dosage	201	40	241
Duration	35	15	50
Strength	168	31	199
Adherence	132	48	180
Current Medication Use	725	185	910
Number of Notes	381	126	507

3. Medication Reconciliation Using NER Model for Current Medications

Finally, we developed a prototype medication reconciliation tool using the optimized NER model. For this purpose, we are focusing only on medications the patient is currently using as documented in progress notes. **Figure 2** demonstrates an example of medication reconciliation using our prototype tool. First, our NER model extracted the patient's medications and "Drug Use" label from the150 sample progress notes which were manually reviewed in phase 1. The "Drug Use" label identified which medications that the patient was currently taking. Next, the current medications were standardized based on RxNorm Ingredient (IN).[220] Finally, the standardized medications were compared to the manually identified medications from phase 1. Both ophthalmologic medications and over-the-counter (OTC) medications (e.g., artificial tears) were included. All medications listed in the notes but not on the medication list or vice versa were flagged.

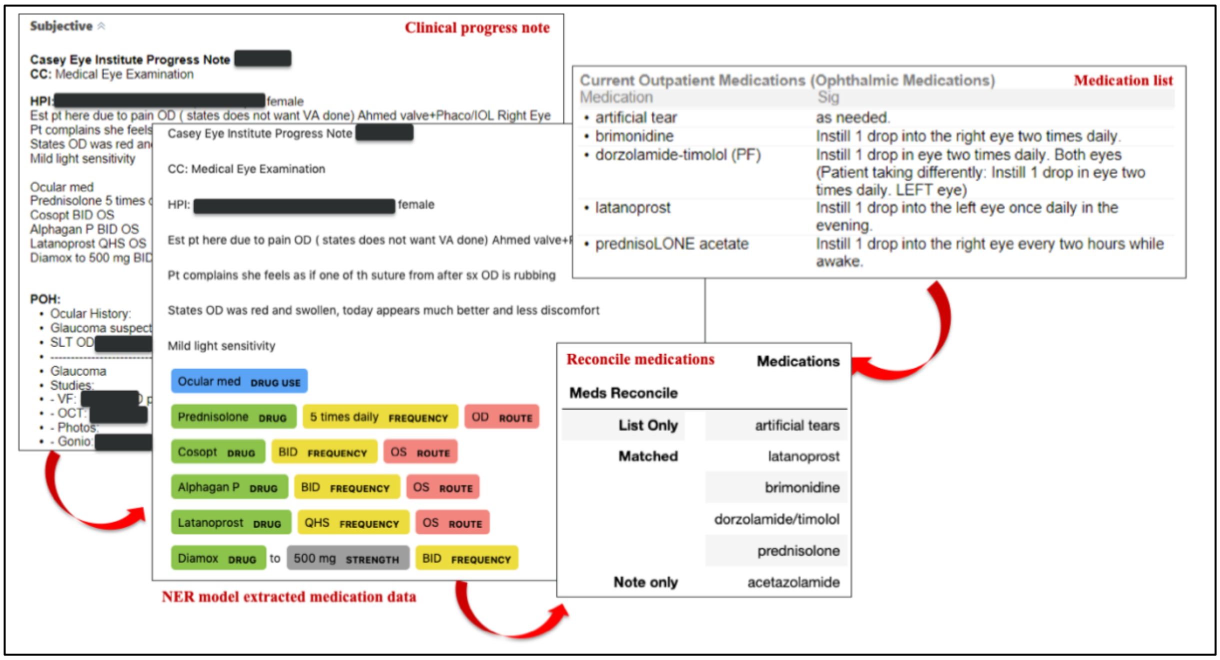

Figure 2. Example of medication reconciliation using the developed NER model

RESULTS

1. Manual Chart Review of Medication Lists and Progress Notes

The randomly sampled 150 patients' notes and medication lists contained a total of 450 medications, including glaucoma eye drops, mydriasis eye drops, antimicrobials, corticosteroids, and OTC medications. Prescription medications were most common (n = 355; 79%), followed by OTC medications (n = 95; 22%). Around 57% of patients had at least one medication mismatch for all categories in their records. However, only 36% of patients had at least one medication mismatch for prescription medications (**Figure 3**). Nearly 66% of medications (n = 298) could be reconciled between the progress notes and medication list. Around 34% (n = 152) of medications are mismatched for various reasons, including medications prescribed by clinicians from different institutions, medications with duplicated prescriptions, medications that were prescribed and entered in the medication list but not recorded in the progress note, and old medications that were not discontinued in the medication list. **Figure 4** displays the distribution of medication mismatches among the two categories in the EHR by location. The most frequent mismatch was found with prescription medications (55%) followed by the OTC medications (45%). The OTC medications were more commonly recorded in the progress notes but not entered into the medication list. In contrast, mismatched prescription medication more often appeared in the medication list but not in the progress notes.

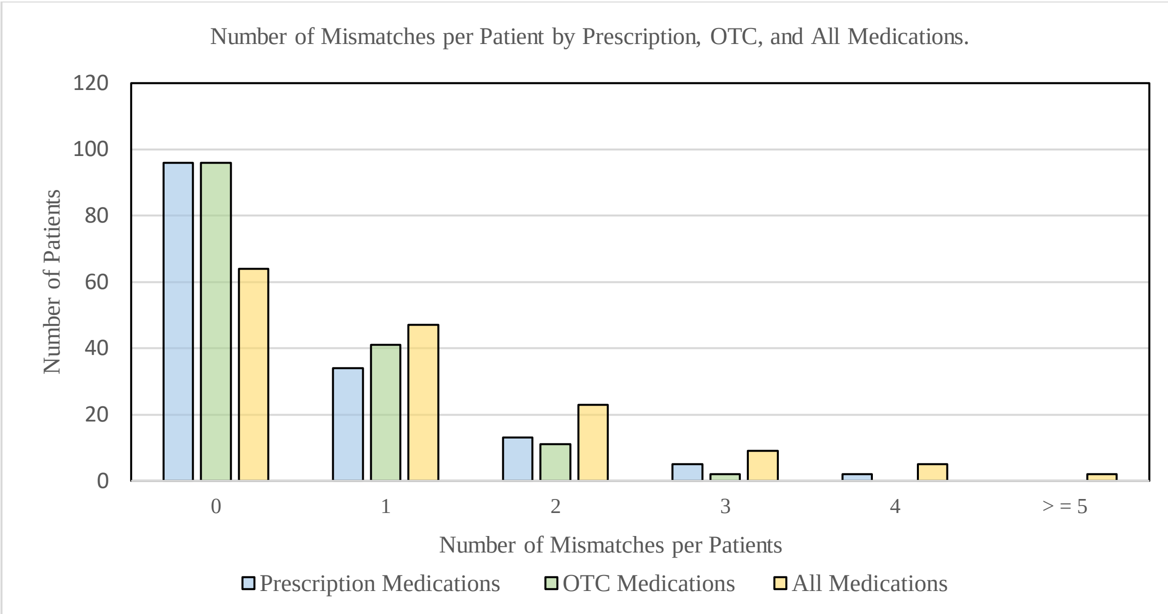

Figure 3. Medication documentation mismatches were stratified based on the number of mismatches that occurred per patient for prescription (blue), OTC (green), and all medications (yellow).

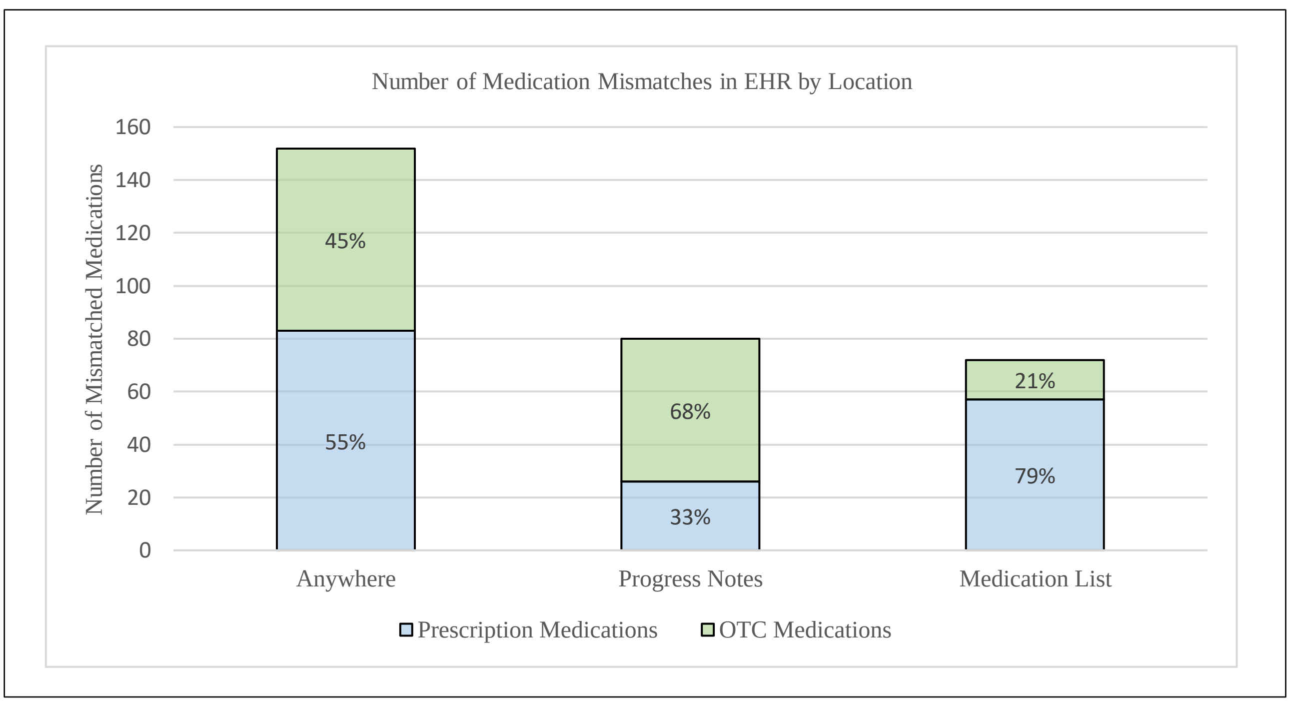

Figure 4. Summary of medication mismatches across 150 patients.

2. NER Model for Extracting Current Ophthalmic Medication

The custom NER model was trained with 381 progress note documents that were manually annotated with eight named entities and then tested on 126 progress notes. **Table 2** presents the overall micro-averaged and per-entity performance for the optimal NER model on test data (126 progress notes). The overall performance of the NER model across all categories was F1 score = 0.955, Precision = 0.951, and Recall = 0.957. Higher performance was observed on medication-related entities: Drug, Name, Route, Frequency, Dosage, and Strength, compared to patient's behavior-related entities: Adherence and Current Medication Use. An error analysis was performed for false negative and positive on Drug Name, Adherence, and Current Medication Use to recognize the source of error predictions. Several causes of errors were identified, such as different wordings for medication adherence, mislabeled current medication use and drug name due to similar sentence structure, eye exams or warm compress mislabeled as drug name, and misclassification when entity information was contained in a short sentence. (**Table 3**).

Table 2. The results of the NER model on the test dataset

Entities	**Performance on Test Data**		
	Precision	**Recall**	**F1-Score**
Drug	0.971	0.971	0.971
Frequency	0.972	0.969	0.970
Route	0.948	0.986	0.966
Dosage	0.987	0.998	0.991
Duration	1.000	0.600	0.749
Strength	0.969	0.997	0.982
Adherence	0.803	0.758	0.779
Current Medication Use	0.899	0.919	0.909
Average (micro)	0.951	0.957	0.955

Table 3. Error analysis from NER predictions related to Drug Name, Current Medication Use, and Adherence labels

Error category	Example	Explanation
Mislabeled Current Medication Use	**“Urgent add on** - Last seen Dr. X on X/X/XXXX”	Unexplained error, “Urgent add on” was labeled as Current Medication Use
	“**encouraged** PFATs at least BID OU - discussed to space at least 5 mins from glaucoma drops”	“Encouraged” was mislabeled as Current Medication Use due to the similar sentence structure
Mislabeled Adherence	“- History *inconsistent drop adherence*” “No eye pain/discomfort but patient admits to *forgetting his drops frequently*.”	There are many different wordings for medication adherence, and Adherence label was not assigned
Mislabeled Drug Name	“Cont **warm compresses** BID ou”	“warm compresses” was mislabeled as Drug Name due to a similar sentence structure
	“Vision has been good. Just **using OTC readers**.”	“Using” was mislabeled as Current Medication Use, and OTC readers was mislabeled as Drug Name due to the similar sentence structure

3. Medication Reconciliation Using the NER Model

The prototype medication reconciliation tool identified 408 current medications from the 150 progress notes that were manually reviewed in phase 1. After standardizing the medications to RxNorm, 14 medications were removed for a final list of 394 medications. Among the 394 medications, there were 379 medications matched with the manually abstracted current medications. The prototype tool achieved a good performance of F1 score = 0.969, Precision = 0.959, and Recall = 0.979.

DISCUSSION

In this study, we explored medication discrepancies in the EHR data and evaluated the performance of a custom NER model's applicability to extract current medication for glaucoma patients. We also used the developed NER model in a proof-of-concept application to perform medication reconciliation in a subset of our patients. The key findings from our study were (1) Medication discrepancies in patient charts were found to be present in a large proportion of office visits; (2) The custom NER model can accurately extract current medication and adherence for glaucoma patients; (3) The NER model can be used to reconcile the medication documentation.

The first key finding is that medication discrepancies were found to be present in a large proportion of office visits. Our study shows that approximately twenty percent of medications prescribed to glaucoma patients had at least one discrepancy between the medication list and the progress note. Overall, more than one-third of patients in this study had at least one medication mismatch between both data sources. These inconsistencies in the EHR medication records may increase the risk of medication errors[221] and affect the reliability of research that relies on this data. These findings are similar to other studies, including a study for microbial keratitis demonstrating 76.9% of medication agreement between progress notes and medication lists[190] and another study for inflammatory bowel disease reporting 78.6% of medications agreement between clinical narrative and medication list.[193] The findings from these studies indicate that the accuracy of the medication list is a common problem. An accurate tool for medication reconciliation of medication lists and further qualitative studies to understand the causes of medication data discrepancies is needed.

The second key finding is that our NER model can accurately identify current medication and adherence from progress notes from outpatient glaucoma visits. In our study, the model reached a micro-averaged F1 score of 0.955 across all categories. The NER model was developed to recognize eight categories from free-text progress notes, including drug names (including generic, brand, and abbreviation names), the route of administration, prescription frequency, the dosage of the drugs, drug strength, duration, medication adherence, and current medication use. The NER model could accurately identify medication-related entities (except duration) but showed lower performance on patient behavior-related entities, such as adherence and current medication use. The difference could be ascribed to the limited number of training cases and the higher variety of wordings. As shown in **Table 1**, there are only 35 annotated duration entities and 132 annotated adherence entities in the training data. In addition, the words and phrases to indicate adherence and current medication use are various, and some of these phrases are located in different sentences than the medications. Nevertheless, the most common error of drug name identification is mislabeling other terms such as “warm compress” or “OTC readers” as a drug name due to similar sentence structure. For example, “warm compress left eye PRN” or “Vision has been good. Just using OTC readers.” In these cases, these mislabeled drugs will easily be filtered out of the results in practice during the conversion to RxNorm names.

Finally, our NER model can be used to reconcile medication documentation. As shown in the phase one study, we can manually abstract the medication records from their progress notes to compare with their medication list. Similarly, the NER model was able to recognize common medications as well as identify text related to current medication use. This is the first study, to our knowledge, to develop NLP models to recognize *current medication* use from free-text progress notes. With the ability to identify the current medication use, we are able to capture the whole picture of current medications for the target patients and reconcile it with their medication list. As previously mentioned, the medication reconciliation between progress notes and medication lists was only reported from 76.9% to 79.6% for three different diseases, including microbial keratitis, inflammatory bowel disease, and glaucoma.[190, 193] And more than one-third of patients had at least one discrepancy for ophthalmic prescription medications.[190] In our study, the NLP tool can correctly identify current medications for glaucoma patients on 150 sample progress notes (F1 score = 0.969). **Figure 2** displays an example of medication reconciliation using the NLP tool. In this prototype tool, we focused on reconciling the drug names since physicians did not always record the other attributes, such as route, frequency, and dosage along with the medications. In future work, we plan to extend this medication reconciliation method to use the information from both narrative progress notes and medication lists to construct a current medication list for glaucoma patients.

Our study has limitations future work may address. First, some of the entities are naturally less frequently recorded in the progress notes that affect the performance of the NER model. For example, text related to drug duration appeared much less frequently than other entities, such as drug name, route, and frequency. Thus, it is challenging to train the model correctly recognize these entities. A similar finding was reported in another study.[208] Second, the model was trained on a set of notes for glaucoma patients from a single institution; it is unclear if the model can be generalizable to other subspecialties within ophthalmology or other healthcare systems. Finally, the application of the custom NER model for medication reconciliation is a proof of concept. We conducted the test of medication reconciliation using the NER model on a limited number of samples. Our intention is to extend and replicate these study methods to different specialties and institutions to increase the generalizability of our model. In the future, the custom model could be incorporated into the EHR system to help with medication reconciliation.

CONCLUSION

Discrepancies in medication documented in the medication and in progress notes were observed in more than one-third of encounters for glaucoma patients. Inaccurate medication lists in the EHR may affect the reliability of the research or clinical decision support using this data. Since physicians often record current medication information in the progress notes this data could be used for medication reconciliation. In this study, we developed an NLP model to accurately identify current medication information from free-text EHR data that can be applied to perform automated medication reconciliation; the performance of the model is similar to the best performing published NLP models for medication extraction studies.[99, 205-208, 222] This has implications in improving the data quality and usefulness for medication data in both research and clinical care.

Chapter 5: Prediction of Multiclass Surgical Outcomes in Glaucoma Using Multimodal Deep Learning with Operation Notes and EHR Data

ABSTRACT

Objective

The objective of this study was to predict long-term multiclass surgical outcomes of trabeculectomy using multimodal models and explore the predictive effect of operative notes. Also, we compared different deep learning techniques to find the best method to extract text information in a multimodal neural network with limited sample size.

Materials and Methods

Three classes of surgical outcomes were defined, including surgical success, elevated intraocular pressure (IOP) and hypotony surgical failure. Operation notes of primary trabeculectomy were collected and mapped to pre-trained word embeddings. The structured features contain three groups: pre-operative EHR features, operation-related features, and early post-operative features. We developed several deep learning models to predict long-term multiclass surgical outcomes using operation notes, structured input features, or combined input features.

Results

At one year, 193 eyes were considered hypotony surgical failure, 183 eyes were considered elevated IOP surgical failure, and 1164 eyes were defined as surgical success. The transformer multimodal neural network had the highest macro AUROC (0.752) and macro F1 score (0.541),

followed by LSTM multimodal neural network (AUROC = 0.725; macro F1 score = 0.491), and the ANN model used structured input features alone (AUROC = 0.709; macro F1 score = 0.476). Prediction models with text data alone showed lower model performance.

Conclusions

Multimodal deep learning models with both structured EHR data and operation notes can be used to predict long-term multiclass surgical outcomes of trabeculectomy. Also, we explored the prediction power of operation notes and the better way to extract text information in a multimodal prediction model. The work has implications for improving post-operative management. In the future, we may incorporate imaging data into models to improve prediction accuracy.

INTRODUCTION

Glaucoma is a group of eye diseases characterized by optic nerve damage and visual field loss.[223] It is the second leading cause of irreversible blindness worldwide and associated with deterioration in the quality of life.[224] Intraocular pressure (IOP) control through medication is the primary intervention for preventing the progression of glaucoma.[15] Surgical intervention may be needed if the maximum dose of glaucoma medications can not halt disease progression. Trabeculectomy remains one of the most common surgical procedures for glaucoma worldwide.[225, 226] Trabeculectomy is a type of filtering surgery to improve eye fluid drainage, which decreases IOP.[227] However, the long-term surgical failure rates of trabeculectomy have been reported as 22% to 35% in multiple studies.[9-11] [9] Surgical failures were defined as patients needing re-operation, loss of light perception vision, showing consistently elevated IOP (> 21 mmHg or less than 20% reduction below baseline) or hypotony (<=5 mmHg) after 3 months of primary surgery.[9] The outcomes of trabeculectomy highly depend on post-operative management within the first 3 months following surgery.[228] Proper IOP control in the early post-operative period is critical for long-term surgical outcomes and is affected by different surgical complications.[63] Scarring is the most common and challenging complication related to post-operative elevated IOP.[64] Physicians may lower the intraocular pressure by releasing the scleral flap sutures using laser suture lysis or using post-operative antifibrotic agents.[65] On the other hand, excessively low IOP post-surgery is caused by conjunctival wound leaks and over-filtration.[66, 228] Physicians may postpone laser suture lysis or decrease the dose and frequency of topical steroid use to reduce the risk of hypotony. Yet, the fluctuation of intraoperative pressure during the early post-operative period and the complexity of surgical recoveries make identifying which patient has a higher risk of long-term surgical failure more difficult. Therefore, there is a strong need for a quantitative model for identifying a patient's risk of surgical failures due to high or low IOP, which could aid the decision-making of post-operative management.

Artificial intelligence has been applied for surgical outcome prediction studies in many specialties such as ophthalmology, neurosurgery, cardiovascular, and renal disease.[123, 229-232] For example, Jeong et al. developed several machine learning models to predict post-operative complications for end-stage renal disease (ESRD) patients who underwent any type of surgery. They concatenated three groups of features that came from different sources: pre-op electronic health records (EHR) features, peri-op features, and text features. Two categories of text features (binary and numeric) were extracted from the pre-anesthetic assessment document using the rule-based natural language processing method. Their best-performing model achieved F1 score of 0.797 with the random forest model.[229] Also, applied machine learning was used to predict the improvement of quality of life after surgery for degenerative cervical myelopathy patients. The random forest model showed the highest AUC of 0.70 with accuracy of 0.77.[231] In ophthalmology, a recently published study described a multimodal machine learning approach with convolutional neural networks (ResNet50) and XGBoost to predict myopic regression after corneal refractive surgery. Their final combined machine learning model showed good performance with AUC of 0.75.[123] These studies show the promise of using artificial intelligence to provide surgical outcome predictions on an individual level to help with clinical decision-making.

For trabeculectomy surgical outcome predictions, several studies have used early post-operative IOP to predict long-term eye pressure control.[70, 233, 234] However, classic statistical regression methods based on intraocular pressure alone or with limited features cannot provide sufficient information to predict long-term surgical outcomes accurately. A previous study demonstrated that machine learning algorithms with pre-operative surgical data could predict the higher risk group of surgical failure with AURCO 0.64 - 0.74, but these prediction models were based on small samples and focused on a binary outcome prediction: surgical success or failure.[232] As aforementioned, long-term surgical failures might correlate with different causes, and the post-operative management for the two different causes is often contrary to each other. Thus, it is more clinically useful to predict a patient's surgical failure risk due to the specific cause.

In addition to IOP, there are other structured EHR features such as demographic data and medication usage, intraoperative features and unstructured EHR data are other important predictors for surgical outcomes after trabeculectomy.[229, 235, 236] While previous prediction models for trabeculectomy mainly rely on patients' pre-existing conditions or early post-operative clinical measurements, there is a wealth of intraoperative information reflecting patients' conditions and performed procedures during the surgery that has potential for predicting outcomes. The operative note is a clinical document that records intraoperative information such as surgical findings, procedures performed, and the patient's condition during the surgery. Operative notes are free-text documents that require natural language processing (NLP) for extracting data to be used in models. With recent advancements in machine learning and the large text corpora available in EHR, NLP has been successfully used to process free-text EHR data, for deep contextualized word representations,[237] information extraction,[92, 238] and text classification.[93, 239] Furthermore, in recent years, several studies combined structured data and unstructured text directly through deep learning techniques to develop prediction models.[240-242] Transformer encoder blocks and long short-term memory (LSTM) layers are commonly used to extract information from the text in these multimodal studies.[241, 243] Deep neural networks are flexible and capable of combining different functional blocks in a single model.[241]

Therefore, an advanced prediction tool combining pre-operative, intraoperative, and post-operative features has the potential to predict the multiclass risks of surgical failures after trabeculectomy. We developed multimodal models to predict long-term multiclass surgical outcomes for patients who underwent trabeculectomy and explored the predictive effect of operative notes on the surgical outcome prediction model. Also, to explore the best method to extract information from the operation notes in a multimodal architecture, we compared the model performance of multimodal neural networks with transformer encoder blocks versus the LSTM. Finally, we identified possible risk factors for surgical failures using an analytic tool.

METHODS

This study adheres to the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board at Oregon Health and Science University (OHSU). OHSU is a large academic medical center in Portland, Oregon. This study was conducted at Casey Eye Institute (CEI), OHSU's ophthalmology department serving all major ophthalmology subspecialties. The department performs over 130,000 outpatient examinations annually and is a major referral center in the Pacific Northwest and nationally. In 2006, OHSU implemented an institution-wide EHR (EpicCare; Epic Systems, Verona, WI) to handle all ambulatory practice management, clinical documentation, order entry, medication prescribing, and billing.

The study included patients aged 18 years or older who underwent primary trabeculectomies (Current Procedural Terminology codes 66170 and 66172) from January 1, 2010 to May 31, 2021 at OHSU Casey Eye Institute. We collected EHR data for the study patients from the enterprise-wide clinical warehouse; it included demographic data, pre-operative systemic health data, pre-operative medication data, pre-operative and post-operative ocular data, operation notes, surgery-related data, and post-operative procedures data. Patients were excluded if they (1) were under 18 years old; (2) did not have a complete operation note; (3) had a trabeculectomy combined with other procedures except phacoemulsification; (4) had less than 1 year of follow up.

Outcome

The primary outcome was 3 classes of long-term surgical outcomes at year one: surgical success, surgical failure due to elevated IOP, and surgical failure due to hypotony. The surgical failure due to elevated IOP was defined as post-operative IOP higher than 21 mmHg or less than 20%

reduction below baseline on 2 consecutive follow-up visits after 3 months or reoperation for glaucoma due to continuous high IOP. Similarly, surgical failure due to hypotony was defined as post-operative IOP of 5mmHg or lower on 2 consecutive follow-up visits after 3 months or operation for glaucoma led by hypotony.

Free-text operative notes

The operative notes are free-text clinical documents that record detailed information about the surgery. There are several specific aspects that need to be recorded in the operation note, including the site or type of incision made, surgical findings, all steps carried out in the procedure, the medications or materials utilized, all complications discovered intraoperatively, and the estimated blood loss. We identified the primary trabeculectomy operation notes for each eye. All notes were preprocessed by removing special characters and punctuation, converting to lowercase, tokenized, and removing custom stop words. Then, the length of each operation note was fixed at 512 tokens to ensure the consistency of the input text feature. The cleaned-up operation notes were mapped to custom 50-dimensional word embeddings. To obtain custom word embeddings, we trained the unsupervised word2vec model[100] with 50-dimensional word embeddings. The word2vec model was trained using all glaucoma-related operations notes in the CEI data warehouse to enlarge the training text corpus. We utilized the Gensim toolkit[244] with the Continuous Bag of Words Model (CBOW) to train the word2vec model in the Python environment.

Structured EHR data

The structured features contain three groups: pre-operative EHR features, operation related features, and early post-operative features. The pre-operative EHR features include demographic values, glaucoma diagnosis, active medication usage before the surgery, chronic systemic diseases, conjunctiva conditions, the best visual acuity measures (logMAR),[245] and the highest IOP recorded in 6 months prior to surgery. The operation related features included surgeon and procedure type (trabeculectomy, trabeculectomy with the previous scar, and trabeculectomy with phacoemulsification). The post-operative features include multiple time points of IOP measures (at day 1, day 2 - day 14, and day 15 - day 30) and the best visual acuity measures (logMAR) within 30 days. All categorical features are converted into binary features. Numeric features were normalized and linear imputation was used to handle the missing data. The final input structured dataset contains 75 features.

Models

We developed three groups of classification models to identify glaucoma patients with high risk of surgical failures after 30 days of surgery: 1) models that use only structured EHR data, 2) models that use only unstructured operative notes, and 3) multimodal models that combine the structured and unstructured data models. These models are implemented using Pytorch[246] in the Python environment.[247] All models are trained with Adam optimizer and used ReLU as the activation function. The batch size is 16 with learning rates initiated with 4e-5. Class weights were used to handle the imbalanced nature of the dataset.

EHR structured data classification model

We trained an artificial neural network (ANN) and a random forest with the structured input features as the baseline models. The neural network consisted two dense layers (Dimension: 75D -> 256D -> 256D -> 64D) with a dropout rate of 0.5 and an output layer (Dimension: 64D -> 3D) with a softmax function to predict the probability of surgical outcomes. The random forest is a bootstrap aggregating-based ensemble method that is popularly used in many clinical prediction models.[248] As previously mentioned, several studies applied random forest models for surgical outcome predictions and have shown good results.[229, 231, 232] Five-fold cross-validation was used to tune the hyperparameters of the random forest model and avoid overfitting.

Text classification model

To investigate the prediction power of the operation notes, we used the preprocessed operation notes to develop the text classification model. We compared two popular text classification models including transformer encoder block and long short-term memory (LSTM) neural networks where they have been previously shown to perform well in text classification tasks.[89] In our study, the operative notes were mapped to the custom word embeddings. These pre-trained word embeddings were input to the transformer encoder blocks and LSTM layer (50 hidden units), then connected to the other two dense layers and the softmax output layer (Dimension: 50D -> 256D -> 64D -> 3D). Batch normalization and the dropout (0.5) layers were used to prevent the gradient vanishing and overfitting. The transformer model consists of 10 attention heads, 12 layers of transformer blocks, and 768 hidden units.

Multimodal model

We developed multimodal neural networks to verify our hypothesis that operation notes can improve the predictive model performance by incorporating structured input features. To explore the better method to process text data in a multimodal predictive model, we compared the transformer encoder blocks with the LSTM models.

The transformer model architecture is demonstrated in **Figure 1(A)**, which combined both structured input features and operation notes. The early-fusion strategy was used to concatenate these two types of data. The operation notes were mapped to the aforementioned custom word embeddings and then passed through transformer encoder blocks (12 layers) with 10 attention heads and 768 hidden units. One more global average pooling layer takes the output of transformer encoder blocks as input and outputs the final text vector. The structured features were input to the model and concatenated with the final text vector and connected with two more dense layers and the final out layer with a softmax function (Dimension: 125D -> 256D -> 48D -> 3D). Batch normalization and the dropout (0.5) layers were used to prevent the gradient vanishing and overfitting.

Figure 1(B) shows the structure of the LSTM-based multimodal neural network. The LSTM layer contains 50 hidden units. One more global average pooling layer was used to capture the information from all hidden layers. The structured features were input into the model and concatenated with the output of the LSTM layers. Two more dense layers and a final output layer with a softmax function were connected (Dimension: 125D -> 256D -> 48D -> 3D). Batch normalization and the dropout (0.5) layers were used as well.

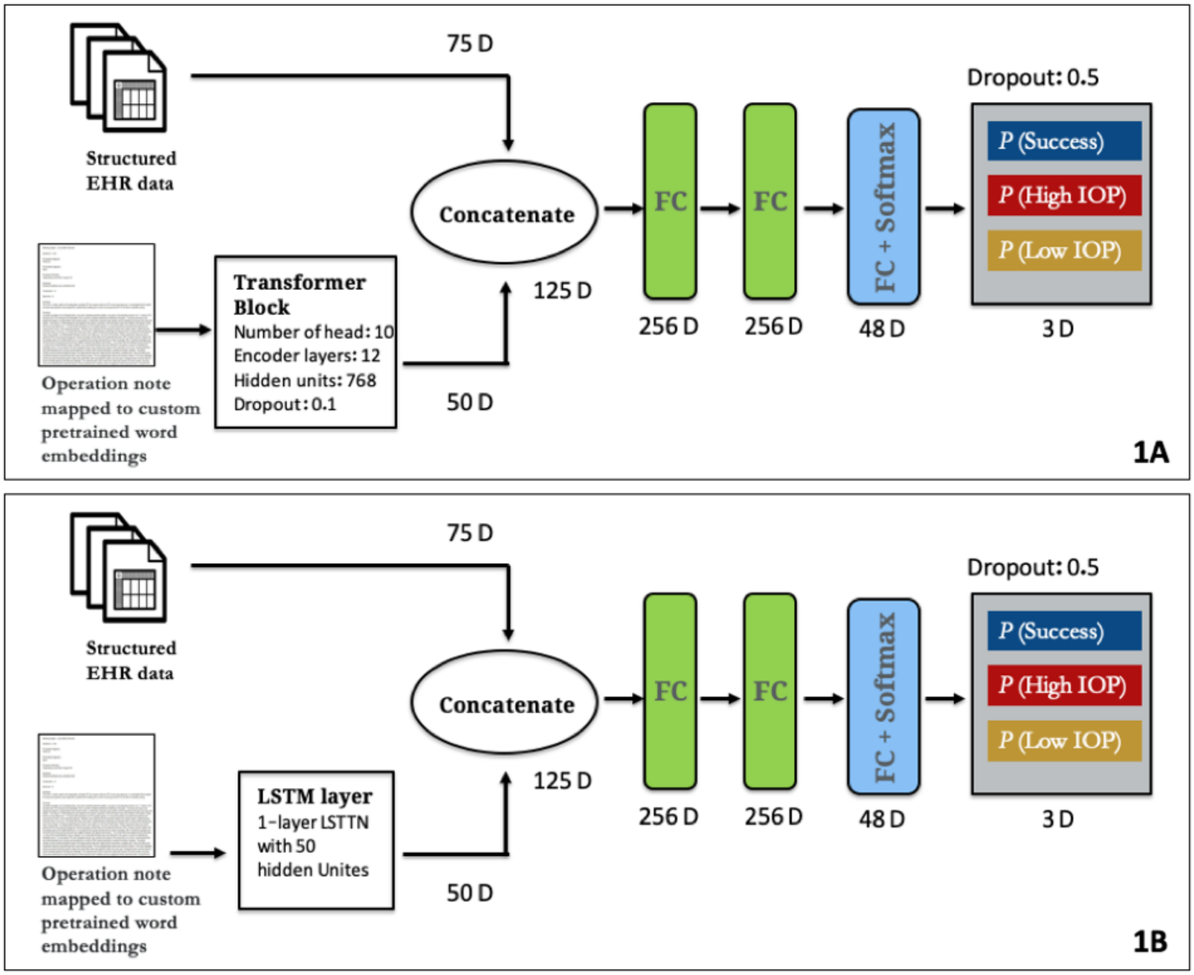

Figure 1. Overview of multimodal neural network architectures used, (A) transformer encoder block model, (B) LSTM-based model. LSTM: long short-term memory

Model evaluation

The dataset was randomly split on the patient level; 70% of the data were used for training, and 10% for validation, 20% for testing. We used the area under the receiver operating characteristic curve (AUROC), precision, recall, and F1 score as the main evaluation metrics on the test dataset. For the multiclass classification task, we calculated macro average and One-vs-Rest (ovr) of AUROC and calculated macro average and per class of precision, recall, and F1 score. We also performed model interpretations for the random forest model and the text classification model using SHAP (SHapley Additive exPlanations) toolkit[249] in Python to explore the important structured input features and the important words in the operation notes related to surgical outcome prediction.

RESULTS

Table 1 shows the descriptive characteristics of the patients in the study cohort. A total of 1540 eyes from 1326 patients who underwent trabeculectomy between January 2010 and May 2021 met the inclusion criteria. At one year, 193 (13%) eyes were defined as surgical failure due to hypotony, 183 (12%) eyes were defined as surgical failure due to elevated IOP, and 1164 (75%) eyes were defined as surgical success. The patient demographic showed that the majority of patients were Caucasians (86%) and females (57%), with a diagnosis of primary open-angle glaucoma (72%). The patient's average age was 64 years and most patients used Medicaid (49%) and commercial insurance (42%). The mean IOP before the surgery was near 21 mmHg and the mean logMAR visual acuity was near 0.25, which is about 20/36 as a Snellen equivalent.

Table 1. **Baseline Demographic and Clinical Characteristics**

	Total (1540 eyes)
Age, years	
Mean (SD)	63.55 (15.69)
Sex	
Male	661 (43%)
Female	879 (57%)
Race	
White	1329 (86%)
Non-White Hispanics	56 (4%)
Black	50 (3%)
Asians	60 (4%)
Others	45 (3%)
Clinical Characteristics	
Intraocular Pressure (mmHg)	21.07 (8.6)
Visual acuity logMAR	0.25 (0.41)
Number of Glaucoma Medications	2.59 (1.41)
Surgical Outcomes	
Success	1164 (75%)
Hypotony surgical failure	193 (13%)
Elevated IOP surgical failure	183 (12%)
Healthcare Insurance	
Medicaid	62 (4%)
Medicare	756 (49%)
Commercial insurance	641 (42%)
Unknown	81 (5%)

Macro receiver operating characteristic curves on the test dataset for the ANN with structured input features, text classification model with operation notes, and transformer and LSTM multimodal neural networks with both data are presented in **Figure 2**. Also, **Table 2** shows other evaluation metrics including the precision, recall, F1 score (macro average) and AUROC (macro average) for the three models. The transformer multimodal neural network had the highest macro AUROC (0.752) and macro F1 score (0.541), followed by LSTM multimodal neural network (AUROC = 0.725; macro F1 score = 0.491), the random forest model (AUROC = 0.712; macro

F1 score = 0.486), and the ANN model (AUROC = 0.709; macro F1 score = 0.476). The text classification models showed lower model performance for both AUROC and F1 score.

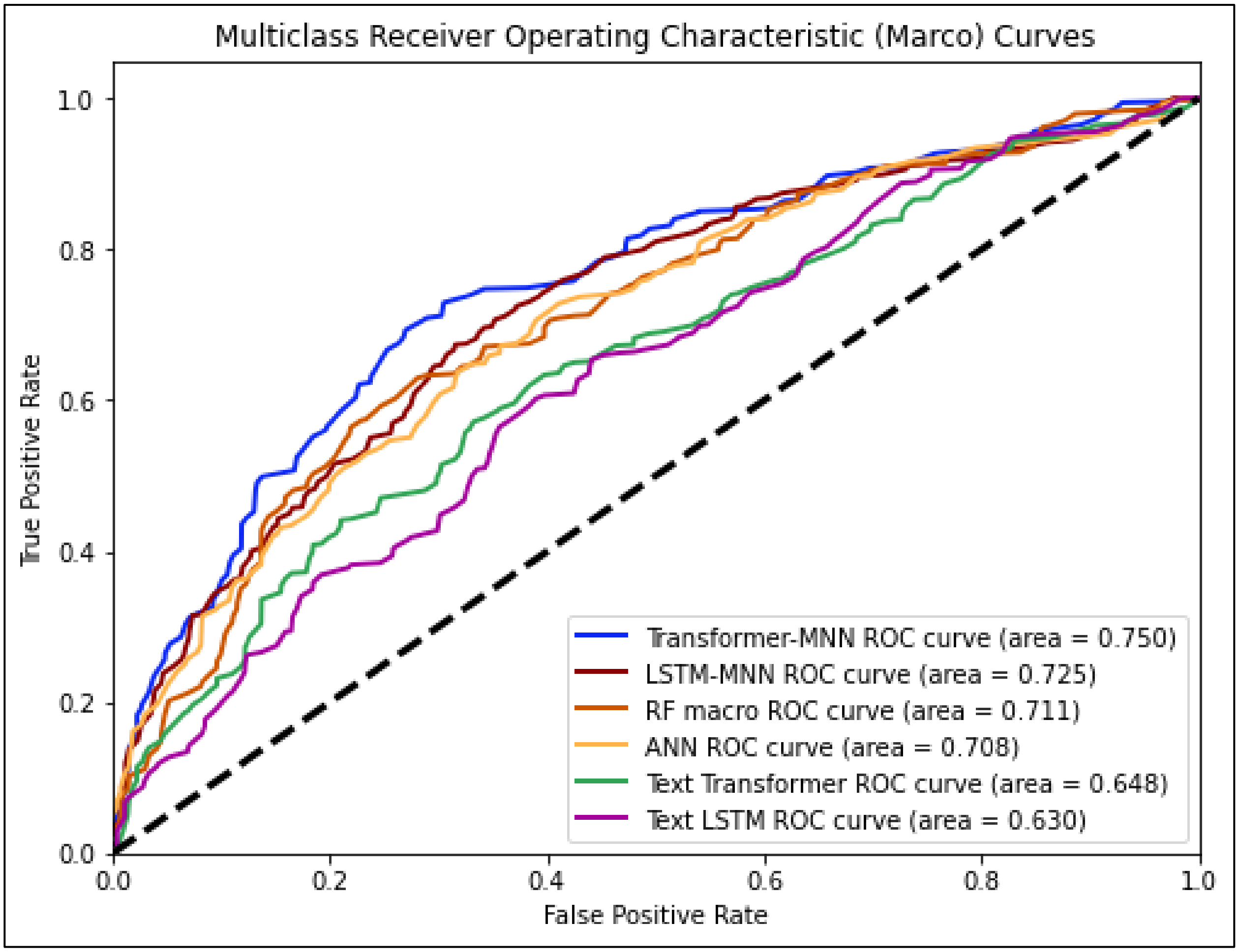

Figure 2. Multiclass receiver operating characteristic curves on the test dataset for the ANN model, text classification model, LSTM multimodal network, and transformer multimodal network. LSTM: long short-term memory; MNN: Multimodal neural network

Table 2. Comparison of model performance using macro average metrics

Surgical Outcomes Predictions (Macro)				
	AUC score	Precision	Recall	F1 score
Transformer-MNN	0.750	0.529	0.622	0.541
LSTM-MNN	0.725	0.477	0.527	0.491
ANN	0.708	0.492	0.470	0.476
Random forest	0.712	0.495	0.487	0.486
Text Transformer	0.648	0.461	0.410	0.414
Text LSTM	0.630	0.388	0.409	0.391

MNN: Multimodal Neural Network; LSTM: long short-term memory

Receiver operating characteristic curves of each class and macro average on the test dataset for the transformer multimodal neural network were shown in **Figure 3**. In **Table 3**, the evaluation metrics for each class are depicted for the transformer multimodal neural network. The model shows the highest AUROC (0.787, ovr) for the elevated IOP surgical failure group, followed by the hypotony surgical failure group (0.756, ovr), and the surgical success group (0.707, ovr). In addition, the model had the highest recall (0.691) for hypotony surgical failure, while the surgical success group had the highest precision (0.876) and F1 score (0.707). Overall, the model showed a better discriminate ability to predict the elevated IOP surgical failure (AUOC: 0.787; F1-score: 0.482) than hypotony surgical failure (AUROC: 0.756; F1-score: 0.409).

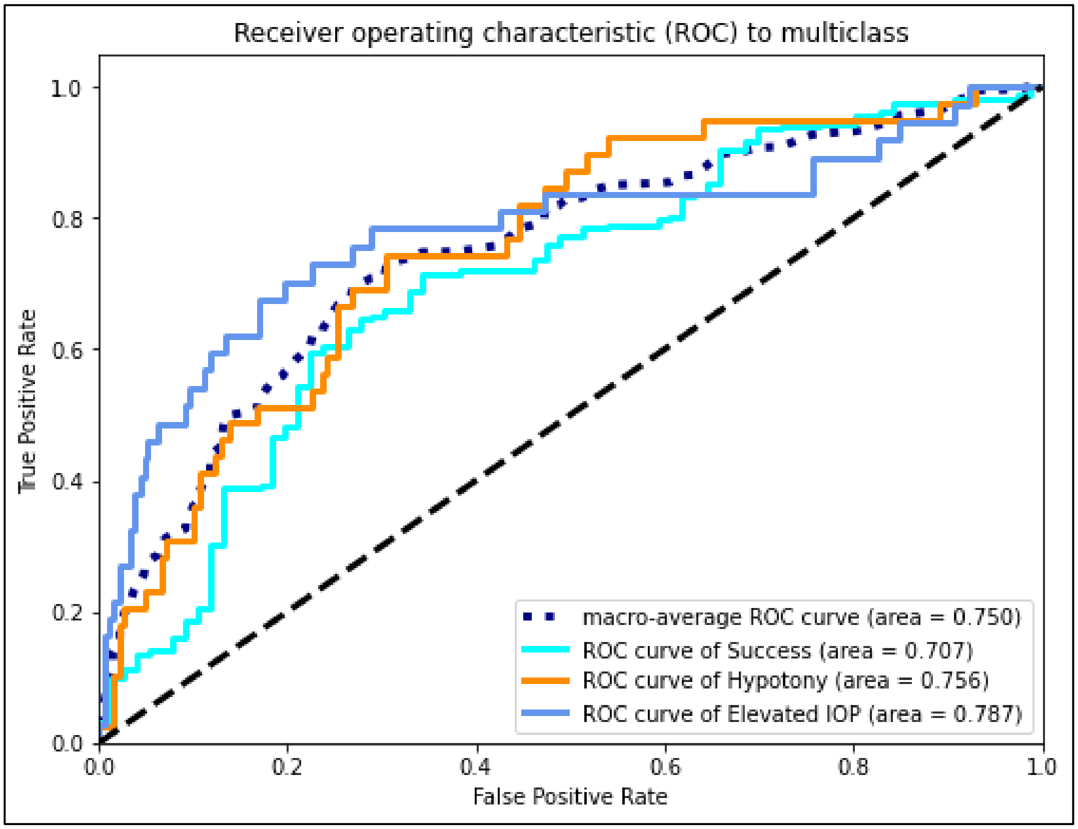

Figure 3. Receiver operating characteristic curves of each class and macro average on the test dataset for the transformer multimodal neural network

Table 3. Performance metrics for each class for the multimodal neural networks

Transformer Multimodal Neural Network				
	AUC score	Precision	Recall	F1 score
Success	0.707	0.876	0.623	0.734
Low IOP	0.756	0.294	0.691	0.409
High IOP	0.787	0.437	0.544	0.482
Macro average	0.750	0.529	0.622	0.541
LSTM Multimodal Neural Network				
	AUC score	Precision	Recall	F1 score
Success	0.654	0.783	0.649	0.709
Low IOP	0.710	0.314	0.458	0.373
High IOP	0.797	0.333	0.474	0.391
Macro average	0.725	0.477	0.527	0.491

LSTM: long short-term memory

In addition, we are also interested in which structured input features are more important to help with the model prediction. We applied the SHAP toolkit to the random forest model, which showed higher AURCO and F1 scores than the ANN model. We calculate the global importance of each feature contributing to the model prediction for all outcome classes. **Figure 4** shows the top 15 most important structured features related to all outcome classes prediction. Important features include several IOP measures before and after the surgery, use or non-use of prostaglandin eye drops, with or without previous cataract surgery, and patient's age. Also, we used the SHAP summary plots to explore the feature effects on each outcome class. **Figure 5** and **Figure 6** present the top 14 most important features and their effects on elevated IOP surgical failure and hypotony surgical failure, respectively. Several features increased the predicted risk of elevated surgical failure, including males, using angiotensin-converting enzyme (ACE) inhibitors, prior cataract surgery, and higher pre and post-operative IOP. On the other hand, patients with lower pre and post-operative IOP, used prostaglandin eye drops, and had prior cataract surgery had a higher predicted risk of hypotony surgical failure. Lastly, **Figure 7** shows the top 14 most important features and their effects on surgical success. Post-operative IOP from week 2 to week 4 was still the most important feature, but its dependence plot (**Figure 8A**) shows no clear linear relationship between the IOP value and the probability of surgical success. Many instances with the same IOP value show both positive and negative impacts on the probability of surgical success. Besides, **Figure 8B** shows the clear linear relationship between the patient's age and the probability of surgical success. In this plot, younger patients had a lower surgical success rate.

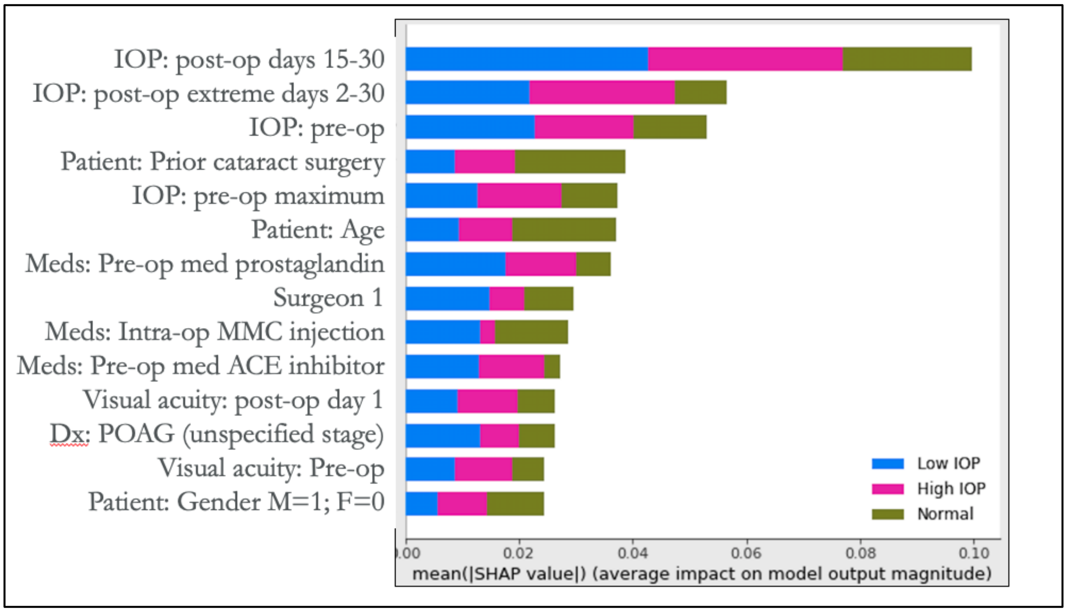

Figure 4. Important features in the random forest model as determined by average SHAP values. The Blue label shows the features' effect on hypotony surgical failure. The red label indicates the features' effect on elevated surgical failure. And the yellow label means the features' effect on surgical success.

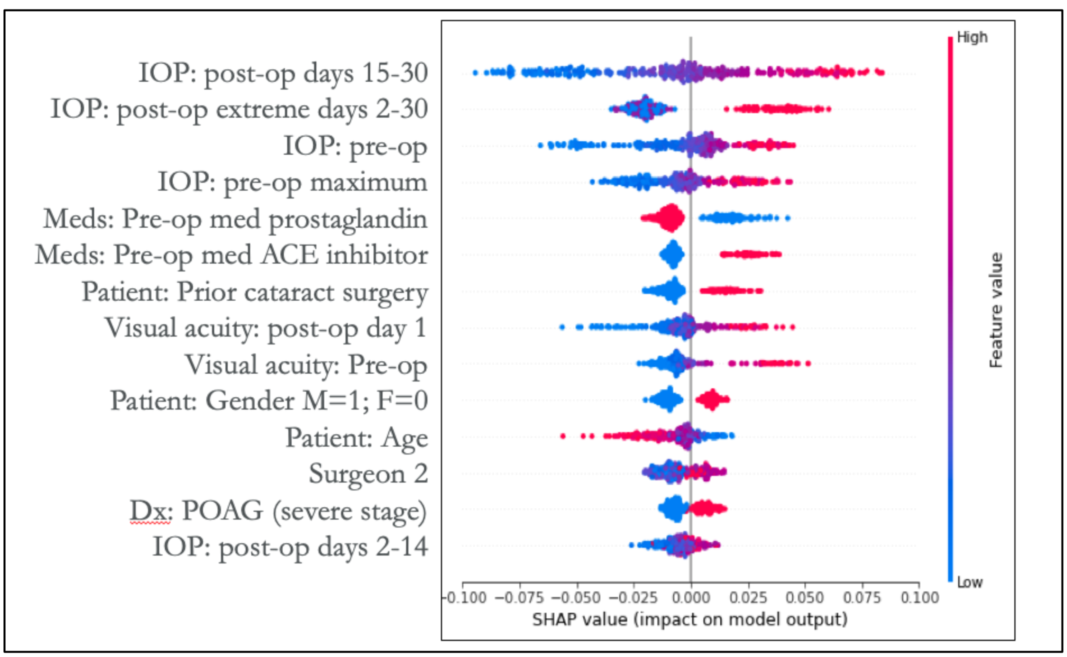

Figure 5. SHAP summary plot of important features and their impact elevated IOP surgical failure in the random forest model. Each point on the summary plot is a SHAP value of a feature for a data point. The Y-axis is determined by the feature and ordered according to its importance. The X-axis shows the SHAP value, and the color represents the value of the feature. Overlapping points are jittered in the y-axis direction.

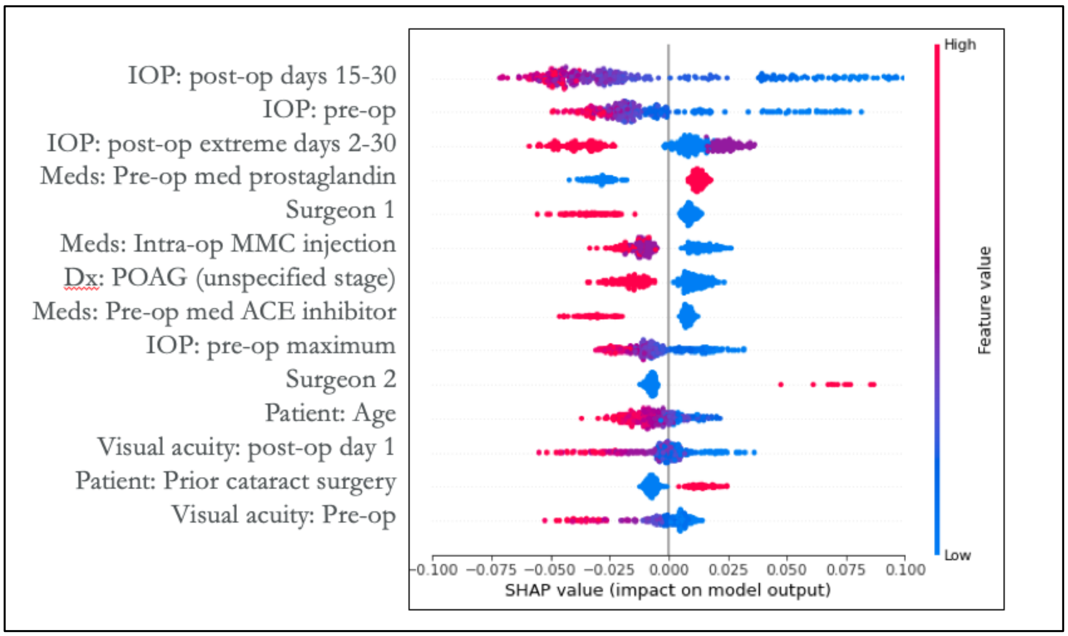

Figure 6. SHAP summary plot of important features and their impact hypotony surgical failure in the random forest model.

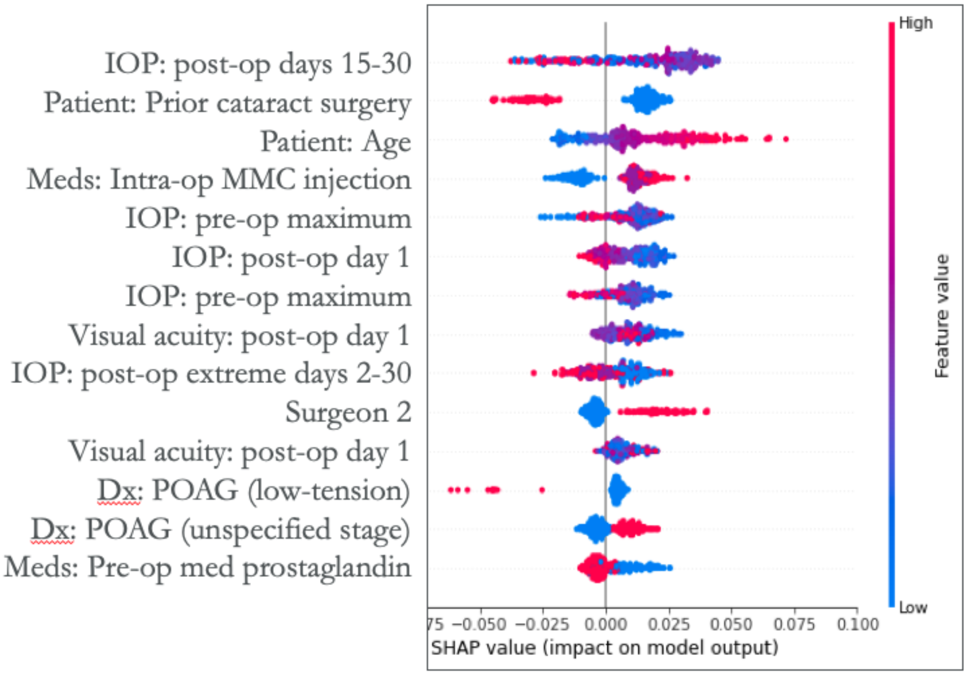

Figure 7. SHAP summary plot of important features and their impact on surgical success in the random forest model.

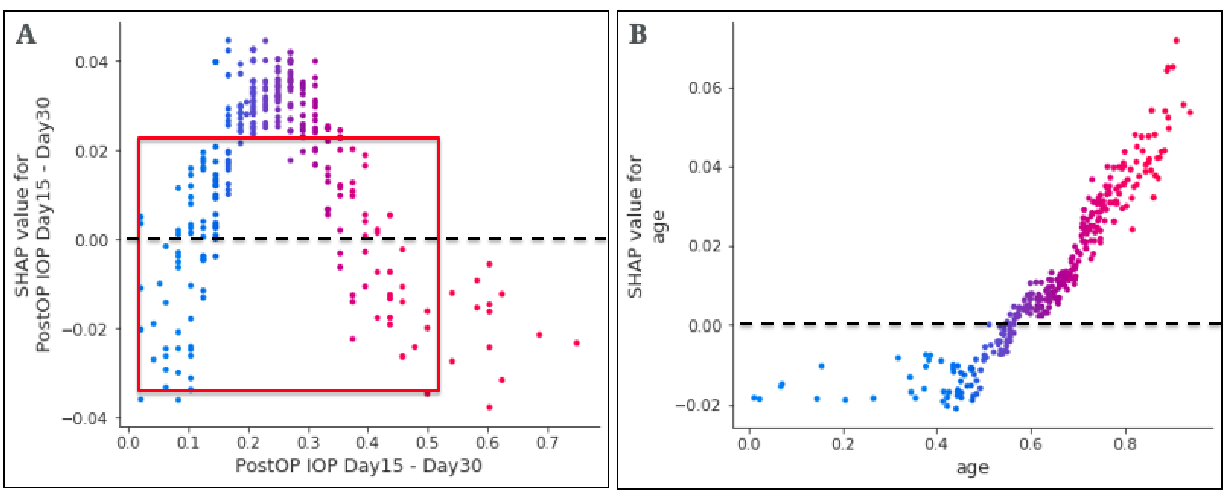

Figure 8. Dependence plot of post-operative IOP from week 2 to week 4 (A) and patient's age (B). Figure A shows no clear linear relationship between the IOP value and the probability of surgical success. In the age dependent plot, there is a clear linear relationship.

DISCUSSION

In this study, we evaluated the applicability of deep learning models to predict multiclass surgical failures for glaucoma patients who underwent trabeculectomies. Also, we explored the prediction ability of operation notes in the outcome prediction model. The key findings from our study were (1) Using multiclass surgical outcome prediction for glaucoma patients provides more information to clinicians; (2) Operative notes can provide important predictive information to help with the model performance; (3) Transformer-based multimodal neural network model outperformed the baseline model and yielded more accurate predictions by incorporating operation notes; (4) Model explainer tool can be used to identify possible risk factors for surgical failures.

Our first key finding—that our model successfully predicted a multiclass outcome—is crucial for trabeculectomy post-operative care. In previous studies, the developed prediction models for long-term surgical outcomes for glaucoma patients were a binary outcome of success or failure. In our study, we have shown that predicting specific causes of surgical failure can provide more practical assistance for clinical decision-making. As previously mentioned, the outcome of trabeculectomy is mainly determined by early post-operative care and IOP control.[63, 228] The post-operative treatment plan varies according to the IOP and post-operative complications. For patients with high risk of scarring and long-term elevated IOP surgical failure, physicians might remove releasable sutures, perform laser suture-lysis, increase steroids dosage, and use antifibrotic agents (5 Fluorouracil). On the contrary, for patients with high risk of chronic hypotony, lower steroid dosage use or delayed suture-lysis should be considered. In our study, with the reasonable threshold setting, the multimodal neural network achieved a recall score of 0.691 for the hypotony surgical failure cohort and 0.543 for the elevated IOP surgical failure cohort. This result indicates the model could correctly identify more than half of patients who actually progressed to surgical failures within one year. With this result, the prediction model might be able to inform physicians of the potential risk for specific surgical failures so that the patients could receive more appropriate therapy.

Our second key finding was that operative notes had predictive power for trabeculectomy outcomes. To the best of our knowledge, this work is the first study investigating the usage of free-text operation notes in predictive models for multiclass surgical outcomes. Our model with structured EHR data included patients' information, pre-operative ocular measures, and post-op ocular measures but lacked intraoperative information. To address this gap, we developed multimodal neural network models to extract intraoperative information from operation notes and incorporate it with structured input features. The result of the study demonstrates the multimodal neural network that used both structured data and unstructured operative notes performed better than the model using structured data alone. This finding indicates operation notes can complement the intraoperative information gap in predictive modeling which leads to performance improvement. This concept can be applied to other specialties to improve the performance of surgical outcome prediction models.

We further investigated critical features from the operative notes using SHapley Additive exPlanations (SHAP) toolkit for transformer text classification model. Several words/phrases were identified including mitomycin c usage, surgeons, and conjunctiva conditions. From clinical perspective, usage and dosage mitomycin c could be an important predictor of the surgical outcome predictions.[250] Conjunctiva conditions were also important, which makes sense since patients' conjunctiva conditions are correlated to wound recoveries.[251] However, there were many words that contributed to the prediction having less clinical meaning. This could be ascribed to the frequency of words and the limited number of training cases. The investigation of keywords in the predictive models might be helpful to validate the trustworthiness of the model and provide possible risk factors for glaucoma surgery for future studies.

Our third key finding was the transformer-based multimodal neural network model outperformed both types of models using structured and unstructured data only. We have used the word2vec based custom word embeddings to preprocess the free-text operation notes. Previous studies have shown that transformer blocks with pre-trained embedding give promising results compared to classical deep learning models for text classification tasks.[89, 243, 252] However, the word embedding models in these studies were trained using large datasets with over 1 million lines of sentences. In our study, we used a high correlation but smaller dataset (around 150 thousand lines of sentences) to develop the custom pre-trained word embeddings with 50 dimensions. To explore the better method to extract information from the operation notes with a custom small text corpus pre-trained word embedding in a multimodal architecture, we compared the transformer encoder blocks with the LSTM layers. The transformer-based multimodal neural network model and text classification model all showed better performance for both AUROC and F1 scores. This is likely due to the sequential nature of LSTM for the long text, where the length of input text data in our model is 512 tokens. In addition, data scarcity is a common problem in the real world and affects model performance. Our study shows using the transformer encoder blocks with custom word embeddings can provide better results with scarce data. This has implications for other classification tasks with limited sample sizes.

The final key finding was the model explainer tool can be used to identify possible risk factors for surgical failures. We used the SHAP toolkit to investigate the importance of structure input features contributing to the model prediction. We calculated the global importance of each feature and its impact on each prediction output class. These results are shown in **Figure 4** to **Figure 8**. As expected, higher IOP before and after the surgery increased the predicted risk of elevated surgical failures and vice versa. The most important feature for all output classes is post-operative IOP from week 2 to week 4, which is similar to the previous studies' findings. Two studies suggested that post-operative IOP at 1 month had a higher predictive value for long-term surgical outcomes. [69, 70] To be noticed, **Figure 8A** shows no clear linear relationship between the IOP value and the probability of surgical success. Many instances with the same IOP value show both positive and negative impacts on the probability of surgical success. This finding indicates that using IOP measures alone to predict long-term surgical outcomes is difficult and unreliable. Also, patients who had prior cataract surgery have a higher predicted risk of surgical failures for both elevated IOP and hypotony. Similar findings were reported in Takihara's study,[253] where they found patients who underwent trabeculectomy in pseudophakic eyes after phacoemulsification had less surgical successful rate compared with that in phakic eyes. Younger patients and male patients are also important for surgical outcome predictions, and similar results were presented in Okimoto's study[233] and Mathew's study.[254] The similarity between these important features and previous studies' findings can be helpful in proving the reliability of the prediction model. Furthermore, several important features related to surgical outcome predictions haven't been clearly investigated or controversial. For example, in the prediction model, the use of prostaglandin eye drops showed a protective effect against elevated IOP surgical failure but increased the predicted risk of hypotony surgical failure. On the contrary, the use of ACE inhibitors shows the opposite effect. Yet, the effect of prostaglandin eye drops and ACE inhibitors on surgical outcomes hasn't been clearly studied. Therefore, these findings can be a direction for future study.

Despite these innovative findings, there are several limitations in our study. Firstly, an important challenge of this work is the naturally inherent imbalance dataset in our study cohort. In our study, less than 25% of glaucoma patients were considered surgical failures at year one, which increases the difficulties of training the prediction model. Predictive models were trained to optimize the loss for overall accuracy, often resulting in making predictions for the major class. To address this problem, class weights and optimal prediction thresholds were used to handle the imbalanced nature of the dataset. Secondly, our model did not include ocular imaging data, as this information was not available for most of the patients in the study cohort. Integrating imaging data into the multimodality predictive model might improve the model performance and is a future direction for this research. Finally, although we collected one of the largest clinical observation datasets with detailed information about trabeculectomy outcomes, the sample size is still limited due to the nature of the study in a single institute. Future studies will ideally include data from multiple institutions.

CONCLUSION

In this study, we have developed a multimodal prediction model for multiclass surgical outcomes of trabeculectomy using both structured EHR data and free-text operation notes to address the need for effective post-operative management. Also, we have investigated the prediction power of operation notes and explored the better method to extract information from text in a multimodal prediction model. We believe that our work will be helpful for clinical decision-making for post-operative care and the study methods can be extended to other specialties to improve surgical outcome predictions. In the future, we may be able to incorporate imaging data as well as multi-site data to improve the model performance.

Chapter 6: DISCUSSION

In this work, we have explored the application of artificial intelligence using EHR data to ocular disease, specifically glaucoma. In the first study, we performed a literature review of applications of AI using ophthalmic EHR data. We found applications for multiple purposes, including diagnosis improvement, disease progression, and risk assessment. In the second study, we conducted a data quality assessment of the medication list for glaucoma patients, finding inconsistencies between medications recorded in progress notes and medication lists. Consequently, we developed an NLP-based algorithm to extract medication information and adherence from the clinical progress notes and extended it to a prototype assist tool for automatic medication reconciliation. In the third study, we developed several multiclass prediction models to predict if a glaucoma patient will have surgical failure due to high or low interocular pressure. In addition, several possible risk factors were identified using a model interpretation analytic tool. There are three key findings derived from this work: (1) Quality of EHR data and data accessibility may affect the secondary use of EHR for AI models; (2) NLP techniques can be used to improve the data quality of medication records and help with medication reconciliation; (3) a multimodal neural network that combined both structured and unstructured EHR data can be used to predict multiclass surgical outcomes for glaucoma patients. Detailed discussions of each aim can be found in Chapters 3 – 5; in the following sections, we will discuss additional findings for each aim.

Aim 1: Identify the possible issues of the secondary use of EHR data in AI applications in ophthalmology. A systematic literature review was conducted to explore the current applications of AI using EHR data in ophthalmology. Our goal was to better understand the potential problems of secondary use of EHR data with AI techniques in ophthalmology, explore the AI techniques used, and the model performance of these works. There were 13 studies that met our inclusion criteria focusing on glaucoma, cataracts, age-related macular degeneration, and diabetic retinopathy. Also, supervised machine learning was the major technique used in these studies, especially the random forest model. Besides this literature review study, we also explored other AI application studies for glaucoma management. Overall, two key findings arose from this study: (1) There are limited studies using AI techniques for disease outcome or progression predictions; (2) Data quality is the major issue for secondary use of EHR data via AI applications.

AI techniques have been broadly applied to ocular diseases for a decade. However, most of the previously published studies in the field of AI application for glaucoma used imaging data and functional data to improve diagnostic accuracy or characterize patterns of glaucoma progression.[255] There are about 80% of the studies focusing on improving diagnosis, but only a limited number of works among these studies focusing on disease outcome or progression predictions with EHR data. In our literature review paper, there were only two studies that used machine learning techniques with EHR data for glaucoma prediction.[116, 118] In the 2 years following this study, there are more studies using EHR data with AI techniques to help with glaucoma management.[92, 232, 255] Compared to the clinical data that was collected for research purposes, EHR data from clinical practice are less structured and are prone to poor data quality.

Aim 2: Approach the challenges of secondary use of EHR data, especially exploring the accuracy of medication lists for glaucoma patients. We performed the EHR medication data quality assessment for glaucoma patients at CEI, checking the medication list accuracy through a manual chart review. To address the inaccurate medication recorded issues, we developed a NER model to extract active medication information and medication adherence from progress notes. The developed model was used to help with medication reconciliation in a prototype tool.

We also conducted additional data quality assessments for other sources of EHR data for glaucoma patients. Multiple issues were noticed, such as incorrect data, duplicated data, ambiguous data, inconsistent data, incomplete data, and data assessing issues. For example, several patients underwent trabeculectomy surgery (CPT66170), but their billing codes were recorded as shunt surgery (CPT66180) and vice versa. These errors could be led by treatment plan change - patients were scheduled for shunt surgery, but eventually, their condition might not be eligible for the procedure; therefore, they received trabeculectomy instead. Duplicated EHR data is another common issue. Many patients' operation notes are duplicated in the EHR system. Ambiguous data issues commonly happen for some description clinical exams, such as visual acuity. Notations or characters in visual acuity exams may not have a consistent definition, which makes it difficult to convert them to quantitative measures (LogMAR). Patients lost to follow-up are another frequent data quality issue. Unlike prospective data collected from a clinical trial study, retrospective EHR data from clinical practice are more commonly incomplete. Furthermore, data accessibility is another major issue for the secondary use of EHR data. For instance, we noticed that accessing visual field data and OCT data in the EHR for ocular diseases is extremely difficult, which creates barriers for AI applications.

To address the accuracy issues of the medication list, we manually reviewed 150 progress notes and found that 36% of patients had at least one medication mismatch for prescription medications. In an extension study, we explored these mismatches in detail for specific groups of medications. Our results demonstrate that 84% of glaucoma medications were accurately documented between the progress note and the medication list. Nonglaucoma ophthalmic medications (57.8%) and over-the-counter (OTC) medications (28.4%) show much lower accuracy.[256] These discrepancies raise concerns about the quality of EHR data used in patient care, research, and billing. There are several possible reasons for these discrepancies. First, physicians may not regularly manually update the medication list, especially OTC medications, which are usually not prescribed by physicians. Also, physicians used to document in paper charts may prefer documenting in progress notes rather than medication lists. The inconsistent medication records may inadvertently lead to unsafe treatment plans and increase the rate of adverse events.[257] These issues also limited the secondary use of medication data. Therefore, we developed a NER model to extract active medication from the progress notes and demonstrated a prototype medication reconciliation tool.

Our model can accurately identify active medication and adherence from progress notes. The model achieved a 0.971 F1 score to identify the medication names and showed similar results for medication frequency, route, and dosage. The performance is close to or better than the best-performing published NLP models for medication extraction studies. In addition, our model can be used to extract "current medication use" labels, and we added "not current medication use" labels in our extension study. These labels can be used to identify if the medications are active use during the visit. As shown in the medication reconciliation tool, the tool can capture and present the mismatched medications between the medication list and progress notes. Our tool shows a promising method to automate medication reconciliation but still needs further validation to be implanted in a clinical setting.

EHR data quality improvement is an essential direction for future studies. Several possible methods may be helpful to improve the data quality of EHR, such as having better standardization of data, enhancing documentation practices, increasing structured data fields, and improving the EHR user interfaces to encourage more accurate and thorough documentation. In addition, most AI studies using EHR data in ophthalmology were performed at a single institution. Thus, the developed algorithms may not be generalizable to other healthcare systems. One possible strategy for extending studies to multiple institutions is federated learning: sharing the model instead of sharing the EHR data. This is done by training the same model at each institution with its own EHR data and sharing the trained model among institutions to optimize the model. Overall, applications of AI techniques are a promising direction to help with clinical care, but there are still several challenges that need to be addressed, especially the quality issues of EHR data.

Aim 3: Predict multiclass long-term surgical outcomes for patients who underwent trabeculectomy and identify possible risk factors. We developed several prediction models using structured input features, free-text notes, or both inputs to predict multiclass long-term surgical outcomes. The multimodal neural network with transformer encoder blocks shows the best performance with both structured EHR data and operation notes. Furthermore, we used the SHAP toolkit to explore the most important structured features in the random forest model and their effects related to different prediction classes.

In this study, we hypothesized that operation notes containing useful intraoperative information could help to fill the information gaps in the prediction model using structured input features only. The operative notes are free-text clinical documents that record detailed information about the surgery that are not recorded elsewhere. Examples of potentially helpful intraoperative information might include mitomycin c dosage and injection rate, conjunctiva conditions, the volume of blood loss, and specific trabeculectomy techniques. Mitomycin c is an anti-fibrotic agent usually used during surgery to prevent postoperative scarring leading to bleb failure. The dosage of mitomycin c and the injection rate might be important predictors of surgical outcomes.

We developed several text classification models, including transformer-based neural network, LSTM, and 1-dimensional convolutional neural network (CNN), to evaluate the prediction power of operation notes alone and determine the best information extraction method. The transformer-based neural network shows the best result. The possible reason can be the limitation of LSTM and CNN for the long text since we used a fixed input length of 512 words. Although the text classification model shows a certain degree of prediction power, the model performance (macro

AURCO of transformer-based neural network = 0.648) was weaker than the baseline model using structured EHR data alone (macro AURCO of ANN model = 0.708). Therefore, we developed several combined models incorporating operation notes and structured EHR data to verify our hypothesis that the operation notes can fill the information gaps in the baseline model.

Our experiments demonstrated that the transformer-based multimodal neural network (macro AURCO = 0.750) outperformed both types of models using structured (ANN; macro AUROC = 0.708) and unstructured data (Text Transformer; macro AUROC = 0.648) alone and other multimodal models (LSTM-MNN; macro AURCO = 0.725). The results are consistent with the previous text classification model - the transformer encoder block has a better ability to extract information from operation notes. This finding can support our hypothesis and provide directions for future studies.

Lastly, we used SHAP explainer to explore the important features and their effects on surgical outcomes in the random forest model. SHAP is a game theoretic approach to explaining the output of a machine learning model. The SHAP value represents the contribution of each feature of each instance to the prediction, which is a type of local explanation. Averaging the absolute SHAP values can get the global importance of the model. This analysis identified some novel risk factors, such as prostaglandin eye drops and oral angiotensin-converting enzyme inhibitors. Further investigation of the effects of these drugs on long-term surgical outcomes may be merited. Overall, identifying risk factors for different surgical failure causes can bring insights into clinical care and future clinical study.

To sum up, in this study, we demonstrated the prediction power of operation notes and developed multimodal neural networks to predict multiclass long-term surgical outcomes with reliable results. We think this work will be helpful for both postoperative clinical care and future research. Also, the identified risk factors can help with clinical decision-making and provide a direction for other clinical studies.

Chapter 7: SUMMARY AND CONCLUSIONS

In the recent decade, AI applications have been popularly applied in a variety of domains, such as disease screening, diagnosis improvement, and outcome predictions in ophthalmology. However, more work is needed to improve the performance of outcome predictions, which may include addressing EHR data quality issues. In this work, we developed multiple models that predict glaucoma surgery outcomes. As previously mentioned, glaucoma is the second leading cause of irreversible blindness globally and often leads to long-term life quality impairment. Medication therapies are the primary treatment to avoid disease progression for glaucoma patients. However, the accuracy of the medication list may be questionable. The discrepant medication documentation may result in unsafe treatment plans and adverse events in addition to limiting its potential for reuse in predictive models. Thus, we developed a robust active medication information extraction model to help with medication reconciliation as well as to improve the quality of the medication list. Next, we investigated the use of AI for supporting early postoperative management for glaucoma patients who undergo trabeculectomy since treatment varies according to different complications, but predicting these longer-term complications is difficult. Therefore, we developed a prediction model to classify which patient has a high risk of long-term surgical failure due to specific causes. The model combined structured EHR data and free-text operation notes to make multiclass predictions. The results indicate the model can identify more than half of the patients who may need reoperation or close follow-up due to specific complications. This model can be helpful for clinical decision-making for postoperative care and effective medical resource allocation. Finally, to better understand the risk factors of glaucoma surgical failures, we explored the important risk factors and their effects associated with surgical outcomes in the prediction model. In the future, to improve the model's performance and generalizability, we plan to combine image data in our multimodal neural network and include data from other institutions. In conclusion, we developed artificial intelligence models based on EHR data to improve glaucoma management for two common treatments: medications and trabeculectomy surgery.

References

1. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *British journal of ophthalmology*. 2006;90(3):262-267.

2. Tham Y-C, Li X, Wong TY, Quigley HA, Aung T, Cheng C-Y. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121(11):2081-2090.

3. Flaxman SR, Bourne RR, Resnikoff S, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *The Lancet Global Health*. 2017;5(12):e1221-e1234.

4. Schmier JK, Halpern MT, Jones ML. The Economic Implications of Glaucoma. *Pharmacoeconomics*. 2007;25(4):287-308.

5. Newman-Casey PA, Robin AL, Blachley T, et al. The most common barriers to glaucoma medication adherence: a cross-sectional survey. *Ophthalmology*. 2015;122(7):1308-1316.

6. Lacey J, Cate H, Broadway D. Barriers to adherence with glaucoma medications: a qualitative research study. *Eye*. 2009;23(4):924-932.

7. Ashfaq HA, Lester CA, Ballouz D, Errickson J, Woodward MA. Medication accuracy in electronic health records for microbial keratitis. *JAMA ophthalmology*. 2019;137(8):929-931.

8. Linsky A, Simon SR. Medication discrepancies in integrated electronic health records. *BMJ quality & safety*. 2013;22(2):103-109.

9. Gedde SJ, Schiffman JC, Feuer WJ, et al. Three-year follow-up of the tube versus trabeculectomy study. *American journal of ophthalmology*. 2009;148(5):670-684.

10. Garris WJ, Le C, Zurakowski D, Ayyala RS. Comparison of surgical outcomes between canaloplasty and trabeculectomy with mitomycin C at 2-year follow-up: A longitudinal cohort study. *Indian Journal of Ophthalmology*. 2018;66(1):66.

11. Kirwan JF, Lockwood AJ, Shah P, et al. Trabeculectomy in the 21st century: a multicenter analysis. *Ophthalmology*. 2013;120(12):2532-2539.

12. Suzuki R, Dickens CJ, Iwach AG, et al. Long-term follow-up of initially successful trabeculectomy with 5-fluorouracil injections. *Ophthalmology*. 2002;109(10):1921-1924.

13. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of clinical epidemiology*. 2009;62(10):e1-e34.

14. Weinreb RN, Khaw PT. Primary open-angle glaucoma. *The lancet*. 2004;363(9422):1711-1720.

15. Caprioli J, Varma R. Intraocular pressure: modulation as treatment for glaucoma. *American journal of ophthalmology*. 2011;152(3):340-344. e2.

16. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *Jama*. 2014;311(18):1901-1911.

17. Nuzzi R, Tridico F. Glaucoma: biological trabecular and neuroretinal pathology with perspectives of therapy innovation and preventive diagnosis. *Frontiers in Neuroscience*. 2017;11:494.

18. Group EDPR. Prevalence of open-angle glaucoma among adults in the United States. *Archives of ophthalmology*. 2004;122(4):532.

19. Ko Y-C, Hwang D-K, Chen W-T, Lee C-C, Liu CJ. Impact of socioeconomic status on the diagnosis of primary open-angle glaucoma and primary angle closure glaucoma: a nationwide population-based study in Taiwan. *PloS one*. 2016;11(2):e0149698.

20. Hood DC. Improving our understanding, and detection, of glaucomatous damage: an approach based upon optical coherence tomography (OCT). *Progress in retinal and eye research*. 2017;57:46-75.

21. Smith C, Vianna J, Chauhan B. Assessing retinal ganglion cell damage. *Eye*. 2017;31(2):209-217.

22. Chauhan BC, Malik R, Shuba LM, Rafuse PE, Nicolela MT, Artes PH. Rates of glaucomatous visual field change in a large clinical population. *Investigative ophthalmology & visual science*. 2014;55(7):4135-4143.

23. Heijl A, Buchholz P, Norrgren G, Bengtsson B. Rates of visual field progression in clinical glaucoma care. *Acta ophthalmologica*. 2013;91(5):406-412.

24. Saunders LJ, Russell RA, Kirwan JF, McNaught AI, Crabb DP. Examining visual field loss in patients in glaucoma clinics during their predicted remaining lifetime. *Investigative ophthalmology & visual science*. 2014;55(1):102-109.

25. Heijl A, Bengtsson B, Hyman L, Leske MC, Group EMGT. Natural history of open-angle glaucoma. *Ophthalmology*. 2009;116(12):2271-2276.

26. Aptel F, Aryal-Charles N, Giraud JM, et al. Progression of visual field in patients with primary open-angle glaucoma–P rog F study 1. *Acta Ophthalmologica*. 2015;93(8):e615-e620.

27. Bengtsson B, Patella VM, Heijl A. Prediction of glaucomatous visual field loss by extrapolation of linear trends. *Archives of ophthalmology*. 2009;127(12):1610-1615.

28. Spry PG, Johnson CA. Senescent changes of the normal visual field: an age-old problem. *Optometry and Vision Science*. 2001;78(6):436-441.

29. Drance S, Anderson DR, Schulzer M, Group CN-TGS. Risk factors for progression of visual field abnormalities in normal-tension glaucoma. *American journal of ophthalmology*. 2001;131(6):699-708.

30. Medeiros FA, Zangwill LM, Mansouri K, Lisboa R, Tafreshi A, Weinreb RN. Incorporating risk factors to improve the assessment of rates of glaucomatous progression. *Investigative Ophthalmology & Visual Science*. 2012;53(4):2199-2207.

31. Bengtsson B, Heijl A. A visual field index for calculation of glaucoma rate of progression. *American journal of ophthalmology*. 2008;145(2):343-353.

32. Leung CK, Yu M, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a prospective analysis of age-related loss. *Ophthalmology*. 2012;119(4):731-737.

33. Liu T, Tatham AJ, Gracitelli CP, Zangwill LM, Weinreb RN, Medeiros FA. Rates of retinal nerve fiber layer loss in contralateral eyes of glaucoma patients with unilateral progression by conventional methods. *Ophthalmology*. 2015;122(11):2243-2251.

34. Miki A, Medeiros FA, Weinreb RN, et al. Rates of retinal nerve fiber layer thinning in glaucoma suspect eyes. *Ophthalmology*. 2014;121(7):1350-1358.

35. Vianna JR, Danthurebandara VM, Sharpe GP, et al. Importance of normal aging in estimating the rate of glaucomatous neuroretinal rim and retinal nerve fiber layer loss. *Ophthalmology*. 2015;122(12):2392-2398.

36. Boland MV, Ervin A-M, Friedman DS, et al. Comparative effectiveness of treatments for open-angle glaucoma: a systematic review for the US Preventive Services Task Force. *Annals of internal medicine*. 2013;158(4):271-279.

37. Kass MA, Heuer DK, Higginbotham EJ, et al. The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Archives of ophthalmology*. 2002;120(6):701-713.

38. Heijl A, Leske M, Bengtsson B, Hyman L, Hussein M, Group EMGT. Reduction of intraocular pressure and glaucoma progression. *Evidence-Based Ophthalmology*. 2003;4(3):137-139.

39. Lichter PR, Musch DC, Gillespie BW, et al. Interim clinical outcomes in the Collaborative Initial Glaucoma Treatment Study comparing initial treatment randomized to medications or surgery. *Ophthalmology*. 2001;108(11):1943-1953.

40. Gaton DD, Sagara T, Lindsey JD, Kaufman PL, Weinreb RN. Increased matrix metalloproteinases 1, 2, and 3 in the monkey uveoscleral outflow pathway after topical prostaglandin F2α–isopropyl ester treatment. *Archives of ophthalmology*. 2001;119(8):1165-1170.

41. Liu JH, Kripke DF, Weinreb RN. Comparison of the nocturnal effects of once-daily timolol and latanoprost on intraocular pressure. *American journal of ophthalmology*. 2004;138(3):389-395.

42. Brown P, Marquard J, Amster B, et al. What do physicians read (and ignore) in electronic progress notes? *Applied clinical informatics*. 2014;5(02):430-444.

43. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *Journal of the American Medical Informatics Association*. 2011;18(2):181-186.

44. Tsai JC. Medication adherence in glaucoma: approaches for optimizing patient compliance. *Current opinion in ophthalmology*. 2006;17(2):190-195.

45. Tsai JC, Kanner EM. Current and emerging medical therapies for glaucoma. *Expert opinion on emerging drugs*. 2005;10(1):109-118.

46. Gurwitz JH, Glynn RJ, Monane M, et al. Treatment for glaucoma: adherence by the elderly. *American journal of public health*. 1993;83(5):711-716.

47. Patel SC, Spaeth GL. Compliance in patients prescribed eyedrops for glaucoma. SLACK Incorporated Thorofare, NJ; 1995. p. 233-236.

48. Osterberg L, Blaschke T. Adherence to medication. *New England journal of medicine*. 2005;353(5):487-497.

49. Muir KW, Lee PP. Glaucoma medication adherence: room for improvement in both performance and measurement. *Archives of ophthalmology*. 2011;129(2):243-245.

50. Hamid MS, Valicevic A, Brenneman B, Niziol LM, Stein JD, Newman-Casey PA. Text Parsing-Based Identification of Patients with Poor Glaucoma Medication Adherence in the Electronic Health Record. *American journal of ophthalmology*. 2021;222:54-59.

51. Mansouri K, Medeiros FA, Weinreb RN. Global rates of glaucoma surgery. *Graefe's Archive for Clinical and Experimental Ophthalmology*. 2013;251(11):2609-2615.

52. Odberg T, Sandvik L. The medium and long-term efficacy of primary argon laser trabeculoplasty in avoiding topical medication in open angle glaucoma. *Acta Ophthalmologica Scandinavica*. 1999;77(2):176-181.

53. Shingleton BJ, Richter CU, Dharma SK, et al. Long-term efficacy of argon laser trabeculoplasty: a 10-year follow-up study. *Ophthalmology*. 1993;100(9):1324-1329.

54. Christakis PG, Zhang D, Budenz DL, et al. Five-year pooled data analysis of the Ahmed Baerveldt comparison study and the Ahmed versus Baerveldt study. *American Journal of Ophthalmology*. 2017;176:118-126.

55. Fargione RA, Tansuebchueasai N, Lee R, Tai TYT. Etiology and management of the hypertensive phase in glaucoma drainage-device surgery. *survey of ophthalmology*. 2019;64(2):217-224.

56. Stephenson M. Is It Time to Retire the Trabeculectomy? *Review of Ophthalmology*. 2009;16(6):36.

57. Vinod K, Gedde SJ, Feuer WJ, et al. Practice preferences for glaucoma surgery: a survey of the American Glaucoma Society. *Journal of glaucoma*. 2017;26(8):687.

58. Jinza K, Saika S, Kin K, Ohnishi Y. Relationship between formation of a filtering bleb and an intrascleral aqueous drainage route after trabeculectomy: evaluation using ultrasound biomicroscopy. *Ophthalmic Research*. 2000;32(5):240-243.

59. Ichhpujani PB, Shibal. Releasable Sutures in Trabeculectomy. In: Tarek Shaarawy TD, Shibal Bhartiya, ed. *ISGS Textbook of Glaucoma Surgery*. Jaypee Pub; 2014:128-135:chap 15.

60. Keicher AS, Stoffelns BM. Argon laser suture lysis following glaucoma filtering surgery–A short introduction to the procedure. *Medical Laser Application*. 2010;25(4):209-213.

61. Vijaya L, Manish P, Ronnie G, Shantha B. Management of complications in glaucoma surgery. *Indian journal of ophthalmology*. 2011;59(Suppl1):S131-S140.

62. Shaarawy T, Sherwood MB, Hitchings RA, Crowston JG. *Glaucoma E-Book*. Elsevier Health Sciences; 2014.

63. Rong SS, Meng HL, Fan SJ, et al. Can intraoperative intraocular pressure during primary trabeculectomy predict early postoperative pressure? *Journal of glaucoma*. 2014;23(9):653-657.

64. Murdoch I. Post-operative management of trabeculectomy in the first three months. *Community Eye Health*. 2012;25(79-80):73.

65. Wormald R, Wilkins M, Bunce C. Postoperative 5-Fluorouracil for glaucoma surgery. *Cochrane Database of Systematic Reviews*. 2001;(3)

66. Panarelli JF, Nayak NV, Sidoti PA. Postoperative management of trabeculectomy and glaucoma drainage implant surgery. *Current Opinion in Ophthalmology*. 2016;27(2):170-176.

67. Broadway DC, Chang LP. Trabeculectomy, risk factors for failure and the preoperative state of the conjunctiva. *Journal of glaucoma*. 2001;10(3):237-249.

68. Landers J, Martin K, Sarkies N, Bourne R, Watson P. A twenty-year follow-up study of trabeculectomy: risk factors and outcomes. *Ophthalmology*. 2012;119(4):694-702.

69. Rong S, Feng M, Wang N, et al. Can early postoperative intraocular pressure predict success following mitomycin-C augmented trabeculectomy in primary angle-closure glaucoma. *Eye*. 2013;27(3):403-409.

70. Esfandiari H, Pakravan M, Loewen NA, Yaseri M. Predictive value of early postoperative IOP and bleb morphology in Mitomycin-C augmented trabeculectomy. *F1000Research*. 2017;6

71. Nesaratnam N, Sarkies N, Martin KR, Shahid H. Pre-operative intraocular pressure does not influence outcome of trabeculectomy surgery: a retrospective cohort study. *BMC ophthalmology*. 2015;15(1):1-8.

72. James G, Witten D, Hastie T, Tibshirani R. *An introduction to statistical learning*. vol 112. Springer; 2013.

73. Hastie T, Tibshirani R, Friedman JH, Friedman JH. *The elements of statistical learning: data mining, inference, and prediction*. vol 2. Springer; 2009.

74. Lin W-C, Chen JS, Chiang MF, Hribar MR. Applications of artificial intelligence to electronic health record data in ophthalmology. *Translational vision science & technology*. 2020;9(2):13-13.

75. Liaw A, Wiener M. Classification and regression by randomForest. *R news*. 2002;2(3):18-22.

76. Kleinbaum DG, Dietz K, Gail M, Klein M, Klein M. *Logistic regression*. Springer; 2002.

77. Burges CJ. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*. 1998;2(2):121-167.

78. Friedman JH. Stochastic gradient boosting. *Computational statistics & data analysis*. 2002;38(4):367-378.

79. Kukreja SL, Löfberg J, Brenner MJ. A least absolute shrinkage and selection operator (LASSO) for nonlinear system identification. *IFAC proceedings volumes*. 2006;39(1):814-819.

80. Schapire RE. Explaining adaboost. *Empirical inference*. Springer; 2013:37-52.

81. Goodfellow I, Bengio Y, Courville A. *Deep learning*. MIT press; 2016.

82. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association*. 2017;24(2):361-370.

83. Ruan T, Lei L, Zhou Y, et al. Representation learning for clinical time series prediction tasks in electronic health records. *BMC medical informatics and decision making*. 2019;19(8):1-14.

84. Chen R, Stewart WF, Sun J, Ng K, Yan X. Recurrent neural networks for early detection of heart failure from longitudinal electronic health record data: implications for temporal modeling with respect to time before diagnosis, data density, data quantity, and data type. *Circulation: Cardiovascular Quality and Outcomes*. 2019;12(10):e005114.

85. Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 1998;6(02):107-116.

86. Yu Y, Si X, Hu C, Zhang J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation*. 2019;31(7):1235-1270.

87. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:14123555*. 2014;

88. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Advances in neural information processing systems*. 2017;30

89. Soyalp G, Alar A, Ozkanli K, Yildiz B. Improving Text Classification with Transformer. IEEE; 2021:707-712.

90. Stahlschmidt SR, Ulfenborg B, Synnergren J. Multimodal deep learning for biomedical data fusion: a review. *Briefings in Bioinformatics*. 2022;23(2):bbab569.

91. Peters ME, Neumann M, Zettlemoyer L, Yih W-t. Dissecting contextual word embeddings: Architecture and representation. *arXiv preprint arXiv:180808949*. 2018;

92. Lin W-C, Chen JS, Kaluzny J, Chen A, Chiang MF, Hribar MR. Extraction of Active Medications and Adherence Using Natural Language Processing for Glaucoma Patients. American Medical Informatics Association; 2021:773.

93. del Carmen Legaz-García M, Martínez-Costa C, Menárguez-Tortosa M, Fernández-Breis JT. A semantic web based framework for the interoperability and exploitation of clinical models and EHR data. *Knowledge-Based Systems*. 2016;105:175-189.

94. Kadariya D, Venkataramanan R, Yip HY, Kalra M, Thirunarayanan K, Sheth A. kBot: knowledge-enabled personalized chatbot for asthma self-management. IEEE; 2019:138-143.

95. Dalianis H. *Clinical text mining: Secondary use of electronic patient records*. Springer Nature; 2018.

96. Wu S, Roberts K, Datta S, et al. Deep learning in clinical natural language processing: a methodical review. *Journal of the American Medical Informatics Association*. 2020;27(3):457-470.

97. Schütze H, Manning CD, Raghavan P. *Introduction to information retrieval*. vol 39. Cambridge University Press Cambridge; 2008.

98. Kormilitzin A, Vaci N, Liu Q, Nevado-Holgado A. Med7: A transferable clinical natural language processing model for electronic health records. *Artificial Intelligence in Medicine*. 2021;118:102086.

99. Tarcar AK, Tiwari A, Dhaimodker VN, Rebelo P, Desai R, Rao D. Healthcare NER models using language model pretraining. *arXiv preprint arXiv:191011241*. 2019;

100. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:13013781*. 2013;

101. Pennington J, Socher R, Manning CD. Glove: Global vectors for word representation. 2014:1532-1543.

102. Kowsari K, Jafari Meimandi K, Heidarysafa M, Mendu S, Barnes L, Brown D. Text classification algorithms: A survey. *Information*. 2019;10(4):150.

103. García-Laencina PJ, Sancho-Gómez J-L, Figueiras-Vidal AR. Pattern classification with missing data: a review. *Neural Computing and Applications*. 2010;19(2):263-282.

104. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychological methods*. 2002;7(2):147.

105. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research*. 2011;20(1):40-49.

106. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112-118.

107. Aksman LM, Scelsi MA, Marquand AF, et al. Modeling longitudinal imaging biomarkers with parametric Bayesian multi-task learning. *Human brain mapping*. 2019;40(13):3982-4000.

108. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*. 2018;8(1):1-12.

109. Junninen H, Niska H, Tuppurainen K, Ruuskanen J, Kolehmainen M. Methods for imputation of missing values in air quality data sets. *Atmospheric Environment*. 2004;38(18):2895-2907.

110. Liu Y, Brown SD. Comparison of five iterative imputation methods for multivariate classification. *Chemometrics and Intelligent Laboratory Systems*. 2013;120:106-115.

111. Refaeilzadeh P, Tang L, Liu H. Cross-validation. *Encyclopedia of database systems*. 2009;5:532-538.

112. Koehrsen W. Overfitting vs. underfitting: A complete example. *Towards Data Science*. 2018;

113. Ying X. An overview of overfitting and its solutions. IOP Publishing; 2019:022022.

114. Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. Springer; 2005:345-359.

115. Gaskin GL, Pershing S, Cole TS, Shah NH. Predictive modeling of risk factors and complications of cataract surgery. *European journal of ophthalmology*. 2016;26(4):328-337.

116. Chaganti S, Nabar KP, Nelson KM, Mawn LA, Landman BA. Phenotype analysis of early risk factors from electronic medical records improves image-derived diagnostic classifiers for optic nerve pathology. SPIE; 2017:99-105.

117. Fraccaro P, Nicolo M, Bonetto M, et al. Combining macula clinical signs and patient characteristics for age-related macular degeneration diagnosis: a machine learning approach. *BMC ophthalmology*. 2015;15(1):1-9.

118. Baxter SL, Marks C, Kuo T-T, Ohno-Machado L, Weinreb RN. Machine learning-based predictive modeling of surgical intervention in glaucoma using systemic data from electronic health records. *American journal of ophthalmology*. 2019;208:30-40.

119. Coyner AS, Swan R, Campbell JP, et al. Automated fundus image quality assessment in retinopathy of prematurity using deep convolutional neural networks. *Ophthalmology Retina*. 2019;3(5):444-450.

120. Coyner AS, Chen J, Campbell JP, et al. Diagnosability of synthetic retinal fundus images for plus disease detection in retinopathy of prematurity. American Medical Informatics Association; 2020:329.

121. Coyner AS, Oh MA, Shah PK, et al. External Validation of a Retinopathy of Prematurity Screening Model Using Artificial Intelligence in 3 Low-and Middle-Income Populations. *JAMA ophthalmology*. 2022;

122. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA ophthalmology*. 2018;136(7):803-810.

123. Kim J, Ryu IH, Kim JK, et al. Machine learning predicting myopic regression after corneal refractive surgery using preoperative data and fundus photography. *Graefe's Archive for Clinical and Experimental Ophthalmology*. 2022:1-10.

124. Safran C, Bloomrosen M, Hammond WE, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *J Am Med Inform Assoc*. 2007;14(1):1-9.

125. Charles D, Gabriel M, Searcy T. Adoption of electronic health record systems among US non-federal acute care hospitals: 2008-2012. *J ONC data brief*. 2013;9:1-9.

126. Coorevits P, Sundgren M, Klein GO, et al. Electronic health records: new opportunities for clinical research. *J Intern Med*. 2013;274(6):547-560.

127. Cano I, Tenyi A, Vela E, Miralles F, Roca J. Perspectives on Big Data applications of health information. *Curr Opin Syst Biol*. 2017;3:36-42.

128. Khennou F, Khamlichi YI, Chaoui NEH. Improving the Use of Big Data Analytics within Electronic Health Records: A Case Study based OpenEHR. *Procedia Comput Sci*. 2018;127:60-68.

129. Meystre S, Lovis C, Bürkle T, Tognola G, Budrionis A, Lehmann C. Clinical data reuse or secondary use: current status and potential future progress. *Yearb Med Inform*. 2017;26(01):38-52.

130. Cowie MR, Blomster JI, Curtis LH, et al. Electronic health records to facilitate clinical research. *Clin Res Cardiol*. Jan 2017;106(1):1-9. doi:10.1007/s00392-016-1025-6

131. Kuperman GJ, Bobb A, Payne TH, et al. Medication-related clinical decision support in computerized provider order entry systems: a review. *J Am Med Inform Assoc*. 2007;14(1):29-40.

132. Jiang M, Chen Y, Liu M, et al. A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. *J Am Med Inform Assoc*. 2011;18(5):601-606.

133. Kononenko IJAIim. Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med*. 2001;23(1):89-109.

134. Persell SD, Dunne AP, Lloyd-Jones DM, Baker DWJMc. Electronic health record-based cardiac risk assessment and identification of unmet preventive needs. *Med Care*. 2009;47(4):418-424.

135. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. *J Am Med Inform Assoc*. 2016;24(2):361-370.

136. Cheng Y, Wang F, Zhang P, Hu J. Risk prediction with electronic health records: A deep learning approach. SIAM; 2016:432-440.

137. Motwani M, Dey D, Berman DS, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J*. 2016;38(7):500-507.

138. Gaskin GL, Pershing S, Cole TS, Shah NH. Predictive modeling of risk factors and complications of cataract surgery. *Eur J Ophthalmol*. 2016;26(4):328-337.

139. Chaganti S, Nabar KP, Nelson KM, Mawn LA, Landman BA. Phenotype analysis of early risk factors from electronic medical records improves image-derived diagnostic classifiers for optic nerve pathology. International Society for Optics and Photonics; 2017:101380F.

140. Fraccaro P, Nicolo M, Bonetto M, et al. Combining macula clinical signs and patient characteristics for age-related macular degeneration diagnosis: a machine learning approach. *BMC Ophthalmol*. 2015;15(1):10.

141. Saleh E, Błaszczyński J, Moreno A, et al. Learning ensemble classifiers for diabetic retinopathy assessment. *Artif Intell Med*. 2018;85:50-63.

142. Caixinha M, Nunes S. Machine learning techniques in clinical vision sciences. *Curr Eye Res*. 2017;42(1):1-15.

143. Lin H, Long E, Ding X, et al. Prediction of myopia development among Chinese school-aged children using refraction data from electronic medical records: A retrospective, multicentre machine learning study. *PLoS Med*. 2018;15(11):e1002674.

144. Lee CS, Baughman DM, Lee AY. Deep learning is effective for classifying normal versus age-related macular degeneration OCT images. *Ophthalmol Retina*. 2017;1(4):322-327.

145. Baxter SL, Marks C, Kuo T-T, Ohno-Machado L, Weinreb RN. Machine learning-based predictive modeling of surgical intervention in glaucoma using systemic data from electronic health records. *Am J Ophthalmol*. 2019/07// 2019;doi:10.1016/j.ajo.2019.07.005

146. Apostolova E, White HA, Morris PA, Eliason DA, Velez T. Open Globe Injury Patient Identification in Warfare Clinical Notes. *AMIA Annual Symposium proceedings AMIA Symposium*. 2017;2017:403-410.

147. Rohm M, Tresp V, Muller M, et al. Predicting Visual Acuity by Using Machine Learning in Patients Treated for Neovascular Age-Related Macular Degeneration. *Ophthalmol*. Jul 2018;125(7):1028-1036. doi:10.1016/j.ophtha.2017.12.034

148. Yoo TK, Park E-C. Diabetic retinopathy risk prediction for fundus examination using sparse learning: a cross-sectional study. *BMC Med Inform Decis Mak*. 2013;13(1):106.

149. Sramka M, Slovak M, Tuckova J, Stodulka P. Improving clinical refractive results of cataract surgery by machine learning. *PeerJ*. 2019;7:e7202.

150. Peissig PL, Rasmussen LV, Berg RL, et al. Importance of multi-modal approaches to effectively identify cataract cases from electronic health records. *J Am Med Inform Assoc*. 2012;19(2):225-234.

151. Skevofilakas M, Zarkogianni K, Karamanos BG, Nikita KS. A hybrid Decision Support System for the risk assessment of retinopathy development as a long term complication of Type 1 Diabetes Mellitus. IEEE; 2010:6713-6716.

152. Anzai Y. *Pattern recognition and machine learning*. Elsevier; 2012.

153. Ramesh A, Kambhampati C, Monson JR, Drew P. Artificial intelligence in medicine. *Ann R Coll Surg Engl*. 2004;86(5):334.

154. Burges CJ. A tutorial on support vector machines for pattern recognition. *Data mining knowledge discovery*. 1998;2(2):121-167.

155. Lewis RJ. An introduction to classification and regression tree (CART) analysis. 2000:

156. Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal*. 2002;38(4):367-378.

157. Kukreja SL, Löfberg J, Brenner M. A least absolute shrinkage and selection operator (LASSO) for nonlinear system identification. *IFAC proceedings volumes*. 2006;39(1):814-819.

158. Jain AK, Mao J, Mohiuddin KM. Artificial neural networks: A tutorial. *Computer*. 1996;29(3):31-44.

159. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. 2012:1097-1105.

160. Kourentzes N, Barrow DK, Crone SF. Neural network ensemble operators for time series forecasting. *Expert Syst Appl*. 2014;41(9):4235-4244.

161. Svozil D, Kvasnicka V, Pospichal J. Introduction to multi-layer feed-forward neural networks. *Chemometr Intell Lab Syst*. 1997;39(1):43-62.

162. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *J Am Med Inform Assoc*. 2011;18(5):544-551.

163. Hanley JA, McNeil B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29-36.

164. Buckland M, Gey F. The relationship between recall and precision. *J Am Med Inform Assoc*. 1994;45(1):12-19.

165. Chawla NV, Japkowicz N, Kotcz A. Special issue on learning from imbalanced data sets. *ACM Sigkdd Explorations Newsletter*. 2004;6(1):1-6.

166. Van Hulse J, Khoshgoftaar TM, Napolitano A. Experimental perspectives on learning from imbalanced data. ACM; 2007:935-942.

167. Marlin BM, Kale DC, Khemani RG, Wetzel RC. Unsupervised pattern discovery in electronic health care data using probabilistic clustering models. ACM; 2012:389-398.

168. Organization WH. Global data on visual impairments 2010. *WHO Press*. 2012;

169. Alghamdi HF. Causes of irreversible unilateral or bilateral blindness in the Al Baha region of the Kingdom of Saudi Arabia. *Saudi J Ophthalmol*. 2016;30(3):189-193.

170. Weinreb RN, Khaw PT. Primary open-angle glaucoma. *Lancet*. 2004;363(9422):1711-1720.

171. Wang W, Yan W, Fotis K, et al. Cataract surgical rate and socioeconomics: a global study. *Invest Ophthalmol Vis Sci*. 2016;57(14):5872-5881.

172. Hersh WR. Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. *Clin Pharmacol Ther*. 2007;81:126-8.

173. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary use of EHR: data quality issues and informatics opportunities. *Summit Transl Bioinform*. 2010;2010:1.

174. Ashfaq HA, Lester CA, Ballouz D, Errickson J, Woodward MA. Medication Accuracy in Electronic Health Records for Microbial Keratitis. *JAMA ophthalmology*. May 30 2019;doi:10.1001/jamaophthalmol.2019.1444

175. Berdahl CT, Moran GJ, McBride O, Santini AM, Verzhbinsky IA, Schriger DL. Concordance Between Electronic Clinical Documentation and Physicians' Observed Behavior. *JAMA network open*. Sep 4 2019;2(9):e1911390. doi:10.1001/jamanetworkopen.2019.11390

176. Garvin MK, Abramoff MD, Wu X, Russell SR, Burns TL, Sonka M. Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *IEEE Trans Med Imaging*. 2009;28(9):1436-1447.

177. Parke DW, Rich WL, Sommer A, Lum F. The American Academy of Ophthalmology's IRIS® Registry (Intelligent Research in Sight Clinical Data): A look back and a look to the future. *Ophthalmol*. 2017;124(11):1572-1574.

178. Clark A, Ng JQ, Morlet N, Semmens JB. Big data and ophthalmic research. *Surv Ophthalmol*. 2016;61(4):443-465.

179. Coleman AL. How big data informs us about cataract surgery: the LXXII Edward Jackson Memorial Lecture. *Am J Ophthalmol*. 2015;160(6):1091-1103. e3.

180. Chute CG, Ullman-Cullere M, Wood GM, Lin SM, He M, Pathak J. Some experiences and opportunities for big data in translational research. *Genet Med*. 2013;15(10):802-809.

181. LePendu P, Iyer SV, Fairon C, Shah NH. Annotation analysis for testing drug safety signals using unstructured clinical notes. Springer; 2012:1-12.

182. Warrer P, Hansen EH, Juhl-Jensen L, Aagaard L. Using text-mining techniques in electronic patient records to identify ADRs from medicine use. *Br J Clin Pharmacol*. 2012;73(5):674-684.

183. Forrest CB, Margolis PA, Bailey LC, et al. PEDSnet: a national pediatric learning health system. *J Am Med Inform Assoc*. 2014;21(4):602-606.

184. Rothman B, Leonard JC, Vigoda MM. Future of electronic health records: implications for decision support. *Mt Sinai J Med*. Nov-Dec 2012;79(6):757-768. doi:10.1002/msj.21351

185. Wagholikar KB, Sundararajan V, Deshpande AW. Modeling paradigms for medical diagnostic decision support: a survey and future directions. *J Med Syst*. Oct 2012;36(5):3029-3049. doi:10.1007/s10916-011-9780-4

186. Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of information science*. 2007;33(2):163-180.

187. Meystre SM, Lovis C, Bürkle T, Tognola G, Budrionis A, Lehmann CU. Clinical data reuse or secondary use: current status and potential future progress. *Yearb Med Inform*. 2017;26(1):38.

188. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Inform*. 2013;46(5):830-836.

189. FitzGerald RJ. Medication errors: the importance of an accurate drug history. *Br J Clin Pharmacol*. Jun 2009;67(6):671-675. doi:10.1111/j.1365-2125.2009.03424.x

190. Ashfaq HA, Lester CA, Ballouz D, Errickson J, Woodward MA. Medication accuracy in electronic health records for microbial keratitis. *JAMA Ophthalmol*. May 30 2019;137(8):929-931. doi:10.1001/jamaophthalmol.2019.1444

191. Mekonnen AB, McLachlan AJ, Brien JaE. Pharmacy-led medication reconciliation programmes at hospital transitions: a systematic review and meta-analysis. *J Clin Pharm Ther*. 2016;41(2):128-144.

192. Woodward MA, Ballouz D, Errickson J, Maganti N, Tuohy M. Medication burden for patients with bacterial keratitis. *Invest Ophthalmol Vis Sci*. 2019;60(9):841-841.

193. Walsh KE, Marsolo KA, Davis C, et al. Accuracy of the medication list in the electronic health record-implications for care, research, and improvement. *J Am Med Inform Assoc*. Jul 1 2018;25(7):909-912. doi:10.1093/jamia/ocy027

194. Caglar S, Henneman PL, Blank FS, Smithline HA, Henneman EA. Emergency department medication lists are not accurate. *J Emerg Med*. 2011;40(6):613-616.

195. Orrico KB. Sources and types of discrepancies between electronic medical records and actual outpatient medication use. *J Manag Care Pharm*. 2008;14(7):626-631.

196. Chan KS, Fowles JB, Weiner JP. Electronic health records and the reliability and validity of quality measures: a review of the literature. *Med Care Res Rev*. Oct 2010;67(5):503-527. doi:10.1177/1077558709359007

197. Brown P, Marquard J, Amster B, et al. What do physicians read (and ignore) in electronic progress notes? *Appl Clin Inform*. 2014;5(2):430.

198. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *J Am Med Inform Assoc*. 2011;18(2):181-186.

199. Kwan JL, Lo L, Sampson M, Shojania KG. Medication reconciliation during transitions of care as a patient safety strategy: a systematic review. *Annals of internal medicine*. 2013;158(5_Part_2):397-403.

200. Moghli MA, Farha RA, Hammour KA. Medication discrepancies in hospitalized cancer patients: Do we need medication reconciliation? *Journal of Oncology Pharmacy Practice*. 2021;27(5):1139-1146.

201. Peters ME, Neumann M, Iyyer M, et al. Deep contextualized word representations. *arXiv preprint arXiv:180205365*. 2018;

202. Wang Y, Wang L, Rastegar-Mojarad M, et al. Clinical information extraction applications: a literature review. *J Biomed Inform*. Jan 2018;77:34-49. doi:10.1016/j.jbi.2017.11.011

203. del Carmen Legaz-García M, Martínez-Costa C, Menárguez-Tortosa M, Fernández-Breis JT. A semantic web based framework for the interoperability and exploitation of clinical models and EHR data. *Knowl Based Syst*. 2016;105:175-189.

204. Mohit B. Named entity recognition. *Natural language processing of semitic languages*. Springer; 2014:221-245.

205. Wunnava S, Qin X, Kakar T, Sen C, Rundensteiner EA, Kong X. Adverse drug event detection from electronic health records using hierarchical recurrent neural networks with dual-level embedding. *Drug safety*. 2019;42(1):113-122.

206. Chapman AB, Peterson KS, Alba PR, DuVall SL, Patterson OV. Detecting adverse drug events with rapidly trained classification models. *Drug saf*. Jan 2019;42(1):147-156. doi:10.1007/s40264-018-0763-y

207. Magge A, Scotch M, Gonzalez-Hernandez G. Clinical NER and relation extraction using bi-char-LSTMs and random forest classifiers. PMLR; 2018:25-30.

208. Kormilitzin A, Vaci N, Liu Q, Nevado-Holgado A. Med7: a transferable clinical natural language processing model for electronic health records. *arXiv preprint arXiv:200301271*. 2020;

209. Lee J, Yoon W, Kim S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. Feb 15 2020;36(4):1234-1240. doi:10.1093/bioinformatics/btz682

210. Flaxman SR, Bourne RR, Resnikoff S, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *Lancet Glob Health*. Dec 2017;5(12):e1221-e1234. doi:10.1016/S2214-109X(17)30393-5

211. Tsai JC. Medication adherence in glaucoma: approaches for optimizing patient compliance. *Curr Opin Ophthalmol*. Apr 2006;17(2):190-195. doi:10.1097/01.icu.0000193078.47616.aa

212. Gurwitz JH, Glynn RJ, Monane M, et al. Treatment for glaucoma: adherence by the elderly. *Am J Public Health*. May 1993;83(5):711-716. doi:10.2105/ajph.83.5.711

213. Osterberg L, Blaschke T. Adherence to medication. *N Engl J Med*. Aug 4 2005;353(5):487-497. doi:10.1056/NEJMra050100

214. Muir KW, Lee PP. Glaucoma medication adherence: room for improvement in both performance and measurement. *Arch Ophthal*. 2011;129(2):243-245.

215. Doccano: Open Source text annotation for machine learning practitioner [Internet]. Updated 2018. Accessed December 15, 2020. https://github.com/doccano/doccano

216. Honnibal M, Montani I. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *To appear*. 2017;7(1):411-420.

217. Van Rossum G. Python Programming Language. 2007:36.

218. Serra J, Karatzoglou A. Getting deep recommenders fit: Bloom embeddings for sparse binary input/output networks. 2017:279-287.

219. Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C. Neural architectures for named entity recognition. *arXiv preprint arXiv:160301360*. 2016;

220. Liu S, Ma W, Moore R, Ganesan V, Nelson S. RxNorm: prescription for electronic drug information exchange. *IT professional*. 2005;7(5):17-23.

221. Hellström LM, Bondesson Å, Höglund P, Eriksson T. Errors in medication history at hospital admission: prevalence and predicting factors. *BMC Clin Pharmacol*. 2012;12(1):1-9.

222. Ngo D-H, Metke-Jimenez A, Nguyen A. Knowledge-based feature engineering for detecting medication and adverse drug events from electronic health records. PMLR; 2018:31-38.

223. Harasymowycz P, Birt C, Gooi P, et al. Medical management of glaucoma in the 21st century from a Canadian perspective. *Journal of ophthalmology*. 2016;2016

224. Allison K, Patel D, Alabi O. Epidemiology of glaucoma: the past, present, and predictions for the future. *Cureus*. 2020;12(11)

225. The A. The Advanced Glaucoma Intervention Study (AGIS): 9. Comparison of glaucoma outcomes in black and white patients within treatment groups. *American journal of ophthalmology*. 2001;132(3):311-320.

226. Jonathan Y, Al-Mugheiry TS, Gale J, Martin KR. Factors predicting the success of trabeculectomy bleb enhancement with needling. *British Journal of Ophthalmology*. 2018;102(12):1667-1671.

227. Razeghinejad MR, Fudemberg SJ, Spaeth GL. The changing conceptual basis of trabeculectomy: a review of past and current surgical techniques. *Survey of ophthalmology*. 2012;57(1):1-25.

228. Vijaya L, Manish P, Ronnie G, Shantha B. Management of complications in glaucoma surgery. *Indian journal of ophthalmology*. 2011;59(Suppl1):S131.

229. Jeong Y-S, Kim J, Kim D, et al. Prediction of postoperative complications for patients of end stage renal disease. *Sensors*. 2021;21(2):544.

230. Jalali A, Lonsdale H, Do N, et al. Deep learning for improved risk prediction in surgical outcomes. *Scientific reports*. 2020;10(1):1-13.

231. Merali ZG, Witiw CD, Badhiwala JH, Wilson JR, Fehlings MG. Using a machine learning approach to predict outcome after surgery for degenerative cervical myelopathy. *PloS one*. 2019;14(4):e0215133.

232. Banna HU, Zanabli A, McMillan B, et al. Evaluation of machine learning algorithms for trabeculectomy outcome prediction in patients with glaucoma. *Scientific Reports*. 2022;12(1):1-11.

233. Okimoto S, Kiuchi Y, Akita T, Tanaka J. Using the early postoperative intraocular pressure to predict pressure control after a trabeculectomy. *Journal of glaucoma*. 2014;23(6):410-414.

234. Tojo N, Hayashi A, Otsuka M. Evaluation of early postoperative intraocular pressure for success after ex-press surgery. *Journal of Current Glaucoma Practice*. 2019;13(2):55.

235. Adhikari L, Ozrazgat-Baslanti T, Ruppert M, et al. Improved predictive models for acute kidney injury with IDEA: Intraoperative Data Embedded Analytics. *PloS one*. 2019;14(4):e0214904.

236. Tseng P-Y, Chen Y-T, Wang C-H, et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Critical Care*. 2020;24(1):1-13.

237. Sarzynska-Wawer J, Wawer A, Pawlak A, et al. Detecting formal thought disorder by deep contextualized word representations. *Psychiatry Research*. 2021;304:114135.

238. Wang Y, Wang L, Rastegar-Mojarad M, et al. Clinical information extraction applications: a literature review. *Journal of biomedical informatics*. 2018;77:34-49.

239. Luan Y, Lin S. Research on text classification based on CNN and LSTM. IEEE; 2019:352-355.

240. Kiela D, Firooz H, Mohan A, et al. The hateful memes challenge: Detecting hate speech in multimodal memes. *Advances in Neural Information Processing Systems*. 2020;33:2611-2624.

241. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. *BMC medical informatics and decision making*. 2020;20(1):1-11.

242. Mahajan SM, Ghani R. Combining Structured and Unstructured Data for Predicting Risk of Readmission for Heart Failure Patients. 2019:238-242.

243. Tezgider M, Yildiz B, Aydin G. Text classification using improved bidirectional transformer. *Concurrency and Computation: Practice and Experience*. 2022;34(9):e6486.

244. Řehůřek R, Sojka P. Gensim—statistical semantics in python. *Retrieved from genism org*. 2011;

245. Holladay JT. Visual acuity measurements. *Journal of Cataract & Refractive Surgery*. 2004;30(2):287-290.

246. Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*. 2019;32

247. Van Rossum G, Drake FL. *Python 3 reference manual*. CreateSpace; 2009.

248. Biau G, Scornet E. A random forest guided tour. *Test*. 2016;25(2):197-227.

249. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems*. 2017;30

250. Shin DH, Kim YY, Ginde SY, et al. Risk factors for failure of 5-fluorouracil needling revision for failed conjunctival filtration blebs. *American journal of ophthalmology*. 2001;132(6):875-880.

251. Yu D-Y, Morgan WH, Sun X, et al. The critical role of the conjunctiva in glaucoma filtration surgery. *Progress in retinal and eye research*. 2009;28(5):303-328.

252. Shaheen Z, Wohlgenannt G, Filtz E. Large scale legal text classification using transformer models. *arXiv preprint arXiv:201012871*. 2020;

253. Takihara Y, Inatani M, Ogata-Iwao M, et al. Trabeculectomy for open-angle glaucoma in phakic eyes vs in pseudophakic eyes after phacoemulsification: a prospective clinical cohort study. *JAMA ophthalmology*. 2014;132(1):69-76.

254. Mathew RG, Parvizi S, Murdoch IE. Success of trabeculectomy surgery in relation to cataract surgery: 5-year outcomes. *British Journal of Ophthalmology*. 2019;103(10):1395-1400.

255. Wang SY, Tseng B, Hernandez-Boussard T. Deep Learning Approaches for Predicting Glaucoma Progression Using Electronic Health Records and Natural Language Processing. *Ophthalmology Science*. 2022;2(2):100127.

256. Chen JS, Lin W-C, Kaluzny JV, Chen A, Chiang MF, Hribar MR. Discrepancies in Ophthalmic Medication Documentation for Glaucoma Patients. *Ophthalmology Science*. 2022;2(1)

257. Gandhi TK, Weingart SN, Seger AC, et al. Outpatient prescribing errors and the impact of computerized prescribing. *Journal of general internal medicine*. 2005;20(9):837-841.