Artificial Intelligence (AI) in Ophthalmology: An Overview

Syeed Mehbub Ul Kadir1*, Rokhsanda Rehnuma2, Md. Siddiquur Rahman3, Mostafa Zahir Raihani4, Mahar Ali1 and Muhammad Moniruzzaman5

1Assistant Professor, Sheikh Fazilatunnesa Mujib Eye Hospital and Training Institute, Bangladesh.
2Consultant, Bangladesh Eye Care Hospital, Dhaka, Bangladesh.
3Professor (Ophthalmology), Northern International Medical College, Dhaka, Bangladesh.
4Senior Consultant, Sheikh Fazilatunnesa Mujib Eye Hospital and Training Institute, Bangladesh.
5Consultant (Vitreo-Retina), Vision Eye Hospital, Dhaka, Bangladesh.

ABSTRACT

Artificial intelligence (AI) is an emerging continuous equipment-based power that is initiating to spread throughout to all aspects of our lifestyles for digital revolution even in medical diagnosis, management of the disease AI application in ophthalmology can help in screening, diagnosing, staging, and providing best possible management planning of various eye diseases especially for sight-threatening eye conditions. This paper demonstrates a review of the art of AI in the ophthalmic field and focusing on the applications of AI for the diagnosis and treatment of ophthalmic diseases including cataract, glaucoma, diabetic retinopathy, age-related macular degeneration, retinopathy of prematurity, and future AI applications with image enhancement in Ophthalmology. Diabetic retinopathy is occurring due to damage of retinal blood. Advanced diabetic retinopathy can cause blindness. So, early detection and management of diabetic retinopathy is crucial to prevent vision loss. AI based technology is playing a role as an auxiliary assistant in screening and diagnostic support for ophthalmologist.

Keywords
Artificial intelligence, Machine learning, Deep learning, Diabetic retinopathy, Age-related Macular degeneration.

Introduction

With the advent of technology, our life expectancy has increased significantly. With this advantage, we are experiencing much more diseases related to the ageing process. Eye diseases are also included in these. Early detection and appropriate treatment of these eye diseases are essential to prevent vision loss and promote living quality. Conventional diagnosis methods are tremendously dependent on physicians’ professional experience and knowledge, which leads to a high misdiagnosis rate and a colossal waste of medical data and cost.

A study published in 2014 estimated that diagnostic errors affect at least 5% of US adults (12 million people) per year [1]. More recently, a systematic review and meta-analysis reported that the rate of diagnostic errors causing adverse events among hospitalized patients was 0.7% [2]. Furthermore, diagnostic error is the most critical reason for malpractice litigation in the United States, accounting for 31% of malpractice lawsuits in 2017 [3]. Creating artificial intelligence (AI) programs to identify and analyze diagnostic errors could be an essential step in addressing this problem [4].

AI has integrated with the Ophthalmology to potentially revolutionize the current disease diagnosis and generate a significant clinical impact. It holds the potential to improve patient
and practitioner outcomes, reduce costs by preventing errors and unnecessary procedures, and provide population-wide health improvements.

Artificial intelligence is the fourth industrial revolution in humanity's history [5]. It was proposed in 1956 by Dartmouth scholar John McCarthy; AI "refers to hardware or software that exhibits behaviour which appears intelligent" [6]. The concept suggests that a machine can think and stimulate human intelligence through behaviour such as learning, interpreting and communicating. This concept is referred to as artificial intelligence.

Artificial Intelligence is a vast field of study encompassing many techniques that allow machines to display ever more intelligent behaviour. Subsets of AI are Machine learning (ML) and Deep learning (DL).

ML, which occurred in 1980, refers to a group of mathematical algorithms that learn from experience (data) by mimicking human learning behaviour to perform new tasks. ML can fit complex data sets to extract new knowledge, imitate complex behaviour, and predict and classify based on preliminary data.

Supervised ML (Figure 1) is an approach that requires three labelled data set used for training, validation, and testing. All are defined and labelled by domain experts. Unsupervised ML IS NOT the same level of performance as supervised ML.

![Figure 1: Supervised ML Process. It requires 3 labelled datasets that are used for training, validation, and testing.](image)

Deep learning (DL), which occurred in the 2000s, is a burgeoning technology of ML and has revolutionized the globe of AI (Table 1). These technologies help in several dimensions of modern society, such as objects' recognition in images, real-time languages translation, device manipulation via speech etc.

This paper discusses the applications of AI in the ophthalmic field and review the art of AI for the diagnosis and treatment plan of ophthalmic diseases including cataract, glaucoma, diabetic retinopathy, age-related macular degeneration, retinopathy of prematurity, and future AI modifications and utilization.

Building AI Models

Multiple types of imaging modalities have been used in AI diagnosis, such as radiology images (Computed tomography scan, Magnetic resonance Images, PET-CT) [7], electrophysiological signal records (electrocardiograph [8] and electroencephalogram [9]), visible wavelength images (dermoscopy images and biopsy images), ultrasound images [10], angiography images [11] are some of the imaging modalities used in AI diagnosis.

Developing an AI model includes preprocessing image data, training, validating, testing the model, and evaluating the trained model's performance.

Data Pre-processing

To improve AI prediction efficiency, raw data need to be preprocessed. The preprocessed work includes the following [12,13]: (1) noise reduction. Denoising can promote the quality of the data set and optimize the learning curve. (2) Data integration and standardization: data collected from different sources should be integrated and adjusted to a standard scale. (3) Selecting features and extraction: the most usual features are selected and extracted to improve the learning performances.

Training, Validation, and Test

To attain a standard performance, the data set is frequently divided into two independent subsets; one is for modelling, and the other one is for testing. In most cases, the data in the previous sets will be divided into training and validation sets. The training set is used to pertinent the appropriate parameters of a model. The validation set is used to evaluate how well the model has been instructed and tune the parameters or to assess the performances of the prediction algorithms set.

Cross-validation methods have been widely used to evaluate and optimize algorithms [14]. The most adopted cross-validation is "K-fold cross-validation."

Evaluation

The receiver operating characteristic curve (ROC) is a valuable tool to emboss algorithms' execution. It is created by plotting the distinguish probability for each algorithm across a continuum of the threshold.

Table 1: Artificial Intelligence techniques.

<table>
<thead>
<tr>
<th>Artificial Intelligence</th>
<th>Machine Learning</th>
<th>Deep Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability of computer that can imitate program to function like human brain</td>
<td>Statistical algorithms that enable AI implementation through data</td>
<td>Subset of machine learning which follows neural networking</td>
</tr>
</tbody>
</table>
**AI Application in Ophthalmology**

---

**Cataract**

A cataract is a disease when the lens becomes opacified, usually due to ageing or other causes in young people. Early detection and management can help gain the visual quality of cataract patients and improve their lifestyles. ML algorithms like RF and SVM have been applied to diagnose and grade cataracts using fundus photographs and visible wavelength eye images [5-17]. The risk prediction model for the posterior capsule opacification (PCO) after phacoemulsification has also been built [18].

Senile cataracts can be diagnosed using DL models [19], but a more impressive work is about pediatric cataracts. Long et al. demonstrated a CNN-based computer-aided diagnosis (CAD) framework to grade pediatric cataracts. A multihospital collaboration has been established where a cloud-based platform integrated AI agent has been used. These proposed methods are serviceable for improving cataract screening for a large population.

**Glaucoma**

Glaucoma is a structural and functional disease of the optic nerve characterized by excavation and erosion of the neuroretinal rim that clinically manifests by increased optic nerve head (ONH) cupping. Glaucoma patient suffers from high intraocular pressure, damage to the optic nerve head (ONH), defect in the retina nerve fibre layer (RNFL), and gradual vision loss.

ONH area varies fivefold, but virtually no cup to disc ratio (CDR) defines pathological cupping, hampering disease detection [20]. Li et al. [21] and Ting et al. [22] trained computer algorithms to detect the glaucoma-like disc, defined as a vertical cup-disc ratio (CDR) of 0.7 and 0.8. Investigator has also applied machine learning technique to distinguish glaucomatous nerve fibre layer damage from regular scans on wide-angle OCTs (9×12 mm) [23].

Spectrum domain OCT (SD-OCT) is another critical imaging modality to evaluate cup-disc ratio (CDR). After locating the coarse disc margin by a spatial correlation smoothness constraint, an SVM model is trained to find a patch on OCT images to determine a reference plane to evaluate the CDR. The proposed algorithm can help achieve high segmentation accuracy and a low CDR evaluation error [24].

In glaucoma, retinal ganglion cell axons atrophy occurs in a confined space within the ONH, and ophthalmologists typically depend on low dimensional psychophysical data to delineate the functional consequences of that damage. The outputs from these tests typically support reliability parameters, age-matched normative comparisons, and global summary indices, but a more detailed analysis of this operational data is lacking. Elze et al. [25] developed an unsupervised computer program to analyze visual field (VF) that recognizes clinically relevant patterns of VF loss and assigns a weighting coefficient. This method has proven helpful in detecting early VF loss from glaucoma [26]. Yousefi et al. [27] developed a machine-based algorithm that detected VF progression earlier than these conventional strategies.

Kazemian et al. [28] developed a clinical forecasting tool that uses tonometric and VF data to project disease courses at different target IOPs. Further purification of this tool that integrates other ophthalmic and non-ophthalmic data would be helpful to develop target IOPs and the best strategies to achieve them on a case-by-case basis.

**Diabetic retinopathy**

Worldwide, 600 million people will have diabetes by 2040, with a third having diabetic retinopathy (DR) [29]. DR, a chronic diabetic complication of diabetes, is a vasculopathy that can lead to irreversible blindness [30].

Coupled with timely referral and treatment, screening for diabetic retinopathy (DR) is a globally accepted strategy to prevent blindness. DR screening can be performed by different ophthalmic professionals and under different methods. In the few years, DL has revolutionized the diagnostic performance in detecting DR [31].

The specific abnormalities such as macular oedema [32-35], exudates [32], cotton-wool [33], microaneurysms [36-38], and neovascularization on optic disk [39] can be detected by CML. Based on these hallmarks, the early diagnosis of DR in an automated fashion has been explored [40]. Furthermore, a system focused on timely and effectively proliferative DR (PDR) detection has been developed to ensure immediate attention and intervention [41,42].

Gulshan et al. were the first to describe the application of DL for DR identification. They reported that the method based on DL techniques had very high sensitivity and specificity [43].

Several DL models with impressive performance have been developed for the automated detection of DR [44-46]. Abrámoff et al. [45] showed that a DL system was applicable to evaluate an area under the receiver operating characteristic curve (AUC) of 0.980, with sensitivity (96.7%) and specificity (87.0%), for the delineate of referable DR (defined as moderate non-proliferative DR or worse, with or without diabetic macular oedema (DMO) on Messidor-2 data set. Garcia and Leng [44] showed an AUC of 0.97 using cross-validation on the same data set and 0.94 and 0.95 in two independent test sets (Messidor 2 and E-Ophtha).

Some studies applied DL to automatically stage DR through fundus images [45-48], making up for the deficiency of Gulshan's study that they only studied refeerable DR. Still, there is no comparable data on vision-threatening DR or other DR stages.

Although several groups have demonstrated promising results using DL systems on available data sets, the DL systems were not tested in real-world DR screening programmes.

Ting et al. [22] showed a clinically acceptable diagnostic performance of a DL system, developed, and tested using the
The timeline of selected articles on Deep Learning in AMD

Age-related Macular Degeneration
AMD is a major cause of irreversible vision loss in older people globally. The Age-Related Eye Disease Study (AREDS) classifies AMD stages into none, early, intermediate, and late AMD using drusen, pseudodrusen, fluid or atrophy [49,50]. With the ageing population, there is an urgent clinical need to have a robust DL system to screen these patients.

The goal of using ML and DL systems is to automatically identify AMD-related lesions to improve AMD diagnosis and treatment. Drusen regression, an anatomic endpoint of moderate AMD and the onset of neovascular AMD, can be evaluated through the specifically designed, fully automated, ML-based classifier. Detection of drusen [51,52] fluid [52,53] reticular pseudo drusen [54] and geographic atrophy [55] from fundus images and SD-OCT using ML 54 has been studied. The accuracy is usually over 80% [51,54-56] and the agreement between the models and retina specialists can reach 90%.

Ting et al. [22] reported a clinically acceptable DL system diagnostic performance in detecting referable AMD. This DL system was trained and tested using 108558 retinal photographic images from 38 189 patients. Fovea-centred images without macula segmentation were studied in this study. For the other two studies [57-59], DL systems were constructed using the AREDS data set, with many referable AMD (intermediate AMD or worse).

Using fivefold cross-validation, Burlina et al. [58] reported diagnostic accuracy from 88.4% to 91.6%, with an AUC of 0.94 and 0.96. Unlike, the authors pre-segmented the macula region before training and testing, with an 80/20 split between the testing and training in every fold [22]. Both AlexNet and OverFeat have been used in the DL architecture, with AlexNet yielding better performance. Utilizing the same AREDS data set, Grassmann et al. (Table 2) [59] reported a sensitivity of 84.2% in detecting any AMD. To train different models, the authors used six convolutional neural networks in this study—AlexNet, GoogleNet, VGG, Inception-V3, ResNet and Inception-ResNet-V2.

Bogunovic et al. build a data-driven interpretable predictive model to predict the progression risk in intermediate AMD [52]. Automated image analysis steps were applied for identifying and characterizing each drusen at baseline, and their progression was monitored at a follow-up visit. For this characterization and analysis, they developed an ML method based on survival analysis to delineate a risk score and speculate the incoming regression of individual drusen.

Using ML to prognosticate anti-vascular endothelial growth factor (anti-VEGF) requirements in eye diseases such as neovascular AMD and PDR can alleviate patients' economic burden and facilitate resource management. Bogunovic et al. used OCT images of patients with low or high anti-VEGF injection requirements for prediction using the ML system. A solid AUC from 70% to 80% was achieved for treatment requirement prediction [60]. Prahs et al. instructed a DCNN neural network by OCT images to facilitate decision-making regarding anti-VEGF injection [61], and the outcomes were better than that of CML [60]. These studies are an essential step toward image-guided prediction of treatment intervals in neovascular AMD or PDR management.

Treder et al. automatically establish a model to detect exudative AMD from SD-OCT [62] automatically. For research studies based on fundus images, images with AMD were assigned into four classes (no evidence of AMD, early-stage AMD, moderate-stage AMD, and CNV AMD), [63] or 2-class classification (no or early-stage AMD and moderate or advanced stage AMD) [58]. The diagnostic accuracy is better in the 2-class classification of the recent studies. The DCNN appears to detect a screening function in these experiments, and the performance is comparable with related healthcare professionals.

### Table 2: The timeline of selected articles on Deep Learning in AMD (2017-2020).

<table>
<thead>
<tr>
<th>Classification of referable and non-referable AMD using CFP [22,33,58]</th>
<th>9-step AREDS severity scale using fundus images and multi-task strategy [51]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification of 4-class AMD severity using CFP and universal features [63]</td>
<td>Classification of AMD severity scale using fundus images of both eyes [65]</td>
</tr>
<tr>
<td>Dry-AMD detection using OCT images [49]</td>
<td>GA detection in color fundus images [67]</td>
</tr>
<tr>
<td>AMD detection using OCT images [64]</td>
<td>Multiple clinical referral suggestions on OCT images [69]</td>
</tr>
<tr>
<td>AMD prognosis in a wide time interval using fundus images [66]</td>
<td></td>
</tr>
</tbody>
</table>

**DM, Choroidal neovascularisation, and other Macular diseases**

OCT has had a tremendous effect on diagnosing and managing macular diseases, specifically wet ARM and diabetic macular oedema. OCT also provides near-microscopic images of the retina in vivo with quick acquisition protocols and astounding structural details that cannot be seen using other ophthalmic examination techniques.

From a DL perspective, macular OCTs possess several attractive qualities as a modality for DL. The augmenting growth in the number of macular OCTs routinely collected worldwide. This sizeable number of OCTs is required to instruct DL systems where having many tutoring examples can aid in coupling many-layered networks with millions of parameters. Macular OCTs have dense three-dimensional structural data-based information that is usually consistently captured. The macular field and the foveal fixation are consistent from one volume scan to another. This significantly lowers the computer vision task's complexity and allows networks to reach meaningful performance with smaller data sets. OCTs provide structural detail that is not usually visible using traditional imaging modalities and provide a promenade for uncovering innovative biomarkers of the disease.
One of the initial applications of DL to macular OCTs was in the automated categorization of AMD [64-73]. Approximately 100000 OCT B-scans were used to train a DL classifier based on VGG-16 to achieve an AUC of 0.97 [64]. Several studies used a technique known as transfer learning, where a neural network is trained on ImageNet and ensuing then trained on OCT B-scans for retinal disease classification [65-68].

DeepMind and the Moorfields Eye Hospital combine the power of neural networks for both segmentation and classification tasks utilizing an encoded AI framework. A dismemberment network is first used to delineate fifteen different retinal morphological features and OCT acquisition artefacts during this approach. The output of this network is then moved to a classification network which makes a frame of referral triage decision from four categories (urgent, semi-urgent, routine, observation) and classifies the presence of ten different OCT pathologies as choroidal neovascularisation (CNV), macular oedema without CNV, drusen, geographic atrophy, epiretinal membrane, vitreomacular traction, full-thickness macular hole, partial-thickness macular hole, central serous retinopathy and ‘normal’) [69]. Using this approach, the Moorfields-DeepMind system reports performance on par with experts for these classification tasks (although in a retrospective setting).

**Retinopathy of prematurity (ROP)**

ROP is one of the foremost causes of childhood blindness, with an annual incidence of ROP-related blindness of 32 000 worldwide [73].

There are two main barriers to effective implementation of ROP screening: first, the diagnosis of ROP is subjective, with significant variability of opinion among the examiners in the diagnosis leading to inconsistent application of evidence-based interventions [74]; and second, there are limited trained examiners in many regions of the world [75]. Telemedicine is emerging as a viable model to cover a large geographical area for addressing the problem by allowing a single physician to examine infants over a sizeable geographical area virtually.

There have been some early attempts to use DL for automated diagnosis of ROP [76,77], which could potentially address for implementation barriers of ROP screening.

Brown et al. [76] studied the results of a fully automated DL system that could evaluate the essential features of severe ROP, with an AUC of 0.98 compared to a consensus reference standard diagnosis with a combining image-based diagnosis and ophthalmoscopy. In ROP diagnosis, the i-ROP DL system has agreed with the consensus that it is more frequent (six out of eight experts). Subsequent work confirmed that the i-ROP DL system is helpful to formulate a severity score for ROP that demonstrates objective monitoring of disease progression, regression, and response to treatment [78]. When compared with the same set of 100 images ranked, the algorithm had 100% sensitivity and 94% specificity in detecting pre-plus or worse disease.

**AI in refractive surgery**

Refractive surgery has undergone rapid advancements in the last decades with sound, visual effects and long-term safety [79]. There are several refractive surgery types available both in the cornea and lens. Among all the refractive surgeries, the most popular surgeries are- excimer laser photorefractive keratectomy (PRK), laser-assisted in situ keratomileuses (LASIK), and small incision lenticule extraction (SMILE) surgery that uses femtosecond laser. Before going through these procedures, the exclusion of keratoconus is significant.

Due to the occult onset of keratoconus, its early stages are often challenging to detect. Furthermore, it dramatically affects the patient’s vision with a risk of blindness. Many ways from different perspectives have proposed using machine learning technology to assist in the research and diagnosis of keratoconus [80].

To improve the diagnostic accuracy of keratoconus, keratectasia, and other related diseases, different screening equipment are often combined during the clinical diagnosis the process to develop algorithms for multi-source diagnostic methods, which include anterior segment OCT devices, optical aberration measuring instruments, confocal microscopy, and in vivo measurement of corneal biomechanics. The results of different functional instruments have different meanings, and there are significant differences in the results of different instruments with the same functional category, analysing these inspection parameters is complex and complicated. Also, different population-based studies have shown that ethnic origin influences the keratoconus incidence and corneal physiological parameters of people in different regions [81]. A compatible, robust, and convenient model is necessary to overcome these different issues.

A study was performed on a retrospective analysis to develop diagnostic algorithms from a single corneal topographic device to the cross-platform data of three various topographic device sources [82]. Another research, where a diagnostic model combined with corneal biomechanical measurement has also proven high diagnostic performance [83,84].

Recently, a study was done among 2000 participants based on the Scheimpflug corneal tomography parameters to establish a model that can diagnose subclinical keratoconus with high accuracy. They used the support vector machine (SVM) and gradient boosted decision tree (GBDT), an iterative machine learning algorithm composed of multiple decision trees that screens attribute features with larger weights, to construct a subclinical keratoconus diagnosis model, and performed a 10-fold cross-validation to verify the accuracy. The model achieved a 95.53% diagnostic accuracy. The accuracy of the model in distinguishing subclinical keratoconus from the normal cornea was 96.67%, and the accuracy in distinguishing keratoconus from the normal cornea was 98.91% [85]. The majority of the research mainly focuses on the analysis of images and data. Some of the most commonly used methods are support vector machine (SVM) [86] decision tree (DT) [87],
multilayer perceptron (MLP), radial basis function (RBFNN) [88], and convolutional neural networks (CNN) [89].

It is essential to ensure the accuracy and predictability of corneal refractive surgery; the risk of overcorrection or undercorrection is reduced [90,91]. Previous nomogram reports of adjusting magnitudes of spherical equivalent and astigmatism before PRK and LASIK have shown that the use of multiple regression analysis to establish a nomogram model that can consider numerous factors, including age, diopter, and corneal curvature, improves the accuracy of PRK and LASIK surgery [92-94]. Some more factors influencing the outcome of surgery have also been observed, such as temperature, humidity [95], wind speed, and air pressure [96]. Unlike PRK and LASIK, the nomogram adjustment of SMILE surgery needs to consider more factors and depends more on the experience of the surgeon. A study on 1146 cases that underwent SMILE surgery with ideal postoperative results. The multilayer perceptron algorithm was used to train the artificial neural network model to predict the SMILE nomogram and conduct clinical control experiments for validation. The study compared the outcomes of the surgeon group with the machine learning group in terms of safety, efficacy, and predictability. The outcomes of all aspects of the machine learning group have reached the level of experienced surgeons or even better [97]. So, for better visual outcomes, refractive surgeons should actively embrace the convenience brought by artificial intelligence to help this discipline develop faster and more accurately.

**Future of AI Application in Ophthalmology**

Most studies regarding the intelligent diagnosis of eye diseases focused on binary classification problems, whereas in a clinical setting, visiting patients suffer from multi categorical retinal diseases. For instance, a model trained to evaluate AMD will fail to consider a patient with glaucoma as diseased because the model only can discriminate between AMD from non-AMD.

Choi and his colleagues carried out work applying DL to automatically detect multiple different retinal diseases with fundus photographs. When only standard and fundus images for DR were used in the proposed DL model, the classification accuracy was 87.4%. However, the accuracy is 30.5% when including all ten categories [98]. The model's accuracy has declined while the number of diseases is increased. To further improve the utility of AI in clinical practices, we should make more efforts to build intelligent systems that can help detect different retinal diseases with high accuracy.

Additionally, a single abnormality detected from one imaging technique cannot always guarantee the correct diagnosis of a specific retinal disease (e.g., DR or glaucoma) in clinical practice. Multimodal clinical photographs, such as optical coherence tomography angiography, visual field, and fundus images should be integrated to build a generalized AI system for more reliable AI diagnosis.

However, the need for a massive amount of data remains the most fundamental problem. Images with severe diseases or rare diseases are particularly insufficient. The population characteristics, various systematic diseases, and the various disease' phenotypes should be considered when selecting input data.

Above all, by building interpretable systematic AI platforms using sufficient high-quality and multimodal data and advanced techniques, we can enhance the applicability of AI in our clinics. We hope we might make it possible to adopt intelligent systems in the inevitable clinical work process with reliable accuracy.

**Conclusions**

DL and ML are state-of-the-art AI that has revolutionized the AI field. DL has shown clinically acceptable diagnostic performance for ophthalmology in detecting many retinal diseases, particularly in DR, ROP and AMD. Research with many data is needed in evaluating the clinical application and cost-effectiveness of different DL systems in clinical practice.

**Reference**

10. Shuang Yu, Tan KK, Ban LS, et al. Feature extraction and


39. Yu S, Xiao D, Kanagasingam Y. Automatic detection of neovascularization on optic disk region with feature ex-


73. Blencowe H, Moxon S, Gilbert C. Update on blindness due to retinopathy of prematurity globally and in India. Indian Pediatr 2016; 53: S89-S92.
