Artificial Intelligence Applications in Diabetic Retinopathy: What We Have Now and What to Expect in the Future

Mingui Kong¹, Su Jeong Song¹,²

¹Department of Ophthalmology, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul; ²Biomedical Institute for Convergence (BICS), Sungkyunkwan University, Suwon, Korea

Diabetic retinopathy (DR) is a major complication of diabetes mellitus and is a leading cause of vision loss globally. A prompt and accurate diagnosis is crucial for ensuring favorable visual outcomes, highlighting the need for increased access to medical care. The recent remarkable advancements in artificial intelligence (AI) have raised high expectations for its role in disease diagnosis and prognosis prediction across various medical fields. In addition to achieving high precision comparable to that of ophthalmologists, AI-based diagnosis of DR has the potential to improve medical accessibility, especially through telemedicine. In this review paper, we aim to examine the current role of AI in the diagnosis of DR and explore future directions.

Keywords: Artificial intelligence; Diabetes; Retinopathy; Telemedicine

INTRODUCTION

Recent research estimates that the prevalence of diabetes mellitus (DM) among individuals aged 20 to 79 was approximately 485 million in 2021, and this number is expected to rise. By 2050, the number of people with DM is projected to reach approximately 1.31 billion [1]. In South Korea, the first epidemiological study on diabetes was conducted in 1971, revealing a prevalence of 1.5% in adults aged 30 and older [2]. However, more recent studies have shown that the prevalence rate of diabetes has increased to 12.4%, as determined by fasting plasma glucose levels of 126 mg/dL or higher in 2018. Notably, about 30% of adults aged 65 and older have diabetes [3].

As the prevalence of DM increases, so too does the risk of diabetes-related complications. Among these complications, diabetic retinopathy (DR) is one of the main causes of blindness worldwide [4]. In South Korea, the prevalence of DR and vision-threatening DR (VTDR) among DM patients was found to be 15.8% and 4.6%, respectively [5]. However, it was reported that only 30% of DM patients received annual dilated fundus examinations by ophthalmologists in 2013, which is the standard method for accurately diagnosing DR [5]. There are several reasons for the low annual DR screening rate among DM patients. Firstly, DR often manifests without noticeable symptoms in its early stages, leading patients to overlook the importance of regular ophthalmic examinations. Second, patients often hesitate to undergo dilated fundus examinations. Furthermore, the misleading belief that good blood glucose control can eliminate the risk of DR development or progression induces patients to skip their regular fundus examinations. Although educating patients about the importance of screening for diabetic complications is necessary to address these issues, diverse efforts have...
been made to improve the efficiency and convenience of regular eye examinations.

DR has been chosen as the first disease to be tested by artificial intelligence (AI), particularly in the context of medical imaging analysis, for several reasons. In addition to the public health impact, the availability of imaging data, potential for automation in screening, early successes, and validation of applying AI to DR screening have demonstrated promising results, with some AI systems achieving accuracy levels comparable to or even exceeding that of human experts. These factors have combined to make DR a logical and strategic starting point for the application of AI in healthcare, particularly in the domain of disease screening and diagnosis. The focus on DR not only addresses a critical need within global health, but also demonstrates the potential of AI to transform medical diagnostics across a range of conditions. Therefore, this paper explores the status of AI-based DR diagnosis and discusses the existing challenges and future directions to pursue.

DEEP LEARNING APPLICATION IN DIABETIC RETINOPATHY

Diagnosis and access to high-risk groups

In the field of medicine, AI has become integral in supporting clinical practice through the application of machine learning (ML) and deep learning (DL) techniques [6]. DL, which is a specialized branch of ML, utilizes a complex architecture of neuron-like nodes arranged in layers that mimic the structure of the mammalian brain. This has given rise to the term “Convolutional Neural Networks” (CNNs) for systems that specialize in image recognition and classification [7]. In CNNs, the initial layers detect simple features such as edges, lines, and colors, while deeper layers discern more complex patterns. Additionally, networks of fully connected hidden layers are situated between the input and output layers to enhance processing. Furthermore, mathematical operations such as convolution and pooling are employed by the software to assign varying levels

Fig. 1. A diagram illustrating the classification of diabetic retinopathy using Convolutional Neural Network (CNN). DR, diabetic retinopathy; NPDR, nonproliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy.
of significance to different features within an image. CNNs are particularly adept at analyzing visual data and are frequently used for this purpose. They can recognize patterns autonomously, without the need for manual feature engineering, which makes DL through CNNs especially effective for automated image analysis [8].

When diagnosing DR, a binary approach can be used to determine the presence or absence of the condition. A more nuanced diagnosis is possible with the use of the Early Treatment Diabetic Retinopathy Study (ETDRS) scale, which measures the severity of DR [9]. AI algorithms that utilize CNNs are designed to automatically detect characteristic lesions, such as microaneurysms and retinal hemorrhages, commonly associated with DR (Fig. 1). The performance of these algorithms is evaluated using various metrics, including accuracy, specificity, area under the curve (AUC), sensitivity, precision, F1 score, and Kappa score. These indicators are used to evaluate algorithms’ diagnostic accuracy and classification performance.

Initially, most AI-driven DR diagnosis efforts were centered on binary classifications, such as determining the presence or absence of DR, yielding accuracy or sensitivity rates that were comparable to or even surpassed those of human graders [10-14]. The accuracy of these diagnoses varies widely, with some studies reporting rates as low as 65% and others as high as 98.2%. However, to fully address the risk of vision loss among individuals with diabetes, it is also necessary for automatic DR screening algorithms to assess the severity of DR. In a recent study, an artificial neural network with radiomic features was employed after imaging enhancement and increasing contrast, with the goal of predicting DR stages. The performance of these algorithms is evaluated using various metrics, including accuracy, specificity, area under the curve (AUC), sensitivity, precision, F1 score, and Kappa score. These indicators are used to evaluate algorithms’ diagnostic accuracy and classification performance.

Using AI for predicting progression and treatment guidelines in DR

AI has the capability to predict future outcomes by recognizing patterns in past data, often with greater efficiency than experienced human experts. This is due to AI’s ability to process and learn from vast quantities of historical data, sometimes revealing insights that might be missed by human observation. Dai et al. [24] conducted DL using 717,308 fundus images obtained from 179,327 participants. They aimed to predict the time to DR progression, achieving concordance indices ranging from 0.754 to 0.846 and integrated Brier scores ranging from 0.153 to 0.241 up to 5 years. When applied to a real-world cohort, they predicted a potential increase in the screening interval from an average of 12 to 31.97 months [24]. Prahs et al. [25] trained a deep CNN to predict the need for anti-vascular endothelial growth factor (VEGF) treatment using macular OCT scans, without human intervention. Additionally, Ye et al. [26] reported that an AI-based system for predicting post-treatment visual acuity following anti-VEGF treatments showed promising results. As these models continue to improve, they offer the potential to answer critical questions, such as which OCT findings suggest the possibility of improvement without anti-VEGF therapy, or which specific type of anti-VEGF treatment is most appropriate. They could also provide recommendations on treatment intervals that optimize therapeutic effectiveness while minimizing interventions. The use of AI in this way for treatment and prognosis predictions can provide valuable insights into patient care optimization, potentially leading to cost savings in overall healthcare expenditures.

U.S. Food and Drug Administration-approved AI systems for DR

The advancement of AI technology has extended beyond research laboratories and into the marketplace. Several AI systems designed for diagnosing DR are now commercially available.
Two notable systems that have received U.S. Food and Drug Administration (FDA) approval are IDx-DR by Digital Diagnostics and EyeArt by Eyenuk Inc. (Los Angeles, CA, USA). IDx-DR was granted FDA approval in 2018, allowing it to assess the severity of DR and recommend further medical evaluation for adult diabetic patients using retinal photographs. This was the first AI-driven DR diagnostic product to receive FDA approval and it has shown an impressive sensitivity and specificity, both nearing 90%, by employing two core algorithms. EyeArt has demonstrated a sensitivity of 96% and a specificity of 88% in detecting mild DR, and for more severe cases of VTDR, it has achieved a sensitivity of 92% and a specificity of 94%. In addition to its accuracy, this algorithm is highly efficient and capable of analyzing data from 100,000 individuals in just 45 hours [27,28].

CHALLENGES AND CONSIDERATIONS

Data equality and standardization
To train an AI algorithm, the first step is constructing a composite image dataset, with the essential requirement being the inclusion of a sufficient number of ground truth images. The subsequent step involves effectively and accurately annotating or labeling the images to classify pathologies. For the automatic detection of DME, it is necessary to include not only fundus photographs but also OCT images to develop a comparable algorithm [29]. Therefore, when integrating AI systems into clinical practice, we must tackle challenges that extend beyond obvious pathologies such as bleeding or exudates. We require AI systems that can capture more subtle aspects, such as shape, fuzzy borders, and changes in contrast, since these features contain critical clinical information. However, differentiating high-risk groups with clinically significant macular edema or proliferative features from the early stages can be challenging. This is because the interpretation of DR in early-stage cases can vary, even among experienced experts. Consequently, AI-driven DR detection in cases of minimal DR has been reported to achieve only a moderate sensitivity of 57% [30]. Furthermore, DL requires vast amounts of data, and the quality of this data significantly affects the model’s performance. Poor data quality, characterized by blurred images, artifacts, incorrect labels, and insufficient data for rare diseases, can present challenges. Additionally, AI cannot inherently consider the psychological and social aspects of patients, which physicians routinely integrate into their assessments.

Another critical consideration involves potential biases, specifically inequality of opportunity and inequality of odds. The performance of AI can vary significantly depending on the quality of the data used during training. Racial differences in ocular characteristics mean that the choice of datasets can dramatically affect an AI system’s reliability. Therefore, it is imperative to train AI using high-quality datasets that encompass a diverse range of patients in terms of age, sex, and ethnicity. Diverse datasets should also be used to accurately measure algorithmic bias. However, the lack of publicly available datasets in many studies poses a challenge to the trustworthiness of their reliability. Recent guidelines such as Consolidated Standards of Reporting Trials (CONSORT-AI) and Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT-AI) are prompting researchers to disclose their methods of data integration and acquisition in clinical settings to address this issue [31].

Legal, regulatory, and ethical aspects
We must also consider the legal and ethical aspects of using AI for diagnosis and treatment. This encompasses potential legal issues such as patient privacy, medical malpractice, and product liability. Protecting patient privacy is critical, as data sharing could potentially compromise it. Although diverse datasets are essential for training AI systems, they also risk exposing patients’ personal information. Federated learning offers a potential solution by training separate datasets with the same algorithm, which eliminates the need to pool data [32,33]. Furthermore, challenges related to misdiagnosis and legal liability are yet to be resolved. Even with a device that has 90% sensitivity and specificity, there is a theoretical 10% chance of false positives and false negatives. Consequently, AI cannot completely replace the gold standard of dilated eye examinations for diagnosing DR at this time. Additionally, when a patient does not have DR but suffers from other conditions such as glaucoma or age-related macular degeneration, an AI system trained solely to diagnose DR may result in delayed diagnoses of these other diseases. For example, the developers of IDx-DR have chosen to obtain medical malpractice insurance to address liability concerns related to misdiagnosis for contexts where it is used in primary care settings without eye care providers [34].

Cost-effectiveness
It is essential to assess whether new AI diagnostic systems are cost-effective in comparison to the standard of care. A study in Singapore demonstrated that training a semi-automated DL system to detect referable DR incurred a cost of $62 per person per year, while the cost of human assessment was $77. According to
this research, with the rising prevalence of diabetes, cost savings of $15 million are projected by the year 2050. Additionally, another study indicates that the implementation of the EyeArt 2.0 system could lead to a 23% reduction in costs over a 5-year period when compared to dilated fundus examinations [35,36].

**FUTURE DIRECTIONS**

**Diagnosis of eye conditions beyond DR**

In addition to DR, AI has been explored for diagnosing a range of ocular diseases, including age-related macular degeneration—one of the leading causes of blindness—and other common conditions such as epiretinal membrane, retinal vascular occlusions, and glaucoma [37]. A study that employed AI to diagnose these conditions using fundus photographs showed promising results. Initially, the AI’s training yielded an AUC of 0.952. Further improvements were made through domain-adaptive training, which increased the AUC to 0.964. A particularly notable aspect of this study was the inclusion of 789 fundus photographs representing cases with two or more diagnoses. The AI system demonstrated impressive accuracy, correctly predicting all diagnoses in 61.6% of these complex cases without any false negatives, and it had a 96.3% success rate in making at least one correct diagnosis prediction [38].

**Predicting systemic risk based on eye examinations**

Currently, AI applications in DR primarily focus on the diagnosis of DR and DME. However, looking to the future, we can anticipate the development of guidelines for predicting disease progression and treatment based on retinal structures and microvascular changes, as previously mentioned. Moreover, by leveraging ocular data, AI could predict the status of systemic diseases, such as cardiovascular, renal, and neuropathic conditions, and potentially forecast their progression. Retinal observation offers a non-invasive assessment of blood vessels, making it a valuable tool as a risk prediction model for cardiovascular diseases, including strokes [39-41]. Additionally, some studies have successfully identified and predicted the progression of chronic kidney disease and estimated glomerular filtration rates using fundus images and clinical data [42,43]. Researchers are also actively exploring the use of fundus and OCT findings to predict the development of neurodegenerative diseases in diabetic patients, including cognitive dysfunctions such as Alzheimer’s disease, considering the embryological origin of the retina from the brain [44-46]. Comprehensive analyses of retinal imaging, systemic parameters, and other serum biomarkers using AI may provide deeper insights into these areas, perhaps even surpassing the conclusions that human intelligence can derive.

**Providing personalized medicine for individuals**

Personalized medicine, also known as personalized treatment, is designed to tailor precision medicine to an individual’s genetics, environment, lifestyle, biomarkers, and other factors [47]. The objective is to improve patient outcomes by providing effective, efficient, and safe targeted interventions [48]. A study that utilized ML to analyze gene expression data successfully predicted responses to anticancer chemotherapy in 175 cancer patients, with an impressive prediction accuracy of over 80% [49]. In another study, electronic health records and AI were used to predict responses to various antidepressants in a cohort of 17,556 patients, yielding strong prediction performance [50]. In the field of DR, it is expected that AI could significantly improve the delivery of precision medicine by taking into account each patient’s unique characteristics. These include the type and duration of diabetes, trends in hemoglobin A1c levels, kidney function, current retinal findings, and retinal vascular morphology. This comprehensive, AI-enabled approach has the potential to revolutionize personalized medicine in the management of DR.

**Integrating telemedicine and AI for DR screening**

Telemedicine refers to a broad spectrum of health platforms for patients, physicians, and other medical service providers, transcending the constraints of location and time. The primary objectives of telemedicine include enhancing access to treatment, increasing follow-up rates, facilitating early intervention, and reducing the cost of care. Studies on the integration of AI-based platforms for DR with traditional telemedicine processes have
demonstrated promising outcomes, achieving sensitivity and specificity rates of 90% or higher. Furthermore, this method required 75% less time for evaluation compared to assessments conducted by an ophthalmologist alone.

Self-diagnosis and the monitoring of disease progression at home represent another significant facet of telemedicine, which is already in practice, especially in the management of blood pressure and blood sugar levels. Telemonitoring employs home monitoring devices to observe the onset and progression of diseases. Several commercial devices, including home OCT systems, have been launched for this purpose (Fig. 2) [51-57].

**Generative large language models in DR**

Large language models (LLMs), exemplified by chatbots such as ChatGPT (OpenAI, San Francisco, CA, USA), represent a subtype of DL systems [58]. LLMs demonstrate versatility across various societal domains, with significant anticipated roles in the field of medicine. From simple consultations and facilitating appointment scheduling to providing advice on diseases, LLMs are expected to serve as decision-making models between doctors and patients [59]. Gopalakrishnan et al. [60] explored the application of LLMs in the domain of DR. They conducted an analysis of DR screening timing through 20 hypothetical case scenarios using inputs from five clinicians and three AI applications. They reported fair inter-rater reliability (κ = 0.21–0.4) between “major clinician response” and “major AI response” [60]. Such endeavors are deemed to be highly beneficial tools for patients who have difficulty accessing ophthalmologists, and they provide insight into the potential evolution of LLMs in the field of DR.

**CONCLUSIONS**

In conclusion, the use of AI in diagnosing DR has seen significant advancements, benefiting both patients and physicians. However, there remain issues to be resolved, including legal and ethical considerations. Despite these challenges, AI is on track to become a significant metric in the diagnosis and treatment of DR. It is also expected to play a vital role as a tool for delivering personalized care for diabetes and other diseases. Finally, when combined with telemedicine, AI has the potential to remove barriers to all aspects of eye care for individuals with diabetes.

**CONFLICTS OF INTEREST**

No potential conflict of interest relevant to this article was reported.

**ACKNOWLEDGMENTS**

This work was supported by the KBSMC-SKKU Future Clinical Convergence Academic Research Program, Kangbuk Samsung Hospital & Sungkyunkwan University, 2023.

**ORCID**

Mingui Kong  https://orcid.org/0000-0003-2277-2207
Su Jeong Song  https://orcid.org/0000-0003-2110-1874

**REFERENCES**

9. Early Treatment Diabetic Retinopathy Study Research Group. Grading diabetic retinopathy from stereoscopic color fundus photographs: an extension of the modified Airlie House clas-
Artificial Intelligence and Diabetic Retinopathy


47. Quazi S. Artificial intelligence and machine learning in precision and genomic medicine. Med Oncol 2022;39:120.


57. Bhaskaranand M, Cuadros J, Ramachandra C, Bhat S, Nit-

