

The Utility of ChatGPT in Diabetic Retinopathy Risk Assessment: A Comparative Study with Clinical Diagnosis

Keerthana Raghu^{1,*}, Tamilselvi S^{2,*}, Chitralekha S Devishamani¹, Suchetha M²,
Ramachandran Rajalakshmi³, Rajiv Raman¹

¹Shri Bhagwan Mahavir Vitreoretinal Services, Sankara Nethralaya, Chennai, Tamil Nadu, India; ²Centre for Health Care Advancement, Innovation, and Research Department, Vellore Institute of Technology, Chennai, Tamil Nadu, India; ³Department of Diabetology, Ophthalmology and Epidemiology, Madras Diabetes Research Foundation & Dr. Mohan's Diabetes Specialities Centre, Chennai, Tamil Nadu, India

*These authors contributed equally to this work

Correspondence: Rajiv Raman, Shri Bhagwan Mahavir Vitreoretinal Services, Sankara Nethralaya, Sankara Nethralaya (Main Campus), No. 41 (Old 18), College Road, Chennai, Tamil Nadu, 600006, India, Tel +91-444288708, Fax +91-44-28254180, Email rajivpgraman@gmail.com

Purpose: To evaluate the ability of an artificial intelligence (AI) model, ChatGPT, in predicting the diabetic retinopathy (DR) risk.

Methods: This retrospective observational study utilized an anonymized dataset of 111 patients with diabetes who underwent a comprehensive eye examination along with clinical and biochemical assessments. Clinical and biochemical data along with and without central subfield thickness (CST) values of the macula from OCT were uploaded to ChatGPT-4, and the response from the ChatGPT was compared to the clinical DR diagnosis made by an ophthalmologist.

Results: The study assessed the consistency of responses provided by ChatGPT, yielding an Intraclass Correlation Coefficient (ICC) value of 0.936 (95% CI, 0.913–0.954, $p < 0.001$) (with CST) and 0.915 (95% CI, 0.706–0.846, $p < 0.001$) (without CST), both situations indicated excellent reliability. The sensitivity and specificity of ChatGPT in predicting the DR cases were evaluated. The results revealed a sensitivity of 67% with CST and 73% without CST. The specificity was 68% with CST and 54% without CST. However, Cohen's kappa revealed only a fair agreement between ChatGPT predictions and clinical DR status in both situations, with CST (kappa = 0.263, $p = 0.005$) and without CST (kappa = 0.351, $p < 0.001$).

Conclusion: This study suggests that ChatGPT has the potential of a preliminary DR screening tool with further optimization needed for clinical use.

Keywords: ChatGPT, artificial intelligence, diabetic retinopathy, diabetes

Introduction

As per 2021 estimates, there are about 529 million people living with diabetes worldwide, and by 2050, the numbers are projected to increase by 1.31 billion, respectively.¹ Diabetic retinopathy (DR) is a common microvascular complication of diabetes that can lead to preventable vision loss if not detected and managed early.² Due to the rapid increase in the diabetes population, the burden of DR is expected to grow, exacerbating health disparities in at-risk populations. As per 2020 estimates, the number of adults with DR, Vision Threatening Diabetic Retinopathy (VTDR) and Clinically Significant Macular Edema (CSME) was 103.12 million, 28.54 million and 18.83 million, by 2045 the numbers are projected to increase by 160.50 million, 44.82 million, 28.61 million, respectively.³ With the rising global burden of DR, early detection and timely treatment of DR are paramount, making effective screening methods a clinical priority.

There are several barriers for DR screening especially for people living in the under-resourced settings where the burden of diabetes and its complications are often the highest. The major barriers of DR screening include limited access to DR screening due to financial obstacles, scarce human expertise, inadequate health literacy, and awareness of diabetes management.^{4,5}

Barriers in access to DR screening could delay diagnosis and management, leading to progression of the disease and need for medical and surgical interventions, with guarded visual prognosis, which lead to socioeconomic burdens on the patient and healthcare system.⁶ This situation presents a clear need for novel, scalable methods to facilitate early DR detection.

The use of artificial intelligence (AI) tools has gained popularity in health care for their ability to analyze vast amounts of data rapidly and draw inferences from medical data.^{7,8} AI, particularly machine learning and deep learning algorithms, has shown promise in DR screening. Multiple studies reported their ability to detect referable DR (worse than mild DR) in automatic DR screening using AI, ranging from 87.2% to 97.5% in sensitivity and from 0.936 to 0.991 in AUC.^{9,10} Therefore, AI-based DR screening algorithms have the potential to reach or even outperform clinical experts and provide health care to large populations, especially in less-developed areas.^{10,11}

Building upon the substantial progress made in the field of AI-based healthcare applications, our study takes a closer look at the potential of ChatGPT, which is emerging as a notable AI model. ChatGPT, a Natural Language Processing (NLP) model developed by OpenAI, has shown promising results in aiding medical education and decision-making.^{7,12} In this study, we used ChatGPT-4 to generate a medical report suggesting the risk of DR based on clinical and biochemical parameters. We employed anonymized data of 111 patients with diabetes for evaluating the prediction accuracy of ChatGPT-4 against clinical diagnoses, while also examining the reliability of the responses from ChatGPT-4. The outcomes of this research could provide insights into the future application of ChatGPT in the management of diabetic complications.

Materials and Methods

Study Design and Population

This retrospective study was conducted in October 2023 using an anonymized dataset of 111 individuals with diabetes who participated in a DR research project (Institution Approval Number 564–2016 P, approved by Vision Research Foundation, Chennai, India) conducted at a tertiary eye care center. All participants provided an informed consent, and the study adhered to the principles of the Declaration of Helsinki.

The population sample for this study comprised participants diagnosed with diabetes, who had undergone comprehensive eye examinations along with clinical and biochemical assessments. The use of this group was particularly relevant to the study due to their potential susceptibility to DR. Given that the study was retrospective, it ensured that existing clinical diagnoses were available to validate the AI predictions.

Data Collection

The clinical and biochemical parameters included age, sex, duration of diabetes, blood pressure, glycated hemoglobin (HbA1c), serum total cholesterol, low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides, microalbuminuria, and central subfield thickness (CST) of the macula from optical coherence tomography (OCT).

To minimize bias and maintain the privacy of the participants, the data were anonymized. This was followed by uploading the anonymized data of each patient to the ChatGPT-4 online AI interface, three separate times by an operator without diabetes or ophthalmology background. This repetitive process was designed to assess the consistency of the AI-generated responses. The generated three responses were meticulously documented for further analysis. Table 1 presents a sample of prompts provided to ChatGPT, along with three corresponding responses obtained from ChatGPT. Same process was repeated for the second time with all the clinical and biochemical parameters excluding CST values. This was done to check the model prediction ability with and without CST values, considering the benefits and challenges in using OCT.

Data Analysis

The ChatGPT responses (all the 3 responses for an individual's data) were evaluated by an optometrist who was blinded to any of the patient data, and the most repeated response (2/3 same response) was taken as the final response. In cases

Table I Sample of Prompt Given and Responses Received from ChatGPT, with CST Value and without CST Value

Prompt	Response 1	Response 2	Response 3
<p>With CST</p> <p>1. Predict the risk of developing DR, A 55-year-old male with 60 months duration of diabetes, 6.9% HbA1c, 154mg/dl total cholesterol, 74.2 mg/dl LDL, 46 mg/dl HDL, 169 mg/dl triglyceride, 16.1 microalbuminuria, 136/78 mm/hg BP, CST_OD 227, CST_OS 228</p>	<p>Based on the provided information, here is the risk assessment for developing diabetic retinopathy:</p> <ol style="list-style-type: none"> 1. Duration of Diabetes: 60 months (5 years) - relatively shorter duration. 2. HbA1c Level (Glycated Hemoglobin): 6.9% - fairly well controlled. 3. Lipid Profile: Total cholesterol, LDL, HDL, and triglyceride levels seem within or close to the normal range. 4. Microalbuminuria: Present, but the value is relatively lower compared to higher risk cases. 5. Blood Pressure: 136/78 mmHg - within the normal range. 6. CST (Central Subfield Thickness) of the eyes: Values for both eyes are within normal limits. <p>Given this information:</p> <ul style="list-style-type: none"> • Duration of diabetes: 60 months (5 years) • HbA1c: 6.9% (fairly good control) • Lipid Profile: Within or close to normal range. • Microalbuminuria: Present, but the value is not significantly high. • Blood Pressure: Within the normal range. • CST (Central Subfield Thickness) of the eyes: Within normal limits. <p>This patient demonstrates relatively well-controlled diabetes, with near-normal values for most parameters, including HbA1c, lipid profile, blood pressure, and eye measurements. The microalbuminuria, while present, is not significantly high.</p> <p>The risk assessment for this individual would likely fall within the category of low to moderate risk for developing diabetic retinopathy. However, it's essential for this individual to maintain regular check-ups and healthy lifestyle practices to minimize the risk further.</p>	<p>Assessing the risk of developing Diabetic Retinopathy (DR) involves considering multiple factors, including age, duration of diabetes, HbA1c level, lipid profile, blood pressure, microalbuminuria, and central subfield thickness (CST) measurements. Based on the information provided, we can estimate the risk as follows:</p> <ol style="list-style-type: none"> 1. Age: The patient is 55 years old, which is a moderate risk factor for DR. 2. Duration of diabetes: The patient has had diabetes for 60 months, which is relatively short and is a lower risk factor. 3. HbA1c level: The HbA1c level is 6.9%, indicating good glycemic control, which lowers the risk. 4. Lipid profile: <ul style="list-style-type: none"> • Total cholesterol: 154 mg/dl is within the normal range. • LDL cholesterol: 74.2 mg/dl is within the normal range. • HDL cholesterol: 46 mg/dl is within the normal range. • Triglycerides: 169 mg/dl is slightly elevated but not excessively high. 5. Microalbuminuria: The microalbuminuria level at 16.1 is relatively low, indicating a lower risk of kidney damage due to diabetes. 6. Blood Pressure: The blood pressure is 136/78 mm/Hg, which is within the normal range. 7. CST measurements: Both CST_OD and CST_OS measurements are within the normal range. <p>Based on this information, we can estimate the risk of developing DR as "low to moderate" risk. However, it's important to note that risk assessment is not absolute and can vary depending on individual factors and local clinical guidelines. Regular eye examinations and consultation with a healthcare professional are essential for a more accurate risk assessment and management of diabetic retinopathy. This individual appears to have several factors that reduce the risk of DR, including good glycemic control and relatively healthy lipid and blood pressure profiles.</p>	<p>To predict the risk of developing Diabetic Retinopathy (DR) based on the provided information, we can use some common risk factors associated with the development of DR. Please note that this is a general assessment and not a definitive diagnosis. The risk assessment can be categorized as follows:</p> <ol style="list-style-type: none"> 1. Age: The patient is 55 years old. Older age is a risk factor for DR. 2. Duration of Diabetes: The patient has had diabetes for 60 months, which is equivalent to 5 years. Longer duration of diabetes is a risk factor for DR. 3. HbA1c Level: The patient has an HbA1c of 6.9%. An HbA1c level above 7% is generally considered a risk factor for DR. While 6.9% is slightly below this threshold, it's still a moderate risk factor. 4. Total Cholesterol: The total cholesterol level is 154 mg/dL, which is within the normal range. 5. LDL Cholesterol: The LDL cholesterol level is 74.2 mg/dL, which is within the normal range. 6. HDL Cholesterol: The HDL cholesterol level is 46 mg/dL, which is within the normal range. 7. Triglycerides: The triglyceride level is 169 mg/dL, which is slightly elevated. Elevated triglycerides are a risk factor for DR. 8. Microalbuminuria: The patient has a microalbuminuria level of 16.1, which indicates kidney damage and is a risk factor for DR. 9. Blood Pressure: The blood pressure is 136/78 mmHg, which is slightly elevated. High blood pressure is a risk factor for DR. <p>Based on the provided information, the risk of developing Diabetic Retinopathy for this patient can be categorized as "Moderate to High." This is because there are multiple risk factors present, including longer duration of diabetes, an HbA1c level approaching 7%, elevated triglycerides, microalbuminuria, and slightly elevated blood pressure. However, it's essential to remember that DR risk can vary from person to person, and this is a general assessment. Regular eye check-ups and consultation with a healthcare professional are crucial for accurate risk assessment and management.</p>

(Continued)

Table 1 (Continued).

Prompt	Response 1	Response 2	Response 3
<p>Without CST</p> <p>2. Predict the risk of developing DR for a 51-year-old Female with 180 months duration of diabetes, 9.9% HbA1c, 197mg/dl total cholesterol, 103.8 mg/dl LDL, 31 mg/dl HDL, 311 mg/dl triglyceride, 43.9 microalbuminuria, 138/90 mm/hg BP</p>	<p>To predict the risk of developing Diabetic Retinopathy (DR) for the individual you described, let us assess the risk based on the provided information:</p> <p>1. Age: The individual is 51 years old, which is a relatively younger age. Age is a risk factor for DR, and being younger can be a protective factor.</p> <p>2. Duration of Diabetes: The individual has had diabetes for 180 months, which is 15 years. A longer duration of diabetes increases the risk of complications, including DR.</p> <p>3. HbA1c: The HbA1c level is 9.9%, which is significantly elevated. A high HbA1c level suggests poorly controlled diabetes, and this is a major risk factor for DR.</p> <p>4. Cholesterol Levels:</p> <ul style="list-style-type: none"> • Total Cholesterol: 197 mg/dl • LDL: 103.8 mg/dl • HDL: 31 mg/dl • Triglycerides: 311 mg/dl <p>These cholesterol levels are elevated, especially the triglycerides. Elevated cholesterol levels can contribute to vascular complications, including DR.</p> <p>5. Microalbuminuria: The individual has a microalbuminuria level of 43.9 micrograms, which is relatively high. Microalbuminuria is an early sign of kidney damage and a risk factor for diabetic complications.</p> <p>6. Blood Pressure: The blood pressure is 138/90 mm/Hg, which indicates elevated blood pressure. Hypertension is a risk factor for DR.</p> <p>Based on these factors, the risk of developing Diabetic Retinopathy for this individual can be assessed as "High" risk. The elevated HbA1c, long duration of diabetes, high cholesterol levels, high microalbuminuria, and elevated blood pressure all contribute to a high risk of DR. It is crucial for this individual to work closely with their healthcare provider to manage their diabetes, cholesterol levels, and blood pressure, as well as undergo regular eye screenings to detect and manage DR early. Intensive management of blood sugar and overall health is essential to reduce the risk.</p>	<p>Predicting the risk of developing Diabetic Retinopathy (DR) involves assessing multiple factors. Here is a general risk assessment based on the provided information:</p> <p>1. Age: 51 years - The individual is in a relatively younger age group, which is associated with a lower risk of DR.</p> <p>2. Duration of Diabetes: 180 months (15 years) - A longer duration of diabetes is generally associated with a higher risk of complications, including DR.</p> <p>3. HbA1c: 9.9% - A high HbA1c level indicates poor blood sugar control, which significantly increases the risk of DR.</p> <p>4. Total Cholesterol: 197 mg/dl - Elevated cholesterol levels can contribute to vascular complications, including DR.</p> <p>5. LDL Cholesterol: 103.8 mg/dl - Elevated LDL cholesterol levels can increase the risk of eye complications.</p> <p>6. HDL Cholesterol: 31 mg/dl - Low HDL cholesterol is generally considered a risk factor for vascular and eye complications.</p> <p>7. Triglycerides: 311 mg/dl - High triglyceride levels may significantly increase the risk of complications, including DR.</p> <p>8. Microalbuminuria: 43.9 - The presence of microalbuminuria is a marker of kidney damage and can increase the risk of eye complications in diabetes.</p> <p>9. Blood Pressure: 138/90 mmHg - Elevated blood pressure can increase the risk of DR, especially when it's not well controlled.</p> <p>Based on the information provided, this individual may be at a high risk of developing Diabetic Retinopathy. The combination of long diabetes duration, high HbA1c, elevated cholesterol levels, low HDL, high triglycerides, microalbuminuria, and elevated blood pressure all contribute to a higher risk of complications.</p> <p>It's important to emphasize that this is a general risk assessment, and individual risk can vary. Regular eye check-ups and discussions with a healthcare provider are essential for managing diabetes and its complications. Immediate attention to improving blood sugar control, cholesterol management, and blood pressure control is crucial in this case.</p>	<p>To predict the risk of developing Diabetic Retinopathy (DR) for this patient based on the provided information, we can use common risk factors associated with the development of DR. Please note that this is a general assessment and not a definitive diagnosis. The risk assessment can be categorized as follows:</p> <p>1. Age: The patient is 51 years old, which is relatively younger. Age is a risk factor, but being younger can be considered a protective factor against DR.</p> <p>2. Duration of Diabetes: The patient has had diabetes for 180 months, which is equivalent to 15 years. Longer duration of diabetes is a risk factor for DR.</p> <p>3. HbA1c Level: The patient has an HbA1c of 9.9%, which is significantly higher than the recommended target of below 7%. High HbA1c is a significant risk factor for DR.</p> <p>4. Total Cholesterol: The total cholesterol level is 197 mg/dL, which is slightly elevated.</p> <p>5. LDL Cholesterol: The LDL cholesterol level is 103.8 mg/dL, which is within the normal range.</p> <p>6. HDL Cholesterol: The HDL cholesterol level is 31 mg/dL, which is lower than desirable.</p> <p>7. Triglycerides: The triglyceride level is 311 mg/dL, which is significantly elevated. Elevated triglycerides are a risk factor for DR.</p> <p>8. Microalbuminuria: The patient has microalbuminuria with a level of 43.9, indicating kidney damage. This is a risk factor for DR.</p> <p>9. Blood Pressure: The blood pressure is 138/90 mmHg, which is elevated. High blood pressure is a risk factor for DR.</p> <p>Based on the provided information, the risk of developing Diabetic Retinopathy for this patient can be categorized as "High." This is because several significant risk factors are present, including a long duration of diabetes, very high HbA1c, elevated triglycerides, microalbuminuria, and elevated blood pressure. These factors collectively indicate a high risk of developing DR.</p> <p>It's crucial for this patient to work closely with their healthcare provider to manage their diabetes and address these risk factors to reduce the risk of DR. Regular eye check-ups and comprehensive diabetes management are essential in this case.</p>

<p>3. Predict the risk of developing DR for a 60-year-old Female with 180 months duration of diabetes, 7.1% HbA1c, 242 mg/dl total cholesterol, 130 mg/dl LDL, 12 mg/dl HDL, 118.0 mg/dl triglyceride, 13.2 microalbuminuria, 120/80 mm/hg BP</p>	<p>Based on the provided data, the risk of developing Diabetic Retinopathy (DR) for this individual can be categorized as "low to moderate risk." Here are the key factors contributing to this assessment:</p> <ol style="list-style-type: none"> 1. Age: The individual is 60 years old, which is a moderate risk factor due to age. Age is a significant factor in DR risk. 2. Duration of Diabetes: The individual has had diabetes for 180 months (15 years), which is a moderately long duration and contributes to the risk. 3. HbA1c: The HbA1c level is 7.1%, indicating fair glycemic control. While it's not extremely high, it's above the target range for good control, which is a moderate risk factor. 4. Total Cholesterol and LDL: Total cholesterol is relatively high at 242, and LDL (bad cholesterol) is also elevated at 130. High cholesterol levels are a moderate risk factor for DR. 5. HDL: The HDL (good cholesterol) level is quite low at 12.0. Low HDL is a moderate risk factor for DR. 6. Triglycerides: Triglycerides are within a normal range, which is a positive factor. 7. Microalbuminuria: The microalbuminuria level is 13.2, which is slightly elevated. Elevated microalbuminuria is a moderate risk factor for DR. 8. Blood Pressure: The blood pressure is within a normal range (120/80mm Hg), which is a positive factor in terms of DR risk. <p>Considering the individual's age, moderately long duration of diabetes, fair glycemic control, elevated cholesterol levels (both total and LDL), low HDL, and slightly elevated microalbuminuria, the overall risk assessment is "low to moderate risk." It's important for this individual to work on improving glycemic control, cholesterol management, and other risk factors to reduce the risk of DR and other diabetes-related complications. Regular monitoring and management are crucial.</p>	<p>Based on the provided data, the risk of developing Diabetic Retinopathy (DR) for this individual can be categorized as "moderate risk." Here are the key factors contributing to this assessment:</p> <ol style="list-style-type: none"> 1. Age: The individual is 60 years old, which is a moderate risk factor due to age. While not extremely high, age is still a significant consideration. 2. Duration of Diabetes: The individual has had diabetes for 180 months (15 years), which is a moderate duration. A longer duration of diabetes can contribute to a higher risk of developing complications. 3. HbA1c: The HbA1c level is 7.1%, indicating somewhat good but not optimal glycemic control. This contributes to a moderate risk of developing DR. 4. Total Cholesterol and LDL: Total cholesterol and LDL levels are elevated, which is a moderate risk factor for the development of DR. High cholesterol levels can contribute to increased risk. 5. HDL: The HDL (good cholesterol) level is quite low, which can be considered a moderate risk factor for DR. 6. Triglycerides: Triglyceride levels are within a normal range, which is a positive factor in terms of DR risk. 7. Microalbuminuria: The microalbuminuria level is slightly elevated, which contributes to a moderate risk of developing DR. 8. Blood Pressure: The blood pressure is within a normal range (120/80mm Hg), which is a positive factor in terms of DR risk. <p>Considering the moderate risk factors such as elevated cholesterol levels, a moderately long duration of diabetes, suboptimal glycemic control, and slightly elevated microalbuminuria, this individual would be categorized as having a "moderate risk" of developing DR. Regular monitoring, appropriate management of diabetes, and lifestyle modifications will be crucial in minimizing the risk and preventing complications like DR.</p>	<p>Based on the provided data, the risk of developing Diabetic Retinopathy (DR) for this individual can be categorized as "low risk." Here are the key factors contributing to this assessment:</p> <ol style="list-style-type: none"> 1. Age: The individual is 60 years old, which is a moderate risk factor due to age. However, the other factors in this case seem to be favorable. 2. Duration of Diabetes: The individual has had diabetes for 180 months (15 years), which is a moderate duration. It's not an extremely long duration. 3. HbA1c: The HbA1c level is 7.1%, which is in the range of good glycemic control. While not optimal, it is relatively well managed. 4. Total Cholesterol and LDL: Total cholesterol and LDL levels are within a normal range, which is a positive factor in terms of DR risk. 5. HDL: The HDL (good cholesterol) level is within a normal range. 6. Triglycerides: Triglyceride levels are within a normal range, which is a positive factor in terms of DR risk. 7. Microalbuminuria: The microalbuminuria level is slightly elevated but not severely so. This is a minor risk factor. 8. Blood Pressure: The blood pressure is within a normal range (120/80mm Hg), which is a positive factor in terms of DR risk. <p>Considering the moderately older age, moderately long duration of diabetes, and good management of other risk factors like cholesterol, blood pressure, and slightly elevated microalbuminuria, the overall risk of developing DR for this individual is assessed as "low risk." However, regular monitoring and continued management of diabetes are essential to maintain this low risk and prevent complications like DR.</p>
---	--	--	--

where there were completely different responses, the average response was taken as the final response. A sample of data given to the optometrist to choose the majority response is shown in Table 2.

For the purposes of this study, the ChatGPT predicted risk categories were divided into two groups: “DR” (Indicating the presence of Diabetic Retinopathy) and “No DR” (Indicating the absence of Diabetic Retinopathy). The ChatGPT predicted risks were arbitrarily divided, with half of the predicted risks falling into the categories of low risk, low to moderate risk, and moderate risk which were classified as “No DR”, while the other half of the risk categories such as moderate to high risk and high risk were classified as “DR” (Table 2).

Dividing the risks predicted into binary categories such as DR and No DR allows the comparison between the ChatGPT model and the actual clinical status of DR diagnosed by an ophthalmologist. This binary classification of risks predicted by ChatGPT simplifies and aligns the study with actual clinical practices, making the model’s (ChatGPT) performance more understandable and relevant to real-world clinical scenarios. While it represents an initial assessment of ChatGPT performance, it is a useful starting point for evaluating the model’s utility in the clinical diagnosis of DR and can guide further research and refinement. The clinical diagnosis was based on the dilated fundus examination done using Indirect Ophthalmoscope in the clinic, as well as an OCT evaluation.

Statistical Analysis

The intra-observer variability of ChatGPT response was assessed using the Intraclass Correlation Coefficient (ICC). The agreement between the ChatGPT prediction and clinical diagnosis of DR was assessed using Cohen’s kappa. SPSS version 25 (IBM Corp., Armonk, NY) was used for statistical analysis. Descriptive statistics were reported based on the mean (SD) and percentage. ICC values <0.5 were indicative of poor reliability, 0.5–0.75 indicated moderate reliability, values between 0.75 and 0.9 indicated good reliability and values >0.90 indicated excellent reliability.¹³ Cohen’s kappa values were interpreted as follows: <0: No agreement, 0.01–0.2: slight agreement; 0.21–0.4: fair agreement; 0.41–0.6: moderate agreement; 0.61–0.8: substantial agreement; 0.81–1.00: almost perfect agreement.¹⁴ A post hoc power analysis was conducted using G Power version 3.1.9.7 to determine the study power. The sensitivity and specificity were calculated based on the given formula:¹⁵

$$\text{Sensitivity} = [\text{True Positive} / (\text{True Positive} + \text{False Negative})] \times 100$$

$$\text{Specificity} = [\text{True Negative} / (\text{False Positive} + \text{True Negative})] \times 100$$

Results

Table 3 presents the anonymized baseline information regarding the clinical, biochemical, and ocular data of 111 patients with diabetes. The average age of the participants in the study was 62.3 years, with a range between 41 and 83 years. Among the participants, majority of 60 participants (54%) were male. The study found that the average duration of diabetes among the participants was 156.7 months (SD ± 79.04). The mean HbA1c level was 8.7% (SD ± 1.95). The average serum total cholesterol level was 188.6 mg/dl (SD ± 51.21). The mean serum low-density lipoprotein cholesterol (LDL-C) level was 110 mg/dl (SD ± 42.36). The mean serum high-density lipoprotein cholesterol (HDL-C) level was 43 mg/dl (SD ± 11.7). The mean serum triglyceride level was 171.3 mg/dl (SD ± 66.95). The average microalbuminuria level among the participants was 186.5 mg/dl (SD ± 331.29). The mean systolic blood pressure was 137.51 mm Hg (SD

Table 2 A Sample of Data Given to Optometrist and the Chosen Final Response

Response 1	Response 2	Response 3	Final Response	Arbitrary Classification
Low to Moderate	Low to moderate	Moderate to High	Low to Moderate risk	No DR
High	High	High	High risk	DR
Low to moderate	Moderate	Low	Low to moderate risk	No DR

Table 3 Clinical, Biochemical, and Ocular Data

Variables	Mean
Age	62.3 ± 8.91 (range 41–83)
Gender: Male, n (%)	60 (54%)
Female, n (%)	51 (46%)
Duration of Diabetes (Months)	156.7 ± 79.04
HbA1c (%)	8.7 ± 1.95
Serum Total Cholesterol (mg/dl)	188.6 ± 51.21
Serum LDL -C (mg/dl)	110 ± 42.36
Serum HDL-C (mg/dl)	43 ± 11.7
Serum Triglycerides (mg/dl)	171.3 ± 66.95
Micro albuminuria (mg/dl)	186.5 ± 331.29
Systolic BP (mm Hg)	137.51 ± 15.7
Diastolic BP (mm Hg)	76.61 ± 8.3
CST_OD (μm)	260 ± 78.03
CST_OS (μm)	268 ± 99

Abbreviations: HbA1c, Glycated Haemoglobin; LDL-C, Low-Density Lipoprotein Cholesterol; HDL-C, High-Density Lipoprotein Cholesterol; BP, Blood Pressure; CST, Central Subfield Thickness.

± 15.7) and the mean diastolic blood pressure was 76.61 mm Hg (SD ± 8.3). Central subfield thickness was measured for both eyes, with CST for the right eye with a mean value of 260 μm (SD ± 78.03), and CST for the left eye with a mean value 268 μm (SD ± 99).

The consistency of ChatGPT responses was assessed, resulting in an ICC value of 0.936 (95% CI, 0.913–0.954, $p < 0.001$) with CST values and 0.915 (95% CI, 0.706–0.846, $p < 0.001$) without CST values, both indicating an excellent reliability.

Table 4 presents the agreement between ChatGPT's prediction (with CST) and clinical DR status diagnosed by an ophthalmologist. When comparing ChatGPT's predictions to the clinical DR status, ChatGPT accurately predicted 37 (67%) cases as DR and 38 (68%) cases as No DR, consistent with the clinical diagnosis. Here, the ChatGPT's sensitivity and specificity was found to be 67% of 68%. However, ChatGPT also demonstrated misclassifications, as it predicted 18 (32%) cases that were clinically diagnosed as DR to be No DR and 18 (33%) cases that were clinically diagnosed as No DR to be DR.

Table 5 presents the agreement between ChatGPT's prediction (without CST values) and clinical DR status diagnosed by an ophthalmologist. When comparing ChatGPT's predictions to the clinical DR status, ChatGPT accurately predicted 40 (73%) cases as DR and 30 (54%) cases as No DR, consistent with the clinical diagnosis. This indicates the ChatGPT's sensitivity of 73% and a specificity of 54%. However, ChatGPT also demonstrated misclassifications, as it predicted 26 (46%) cases that were clinically diagnosed as DR to be No DR and 15 (27%) cases that were clinically diagnosed as No DR to be DR.

Cohen's kappa analysis was selected to determine the proportion of agreement between ChatGPT's prediction and clinical DR status. The results indicated only a fair agreement (both with CST and without CST values) between the ChatGPT prediction and the clinical DR status diagnosed by an ophthalmologist, with a kappa value of 0.263 (without CST) (95% CI, 0.114 to 0.412, $p = 0.005$) and 0.351 (with CST) (95% CI, 0.198 to 0.504, $p < 0.001$). A post hoc power

Table 4 Agreement Between ChatGPT Prediction and Clinical DR Status Diagnosed by Ophthalmologist (with CST)

ChatGPT_prediction	Clinical Diagnosis by Ophthalmologist			
		DR	No DR	Total
	DR	37 (67%)	18 (32%)	55 (49.5%)
	No DR	18 (33%)	38 (68%)	56 (50.5%)
	Total	55 (100%)	56 (100%)	111 (100%)

Table 5 Agreement Between ChatGPT Prediction and Clinical DR Status Diagnosed by Ophthalmologist (without CST)

ChatGPT_prediction	Clinical Diagnosis by Ophthalmologist			
		DR	No DR	Total
	DR	40 (73%)	26 (46%)	66 (59%)
	No DR	15 (27%)	30 (54%)	45 (41%)
	Total	55 (100%)	56 (100%)	111 (100%)

analysis was performed with an alpha value of 0.05, an observed effect size of 0.263, and the sample size of this study. The analysis revealed a study power of 79.12%, indicating that the study had a relatively high chance of detecting the observed effect, given the sample size and significance level.

Discussion

AI in health care has shown promising results in several areas, including diagnosis, disease prediction, and patient management. With the advent of NLP models, such as ChatGPT, the capacity for generating accurate and personalized medical reports has increased significantly.^{16,17} The implications of these advancements are vast, potentially enabling more efficient medical services, improving the accuracy of diagnoses and predictions, and saving considerable human resources.^{18,19}

Furthermore, it is worth considering the broader impact of the introduction of AI in health care, particularly concerning patient-doctor communication. As NLP models continue to improve, especially ChatGPT is constantly updated by Open AI since their launch, they may become an integral part of the healthcare system and can serve as a bridge between healthcare providers and patients, facilitating more effective communication, and improving patient satisfaction and outcomes.²⁰ This AI-based screening tool would be a great help particularly in resource constrained settings. Furthermore, ChatGPT's ability to generate simple and comprehensible advice can make it an invaluable tool for patient education especially for the people with limited awareness about Diabetic retinopathy.

In our study, initially we fed ChatGPT with, age, sex, duration of diabetes, blood pressure, glycated hemoglobin (HbA1c), serum total cholesterol, low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides, microalbuminuria and CST values obtained from OCT of all the patients to predict the risk of diabetic retinopathy. These parameters are well-established biomarkers for DR.²¹ OCT, known for its accuracy, serves as a reliable diagnostic tool for detecting diabetic macular edema (DME), with OCT-derived central retinal thickness being an ocular (clinical) biomarker for DR.^{21,22}

In subsequent analysis, we fed ChatGPT with same parameters excluding CST values, to evaluate its performance in predicting the diabetic retinopathy risk without CST values. This was done considering the challenges associated with utilizing OCT diagnostics, such as dedicated macular OCT scans to the DR screening, such as cost constraints, lack of expertise of operators and relatively low availability of OCT machine.⁹

While many current screening programs are efficient in detecting referable DR, the accurate diagnosis of DME, which causes vision threatening diabetic retinopathy, is more challenging. This is due to the inability to reliably identify retinal thickening from the two-dimensional fundus photographs used for DR screening.

Diabetic retinopathy, if left untreated, can lead to severe vision problems and significant healthcare costs, including the cost of surgeries and interventions. Investing in early detection through OCT can be cost-effective in the long run, as it can prevent the progression of the disease and reduce the overall financial burden on healthcare systems.

From a medical and cost-effectiveness perspective, several studies have investigated the incorporation of macular OCT as part of a DR screening. Prescott et al found that use of OCT in cases where color fundus photos were used for detecting macular edema resulted in direct cost savings of 16–17% because of fewer unnecessary referrals without a measurable decrease in medical benefits.²³

Our study's findings are particularly significant considering the global prevalence of diabetes and its complications. Diabetic retinopathy is one of the leading causes of vision loss worldwide.² The early and accurate prediction of DR risk is crucial in reducing the burden of this disease. In this context, the use of AI models like ChatGPT can greatly aid in the early detection of DR risk, facilitating timely intervention and possibly improving patient outcomes.

Our study also found that ChatGPT has the ability to predict the risk of DR based on clinical and biochemical parameters, highlighting its potential application as a screening tool with a sensitivity and specificity of 73% and 54% when used without CST information. When CST information was included, ChatGPT exhibited a moderate sensitivity and specificity of 67% and 68%.

These findings suggest that ChatGPT with 73% sensitivity (without CST) has some promise as a screening tool for identifying individuals at risk of DR. However, the specificity of 54% without CST highlights that there is room for improvement, as it may generate false-positive results in a substantial number of cases. The sensitivity of 73% without CST is notable, indicating that ChatGPT can correctly identify a significant portion of individuals with DR or at risk of developing it.

Incorporating CST information improved specificity to 68%, reducing the number of false positives, but at the expense of a slight reduction in sensitivity. This trade-off should be carefully considered when implementing ChatGPT as a screening tool. A moderate sensitivity and specificity may be acceptable in certain scenarios, such as in a preliminary screening process, but further refinements may be necessary to increase its diagnostic accuracy.

ChatGPT responses had an excellent reliability of responses [0.911 (without CST) and 0.935 (with CST)]. However, its accuracy in determining the presence or absence of DR is not yet optimal. The Cohen's Kappa analysis also indicated only a fair agreement between ChatGPT's prediction and clinical diagnosis in both the cases with and without CST.

This is the first study to assess the accuracy of the NLP model for evaluating the risk of DR based on various systemic parameters. This novel approach will be a guide for future refinements and fine-tuning of the model for using ChatGPT for DR screening purposes. However, the limitations of this study should also be addressed. First, the chatbot language model used in this study is a research version that was not specifically designed for medical applications. GPT models can sometimes generate nonsensical or incorrect information due to their limitations in their training data and algorithm. The limitations and biases in the training data can affect how accurate and reliable AI models are. Fine-tuning the model specifically for medical applications, using well-curated training data would likely enhance its performance and could potentially result in improved sensitivity and specificity.

Second, there have been reports suggesting that ChatGPT is not compliant with HIPAA regulations and that it has been susceptible to hacking attempts, resulting in the exposure of proprietary information.²⁴ To ensure the privacy and protection of patient data, all personal information pertaining to the patients was anonymized before being inputted into ChatGPT.

Third, due to the retrospective nature of the study, we had a relatively small sample size of 111 patients in our study. While we acknowledge that the number of participants was limited, it is important to clarify the rationale behind this choice and address its implications for the study's conclusions. The selection of these patients was primarily driven by the availability of data from a specific research project. The use of this group was particularly relevant to the study due to their potential susceptibility to DR. We aimed to utilize the existing dataset to explore the potential of ChatGPT in DR risk assessment. Given the resources and data at hand, this sample size allowed us to conduct a preliminary evaluation of the AI model's performance within our available means.

While it would have been ideal to prospectively design the study with a predetermined sample size calculation, the retrospective nature of our investigation constrained this possibility. We acknowledge the importance of this consideration and the potential impact on the study's outcomes. However, we performed a post hoc power analysis based on the available dataset of 111 subjects with diabetes. Meanwhile, the post hoc power analysis indicated a study power of 79.12%, suggesting a relatively high likelihood of detecting the observed effect given the sample size and significance level.

Looking ahead, we emphasize the need for future research endeavors to prioritize larger sample sizes in order to enhance the statistical power and validity of conclusions. A prospective approach would undoubtedly provide valuable insights into ChatGPT's practical applicability and potential as a predictive tool for diabetic retinopathy risk assessment.

Despite its limitations, it may serve as an assistive tool in disease detection and monitoring, complementing the work of medical professionals. Further research is necessary to improve the accuracy of ChatGPT for DR risk assessment.

This study has merely scratched the surface of the possible applications of NLP-based AI model in health care. Future research should seek to expand on these findings, investigating the efficacy of AI models in other areas of health care, and exploring ways to improve the accuracy of these models.

Additionally, a key area for future research is the exploration of the ethical implications of AI in health care. While AI models offer numerous benefits, it is crucial to ensure their ethical use, particularly concerning data privacy and the potential for AI to replace human decision-making in health care. Only with careful consideration of these ethical issues, we can truly harness the potential of AI in healthcare.

Finally, in light of the limitations of ChatGPT observed in our study, future versions of the model might consider incorporating a medical-specific training regimen, perhaps developed in collaboration with healthcare professionals. This could help us to address the model's current limitations, improve its accuracy and reliability, and, in turn, its utility in healthcare.

Conclusion

Diabetic retinopathy, if left untreated, can lead to severe vision loss. Therefore, early and accurate detection of DR is imperative to prevent vision impairment. Our findings have shown promising results, where AI-based tools, such as ChatGPT, can predict the risk of developing DR using only clinical and biochemical parameters. This helps in the early identification of high-risk patients. However, there is still considerable work to be done. As AI and NLP models continue to evolve, they hold enormous potential for enhancing healthcare services.

Funding

There is no funding to report.

Disclosure

No conflicting relationship exists for any author.

References

1. Ong KL, Stafford LK, McLaughlin SA, et al. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet*. 2023;2023:1.
2. Raman R, Vasconcelos JC, Rajalakshmi R, et al. Prevalence of diabetic retinopathy in India stratified by known and undiagnosed diabetes, urban-rural locations, and socioeconomic indices: results from the SMART India population-based cross-sectional screening study. *Lancet Glob Health*. 2022;10(12):e1764–e1773. doi:10.1016/S2214-109X(22)00411-9
3. Teo ZL, Tham YC, Yu M, et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology*. 2021;128(11):1580–1591. doi:10.1016/j.ophtha.2021.04.027
4. Kumar S, Mohanraj R, Raman R, et al. 'I don't need an eye check-up'. A qualitative study using a behavioural model to understand treatment-seeking behaviour of patients with sight threatening diabetic retinopathy (STDR) in India. *PLoS One*. 2023;18(6):e0270562. doi:10.1371/journal.pone.0270562
5. Owusu-Afriyie B, Gende T, Tapilas M, Zimbare N, Kewande J. Patients' perspective on barriers to utilization of a diabetic retinopathy screening service. *Diabetology*. 2023;4(3):393–405. doi:10.3390/diabetology4030033
6. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010. *BMJ Open*. 2014;4(2):e004015. doi:10.1136/bmjopen-2013-004015
7. DiGiorgio AM, Ehrenfeld JM. Artificial intelligence in medicine & ChatGPT: de-tether the physician. *J Med Syst*. 2023;47(1):32. doi:10.1007/s10916-023-01926-3
8. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *J Med Syst*. 2023;47(1):33. doi:10.1007/s10916-023-01925-4
9. Fenner BJ, Wong RL, Lam WC, Tan GS, Cheung GC. Advances in retinal imaging and applications in diabetic retinopathy screening: a review. *Ophthalmol Ther*. 2018;7:333–346. doi:10.1007/s40123-018-0153-7
10. Bhargava M, Cheung CY, Sabanayagam C, et al. Accuracy of diabetic retinopathy screening by trained non-physician graders using non-mydiatic fundus camera. *Singapore Med J*. 2012;53(11):715.
11. Brzezicki MA, Bridger NE, Kobetić MD, et al. Artificial intelligence outperforms human students in conducting neurosurgical audits. *Clin Neurol Neurosurg*. 2020;192:105732. doi:10.1016/j.clineuro.2020.105732
12. Biswas SS. Role of chat gpt in public health. *Ann Biomed Eng*. 2023;51(5):868–869. doi:10.1007/s10439-023-03172-7
13. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–163. doi:10.1016/j.jcm.2016.02.012

14. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med.* 2012;22(3):276–282. doi:10.11613/BM.2012.031
15. Trevethan R. Sensitivity, specificity, and predictive values: foundations, plabilities, and pitfalls in research and practice. *Front Public Health.* 2017;5:307. doi:10.3389/fpubh.2017.00307
16. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J.* 2019;6(2):94. doi:10.7861/futurehosp.6-2-94
17. Sarraju A, Bruemmer D, Van Iterson E, Cho L, Rodriguez F, Laffin L. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model. *JAMA.* 2023;329(10):842–844. doi:10.1001/jama.2023.1044
18. Uy H, Fielding C, Hohlfeld A, et al. Diagnostic test accuracy of artificial intelligence in screening for referable diabetic retinopathy in real-world settings: a systematic review and meta-analysis. *PLOS Global Public Health.* 2023;3(9):e0002160. doi:10.1371/journal.pgph.0002160
19. Were MC. Challenges in digital medicine applications in under-resourced settings. *Nat Commun.* 2022; 13:3020 (2022). doi:10.1038/s41467-022-30728-3
20. Javaid M, Haleem A, Singh RP. ChatGPT for healthcare services: an emerging stage for an innovative perspective. *BenchCouncil Trans Benchmarks Stand Eval.* 2023;3(1):100105. doi:10.1016/j.tbench.2023.100105
21. Jenkins AJ, Joglekar MV, Hardikar AA, Keech AC, O'Neal DN, Januszewski AS. Biomarkers in diabetic retinopathy. The review of diabetic studies. *Rev Diabet Stud.* 2015;12(1–2):159. doi:10.1900/RDS.2015.12.159
22. Virgili G, Menchini F, Casazza G, et al. Optical coherence tomography (OCT) for detection of macular oedema in patients with diabetic retinopathy. *Cochrane Database Syst Rev.* 2015;1. doi:10.1002/14651858.CD008081.pub3
23. Prescott G, Sharp P, Goatman K, et al. Improving the cost-effectiveness of photographic screening for diabetic macular oedema: a prospective, multi-centre, UK study. *Br J Ophthalmol.* 2014;98(8):1042–1049. doi:10.1136/bjophthalmol-2013-304338
24. Priyanshu A, Vijay S, Kumar A, Naidu R, Mireshghallah F. Are chatbots ready for privacy-sensitive applications? An investigation into input regurgitation and prompt-induced sanitization. *arXiv Preprint.* 2023;2023:1.

Clinical Ophthalmology

Dovepress

Publish your work in this journal

Clinical Ophthalmology is an international, peer-reviewed journal covering all subspecialties within ophthalmology. Key topics include: Optometry; Visual science; Pharmacology and drug therapy in eye diseases; Basic Sciences; Primary and Secondary eye care; Patient Safety and Quality of Care Improvements. This journal is indexed on PubMed Central and CAS, and is the official journal of The Society of Clinical Ophthalmology (SCO). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/clinical-ophthalmology-journal>