AMERICAN ACADEMY
OF OPHTHALMOLOGY®

A User-friendly Approach for the Diagnosis of Diabetic Retinopathy Using ChatGPT and Automated Machine Learning

S. Saeed Mohammadi, MD, Quan Dong Nguyen, MD, MSc

Purpose: To assess the capabilities of Chat Generative Pre-trained Transformer (ChatGPT) and Vertex AI in executing code-free preprocessing, training machine learning (ML) models, and analyzing the data.

Design: Evaluation of diagnostic test or technology.

Participants: ChatGPT and Vetrex AI as publicly available large language model and ML platform, respectively.

Methods: ChatGPT was employed to improve the resolution of fundus photography images from the Methods to Evaluate Segmentation and Indexing Techniques in the field of Retinal Ophthalmology (Messidor-2) open-source dataset using the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique by Fiji software. Subsequently, Vertex AI, an automated ML (AutoML) platform, was utilized to develop 2 classification models. The first model served as a binary classifier for detecting the presence of diabetic retinopathy (DR), while the second determined its severity. Finally, ChatGPT was used to provide scripts for R and Python programming languages for data analysis and was also directly employed in analyzing the data in a code-free method.

Main Outcome Measures: Evaluating the utility of ChatGPT in generating scripts for preprocessing images using Fiji and analyzing data across Python and R and assessing its potential in analyzing data through a code-free method. Investigating the capabilities of Vertex AI to train image classification models for detection of DR and its severity.

Results: Two ML models were trained using 1740 images from the Messidor-2 database. The first model, designed to detect the severity of DR, achieved an area under the precision-recall curve (AUPRC) of 0.81, with a precision rate of 81.81% and recall of 72.83%. The second model, tailored for the detection of the presence of DR, recorded a precision and recall of 84.48% with an AUPRC of 0.90.

Conclusions: ChatGPT and Vertex AI have the potential to enable physicians without coding expertise to preprocess images, analyze data, and train ML models.

Financial Disclosure(s): Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article. *Ophthalmology Science 2024;4:100495 © 2024 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).*

Supplemental material available at www.ophthalmologyscience.org.

Chat Generative Pre-trained Transformer (ChatGPT), developed by OpenAI, is a large language model designed to generate responses that closely mimic human conversation.[1] As a part of the Generative Pretrained Transformer family, this model creates text by analyzing preceding text to accurately predict and generate the words that should logically follow.[2] Since its launch, ChatGPT has attracted millions of users, making it a popularly utilized tool.[3] Recently, ChatGPT has been equipped with Data Analyst feature, an advanced tool for data analysis. This addition is ideal for users seeking to explore data and solve problems with the aid of artificial intelligence (AI) tools.[4]

The effective use of AI technologies enhances the analysis of complex images, leading to quicker and more accurate disease detection, which results in improved patient care outcomes.[5,6] In the field of machine learning (ML), various algorithms and methodologies are utilized to allow computers to identify patterns and classify images based on the findings. In the context of diabetic retinopathy (DR) classification, ML algorithms can be effectively trained using large datasets of labeled fundus photographs. This training enables these algorithms to identify patterns and features corresponding to various DR stages, allowing for the classification of new fundus images that have not been previously encountered.[7]

Traditionally, development of ML models has been a complex and time-consuming process, requiring computational expertise to optimize hyperparameters. Such a process has posed challenges for clinicians, who often lack coding skills in ML model development.[8] However, the advent of

https://doi.org/10.1016/j.xops.2024.100495
ISSN 2666-9145/24

automated ML (AutoML) has changed this landscape. Vertex AI, developed by Google Cloud, offer graphical user interfaces that enable users to construct ML models. Impressively, AutoML models exhibit performance comparable to traditionally developed models.[9,10]

Diabetic retinopathy is a microvascular disorder resulting from the long-term effects of diabetes mellitus. Several studies have indicated that the number of DR patients in the United States will reach 16.0 million by 2050, with approximately 3.4 million experiencing vision-threatening complications. Diabetic retinopathy can lead to severe vision loss and, in some cases, blindness, making it the leading cause of visual impairment among working-age adults in the western world.[11,12] Early detection and accurate classification of various stages of DR are essential for prompt and effective treatment.[13−15] Timely intervention can slow down the progression of DR from its early nonproliferative stages, primarily through glycemic and blood pressure control,[16,17] and reduce vision loss in later stages using intravitreal injections and laser photocoagulation.[18]

In this index study, we aimed to evaluate the function of ChatGPT in assisting physicians with enhancing the quality of fundus photographs and analyzing the data. Additionally, we assessed the utility of Vertex AI as an AutoML platform for training ML models for the detection and grading of DR.

Methods

This study adhered to the principles of the Declaration of Helsinki and was conducted using a publicly available deidentified dataset; therefore, institutional review board approval or informed consent was not required.

Dataset

Methods to Evaluate Segmentation and Indexing Techniques in the field of Retinal Ophthalmology (Messidor-2), an open-source dataset, was used to evaluate the efficacy of the proposed strategy.[19] The Messidor-2 dataset was compiled from 3 ophthalmic departments in France, employing a digital video recording camera mounted to a Topcon TRC NW6 retinograph. Within this database, a total of 1748 images were classified based on the severity of DR. The classification system adopted for this study assigned specific labels to the images: 0 for “No DR,” 1 for “mild NPDR” (nonproliferative DR), 2 for “moderate NPDR,” 3 for “severe NPDR,” and finally, 4 for “PDR” (proliferative DR).[20] The categorization was determined following an adjudication protocol, as described in the research conducted by Krause et al.[21]

Preprocessing of Images

Contrast Limited Adaptive Histogram Equalization (CLAHE) is an image enhancement method aimed at improving the visual quality of digital images by enhancing their contrast. Unlike traditional histogram equalization, which stretches the intensity levels of an image across the entire dynamic range, CLAHE addresses the issues of overamplification of noise and artifacts in low-texture regions by dividing the image into small, overlapping tiles or regions. Each tile undergoes histogram equalization separately, with a contrast limiting mechanism in place to ensure that no tile’s histogram is stretched beyond a predefined threshold. Such algorithm prevents the overamplification of noise in smooth areas and yields more natural-looking results by uniformly distributing the color intensities while maintaining the original hues intact.[22,23] The localized contrast enhancement offered by CLAHE allows for the effective improvement of details in both bright and dark areas of an image. Consequently, it proves particularly valuable for images with nonuniform lighting or low contrast.[24,25] To preprocess images using CLAHE, we queried ChatGPT-4, “I have thousands of images that I would like to enhance their resolution using CLAHE. Can you kindly demonstrate how to process this batch of images in Fiji?” (Fig S1). Following the instructions provided by ChatGPT, we successfully processed the batch of images using Fiji (ImageJ2 2.14.0/1.54f)[26] (Fig 2).

Model Training

Two single-label datasets were created to train 2 distinct image classification models. This process involved selecting the “us-central” region on Vertex AI and uploading the preprocessed images to Google Cloud. The first dataset was compiled using 5 labels to grade DR, and the second dataset was created using 2 labels to indicate the presence or absence of DR. Correspondingly, the first model was trained using the first dataset to grade severity of DR, while the second model was trained using the second dataset to detect the presence of DR. The dataset was automatically split into 3 sets: 80% for training, 10% for validation, and the remaining 10% for testing the data. To enhance data security during the training process, Google-managed encryption key services were used. Objectives were set to prioritize higher accuracy, aiming for a latency of 200 to 300 milliseconds. Eight node hours were used for training the model, with "node hour" representing an hour of computation on an individual computing node.

Performance Metrics

The AutoML platform provides performance metrices derived from the testing set, while the AI model predicts categories by assigning probabilities to individual images. To classify an image, a confidence threshold determines the algorithm’s minimum required confidence level. The model’s accuracy is evaluated by computing the area under the precision-recall curve (AUPRC) across a spectrum of confidence levels ranging from 0.0 to 1.0 on the testing dataset. The precision-recall curve highlights varying precision and recall values achievable through adjustable confidence thresholds, providing insights into the model’s performance. Additionally, the platform generates confusion matrices including true-positives, true-negatives, false-positives, and false-negatives which enable calculation of performance indices such as sensitivity (SN), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), accuracy, and F1 scores.

Analysis of Data Using Programming Languages

To analyze the confusion matrix of the first trained model and acquire essential performance indices, ChatGPT-4 was utilized to obtain scripts for both R and Python. For R scripts, we presented the following query to ChatGPT: “I have a confusion matrix with 5 labels, labeled 0 to 4. The first row is the header including the prediction labels and the first column is the true diagnosis. Blank cell means zero value. Can you please share the R script to import a comma-separated values (CSV) file, calculate the SN, SP, PPV, NPV, accuracy, and F1 score, and export the results as a CSV file?” The input and output file paths were specified at the end of the query. Subsequently, ChatGPT provided the script. However, the output metrics were found to be all zeros or not availables, which was not anticipated. We asked ChatGPT to create a script that displays the contents of a CSV file, aiming to gather the necessary information for troubleshooting purposes. Based on the error

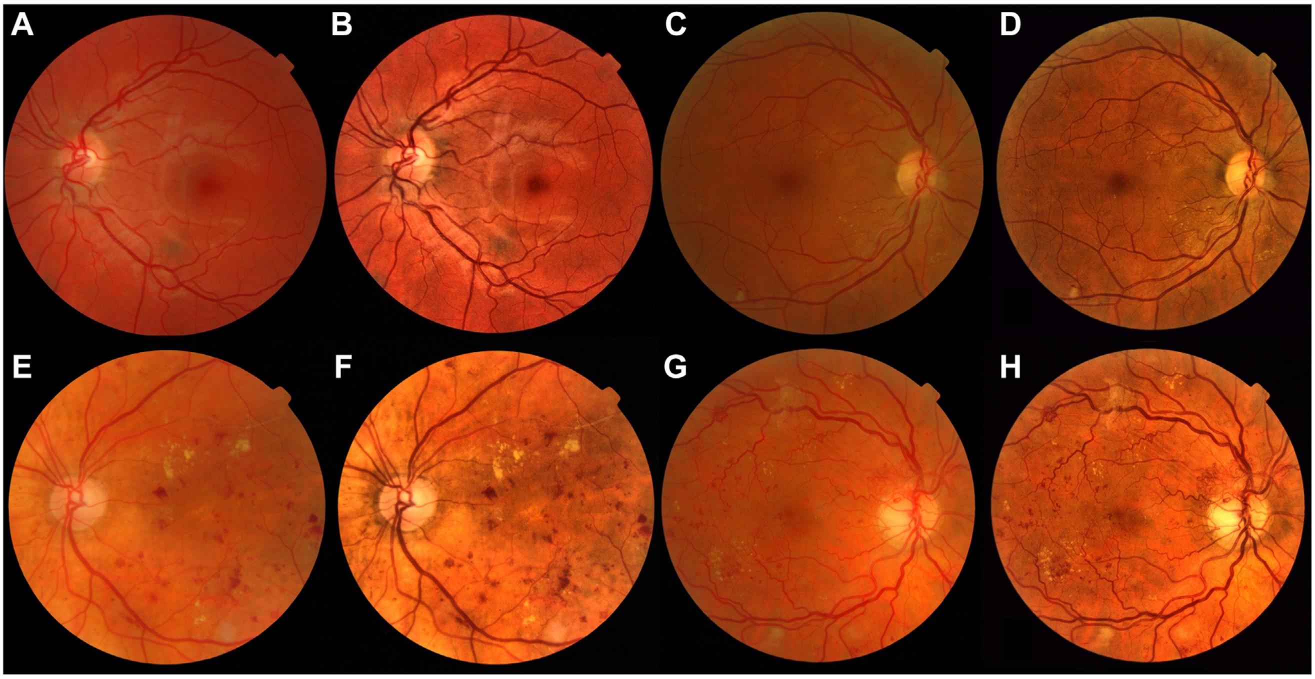

Figure 2. **A,** Original no diabetic retinopathy fundus photo. **B,** Preprocessed no diabetic retinopathy fundus photo. **C,** Original moderate nonproliferative diabetic retinopathy (NPDR) fundus photo. **D,** Preprocessed moderate NPDR fundus photo. **E,** Original severe NPDR fundus photo. **F,** Preprocessed severe NPDR fundus photo. **G,** Original proliferative diabetic retinopathy fundus photo. **H,** Preprocessed proliferative diabetic retinopathy fundus photo.

description from R and ChatGPT's responses, it was determined that blank cells had been converted to "not availables" during import into R and were not recognized as numeric values, leading to miscalculations. Upon running the revised script generated by ChatGPT, a new error was encountered, indicating a mismatch in the number of items being replaced in a data frame compared with the required number. However, after the details of this error were provided to ChatGPT, appropriate edits were made to the script. The final version of the script, executed after these adjustments, successfully yielded the desired results (Fig S3).

For Python3 scripts, we altered our strategy. Initially, we aimed to provide ChatGTP with the data structure by asking, "I have a confusion matrix with 5 labels, labeled 0 to 4. The first row is the header including the prediction labels and the first column is the true diagnosis. Blank cell means zero value. Please provide me with a script to print the contents of confusion_matrix.csv." While trying to run the script we encountered an error: "ModuleNotFoundError: No module named 'pandas.'" This issue was resolved by reporting the exact error details to ChatGPT and proceeding as instructed (Fig S4). The results of the script were shared with ChatGPT. Then, the following inquiry was made and paths for the input and output files were also provided: "I have a confusion matrix with 5 labels, labeled 0 to 4. The first row is the header including the prediction labels and the first column is the true diagnosis. Blank cell means zero value. Can you please share the Python3 script to import a CSV file, calculate the SN, SP, PPV, NPV, accuracy, and F1 score, and export the results as a CSV file?" Subsequently, ChatGPT generated the script, as shown in Fig S5. In line with earlier instructions regarding the "pandas" library, we proceeded to install the "numPy" library and ran the script which resulted in obtaining the desired results.

The scripts provided by ChatGPT were executed using R software (version 4.3.1; R Foundation for Statistical Computing) and Python programming language (version 3.9.7; Python Software Foundation).

Analysis of Data Using Code-Free Method

ChatGPT was employed as a code-free solution for our data analysis following the activation of the ChatGPT Plus subscription plan. The "Data Analyst" Generative Pretrained Transformer was accessed through the "Explore Generative Pretrained Transformers" section to upload a CSV file to ChatGPT. The same query was posted: "I have a confusion matrix with 5 labels, labeled 0 to 4. Blank cell means zero value. Can you please calculate the SN, SP, PPV, NPV, accuracy, and F1 score, and export the results as a CSV file?" ChatGPT conducted the analysis and provided an exportable CSV file with the results (Fig S6).

To ensure accuracy in calculations and metrics, a comprehensive verification process of the results was conducted. The same steps were followed to analyze the data for the second model.

Results

In this study, a comprehensive dataset of retinal images (Messidor-2) was uploaded to Google Cloud for analysis and annotation. The dataset consisted of a total of 1748 images; however, 8 were excluded due to the absence of labels, which could have potentially reduced the accuracy of the models. The dataset was labeled, with 1017 images classified as "no DR," 268 images labeled as "mild NPDR," 345 images labeled as "moderate NPDR," 75 images labeled as "severe NPDR," and 35 images labeled as "PDR." The model's AUPRC for grading the severity of DR was 0.81 with precision of 81.81% and recall of 72.83% using a score

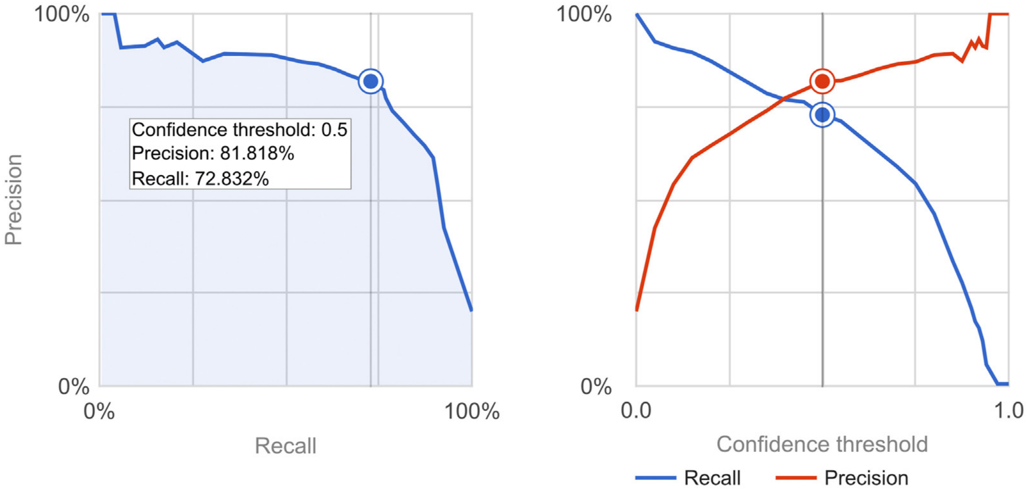

Figure 7. Precision-recall curves for grading diabetic retinopathy.

threshold of 0.5 (Fig 7). Table 1 presents the precision and recall of the model trained for grading the severity of DR at different confidence thresholds, namely 25%, 50%, and 75%.

Table 2 presents the evaluation parameters used for assessing the model trained to detect the severity of DR.

A second model was trained using 2 labels: label 0 for "no DR" and labels 1, 2, 3, and 4 for "DR." The Vertex AI platform provided several performance metrics to evaluate the performance of the trained model to detect DR. The model achieved an AUPRC of 0.90. It also demonstrated an SN of 76% and an SP of 90%. The PPV was recorded at 88.37%, while NPV reached 78.95%. Furthermore, the model attained an F1 score of 0.84 (Fig 8). Table 3 presents the precision and recall of the model trained for detection of the presence of DR at 25%, 50%, and 75% confidence thresholds.

Discussion

In this study, the capabilities of ChatGPT in data preprocessing and analysis were assessed. ChatGPT provided R and Python scripts and instructed us how to enhance image quality using Fiji. Additionally, we explored the potential of AI in a broader context by using Vertex AI to train 2 models with the Messidor-2 open-source database. The primary objectives were twofold: first, to demonstrate the feasibility of integrating AI technologies into existing workflows, and second, to highlight the user-friendly nature of ChatGPT and Vertex AI, which facilitated the preprocessing, training, and analysis processes.

Table 1. Precision and Recall at Different Confidence Thresholds for Grading Diabetic Retinopathy Stages

Confidence Threshold	Precision (%)	Recall (%)
25%	67.59	84.39
50%	81.81	72.83
75%	87.04	54.34

Numerous deep learning models have been developed for DR classification, each employing different datasets.[27–29] However, the challenge of low resolution and suboptimal image quality significantly impacts the performance of the trained models as low-quality images have less distinguishable details.[30–32] Retinal fundus images are commonly captured using different fundus photography devices, leading to intensity variations in the photographs. Therefore, it becomes critical to enhance the quality of fundus images and eliminate different types of noise.[33] One effective technique for improving image quality is CLAHE, which enhances contrast and reduces the impact of uneven illumination.[24,34]

In a study by Faes et al, the feasibility of training AutoML models by health care experts without coding backgrounds was explored. The study utilized the Messidor dataset to develop a model capable of distinguishing between eyes without DR and those with any degree of DR. The outcomes revealed an AUPRC of 0.87, with precision, SN, and SP reported as 73%, 73%, and 67%, respectively.[35] In our study, feasibility of preprocessing, training AutoML models, and analyzing data using ChatGPT and Vertex AI by physicians lacking coding experience was evaluated. We successfully applied CLAHE to enhance image quality, achieving improved detection of DR with SN, SP, PPV, and NPV reported as 76%, 90%, 88.37%, and 78.95%, respectively. Moreover, an AUPRC of 0.90 demonstrated the model's efficacy in correctly identifying DR cases. Additionally, we utilized ChatGPT for data analysis in R and Python, as well as through a noncoding method that involves uploading the data to ChatGPT.

Sanchez et al evaluated the presence of DR in the Messidor dataset using traditional ML techniques and reported an AUPRC of 0.87, successfully distinguishing normal images from those with DR. They also achieved an SN of 92.2% at an SP of 50%.[36] Similarly, Anatal et al conducted

Table 2. Parameters for Evaluation of the Model Trained for Detection of Severity of DR

DR Grading	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)	F1 Score	Accuracy
No DR	92	90.5	70.7	97.8	0.800	0.908
Mild NPDR	44	90.5	53.6	86.6	0.483	0.812
Moderate NPDR	67.3	76.4	41.9	90.2	0.517	0.746
Severe NPDR	57	91	61.2	89.4	0.590	0.842
PDR	33.3	100	100	85.8	0.500	0.868

DR = diabetic retinopathy; NPDR = nonproliferative diabetic retinopathy; PDR = proliferative diabetic retinopathy.

a study on the same dataset, classifying images into "DR/non-DR" categories based on the presence of microaneurysms. Their method achieved an AUPRC of 0.90, with an SN of 76% and an SP of 88%.[37] Furthermore, Seoud et al's[38] study, using traditional ML on the Messidor dataset, could detect DR with an AUPRC of 0.89, SN of 93.9%, and SP of 50%. Our model, with an SN of 76%, SP of 90%, and AUPRC of 0.90, demonstrates comparable efficacy in detecting DR using the same dataset. Considering the reduced time, minimal coding knowledge required, and lower costs associated with training AutoML models, they present a feasible option for ML model training.

The trained model for grading DR achieved an SN of 92% and an SP of 90.5% in detecting “no DR.” It demonstrated an SN of 44% and an SP of 90.5% in identifying “mild NPDR.” In the “moderate NPDR” category, the model reached an SN of 67.3% and an SP of 76.4%. For “severe NPDR,” SN of 57%, and an SP of 91% was observed. Lastly, in the “PDR” group, the model yielded an SN of 33.3% and an SP of 100%.

Python scripts that were generated by ChatGPT for analysis of data included the NumPy and Pandas libraries. ChatGPT generates scripts for various programming languages by combining its training with the specific input it receives. It has been trained on a vast dataset that includes examples from numerous programming languages. When prompted, ChatGPT applies the patterns it has learned to generate code that is syntactically and logically consistent with the requested programming language. It utilizes a database of programming syntax, conventions, and common coding practices to produce these scripts.[39] In the field of data science and AI, particularly during algorithm development, the Pandas and NumPy libraries are frequently used together. The Pandas library is typically employed for initial data handling and preprocessing. Once the data is cleaned and structured, the NumPy library is utilized for more intensive numerical computations. This combination facilitates transition from data preparation to mathematical operations, which is essential in algorithm development.[40]

Recent advancements have made AI tools more accessible and ubiquitous. In this study, Vertex AI was utilized for training ML models while ChatGPT was employed for image preprocessing and data analysis. This included generating R and Python scripts and utilizing the Data Analyst feature. ChatGPT successfully provided scripts for analyzing the confusion matrix generated by Vertex AI. For optimal outcomes and enhanced efficiency, providing ChatGPT with detailed descriptions and specifics of the raw

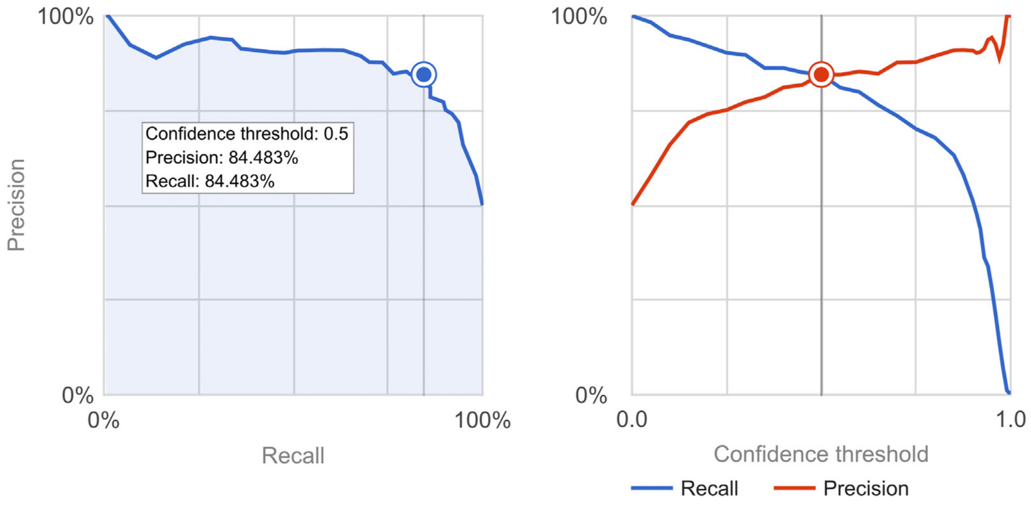

Figure 8. Precision-recall curves for binary detection of diabetic retinopathy.

Table 3. Precision and Recall at Different Confidence Thresholds for Binary Detection of Diabetic Retinopathy

Confidence Threshold	Precision (%)	Recall (%)
25%	75.10	90.20
50%	84.48	84.48
75%	87.80	70.10

data is recommended. However, errors may occur when running the scripts. In such cases, it is crucial to report these errors to ChatGPT and request a revised script. This process may need to be repeated until the expected results are achieved. Nonetheless, the Data Analyst feature could rapidly read and analyze data without the aforementioned issues associated with R and Python scripts. The integration of ChatGPT's processing capabilities with Vertex AI's ML resources demonstrated promising outcomes, highlighting how AI can enhance research and analysis in a user-friendly manner. This study provides insights into the practical applications of AI technologies into medicine, encouraging further exploration and adoption in diverse scientific domains.

Despite its benefits, ChatGPT does have certain limitations. For instance, when ChatGPT encounters identical queries, it may not always provide the exact same answer. Responses might vary, at least to some degree. This variation can be attributed to several factors. Even minor differences in how questions are phrased can lead to variations in the generated responses. Moreover, the order in which questions are presented and their context can influence ChatGPT's responses. Additionally, ChatGPT incorporates a degree of randomness, resulting in variations in its responses even when faced with the same question asked repeatedly.[41] However, large language models like ChatGPT predict each subsequent word based on the preceding context. This allows for multitude of ways to express the same idea with different phrasings. Take, for instance, instructing someone on using CLAHE in Fiji. While ChatGPT might offer varied explanations upon each inquiry, this diversity stems from the multiple methods available to accomplish the same task in Fiji. This principal of varied yet valid responses extends to other concepts and instructions as well. It is important to note that this does not imply that ChatGPT always provides the correct answer. ChatGPT draws on content from the internet, and if this content is inaccurate or misleading, ChatGPT may generate erroneous responses. Therefore, when using ChatGPT, one should be mindful of this possibility and exercise caution. Furthermore, different versions of ChatGPT (3.5 vs. 4.0) can produce different outputs. In a study by Eric Strong and colleagues, ChatGPT-3 was used to respond to free-response case-based clinical reasoning assessments. The study revealed that when given the same case 20 separate times, ChatGPT's performance on that case varied, with scores ranging from 56% to 81%. This indicates a significant degree of variability in ChatGPT's responses, even when faced with identical scenarios.[42] Another study aimed to evaluate ChatGPT's capacity for ongoing clinical decision support. The research involved inputting published clinical vignettes into ChatGPT-3.5 and assessing its accuracy in various areas such as differential diagnoses, diagnostic testing, final diagnosis, and management. The results showed that ChatGPT achieved a 71.7% accuracy overall across all vignettes. It demonstrated the highest performance in making a final diagnosis (76.9% accuracy) and the lowest in generating an initial differential diagnosis (60.3% accuracy).[43]

Another critical factor to consider is the most recent point at which ChatGPT had access to internet content. Although it has been updated with information up until April 2023, some of this data might be outdated today.[44]

Conclusion

ChatGPT and Vertex AI together provide an efficient AI implementation in medical research, allowing researchers and physicians to employ advanced AI for tasks like image analysis, diagnostics, and patient care without needing extensive programming skills.

Data Availability Statement

The introduced public datasets are available: Messidor-2 dataset at https://www.adcis.net/en/third-party/messidor2/ (accessed on July 10, 2023).

Acknowledgments

The Byers Eye Institute is supported by funding from the National Eye Institute (NEI P30-EY026877) and from the Research to Prevent Blindness, Inc.

Footnotes and Disclosures

Originally received: August 15, 2023.
Final revision: February 15, 2024.
Accepted: February 15, 2024.
Available online: February 21, 2024. Manuscript no. XOPS-D-23-00201.

Department of Ophthalmology, Byers Eye Institute, Stanford University, Palo Alto, California.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have made the following disclosure(s):

Q.D.N.: Payments to institution—Acelyrin, Alumis, Belite Bio, Novartis, Priovant, Regeneron; Consulting fees—Bausch and Lomb, Regeneron; Participation on a Data Safety Monitoring Board or Advisory Board—Alvotech; Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid—Tourmaline Scientific Advisory Board.

HUMAN SUBJECTS: No human subjects were included in this study.

This study adhered to the principles of the Declaration of Helsinki and was conducted using a publicly available deidentified dataset; therefore, institutional review board approval or informed consent was not required.

Author Contributions:

Conception and design: Mohammadi, Nguyen.

Data collection: Mohammadi, Nguyen

Analysis and interpretation: Mohammadi

Obtained funding: XXX

Overall responsibility: XXX

Abbreviations & Acronyms:
AI = artificial intelligence; **AUPRC** = area under the precision-recall curve; **AutoML** = automated machine learning; **CLAHE** = Contrast Limited Adaptive Histogram Equalization; **CSV** = comma-separated values; **DR** = diabetic retinopathy; **Messidor** = Methods to Evaluate Segmentation and Indexing Techniques in the field of Retinal Ophthalmology; **ML** = machine learning; **NPDR** = nonproliferative diabetic retinopathy; **NPV** = negative predictive value; **PDR** = proliferative diabetic retinopathy; **PPV** = positive predictive value; **SN** = sensitivity; **SP** = specificity.

Keywords:
artificial intelligence, ChatGPT, Generative Pretrained Transformer, image classification, machine learning.

Correspondence:
Quan Dong Nguyen, MD, MSc, Spencer Center for Vision Research, Byers Eye Institute at Stanford University, 2370 Watson Court, Suite 200, Palo Alto, CA 94303. E-mail: ndquan@stanford.edu.

References

1. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell*. 2023;6:1169595.
2. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9:e45312.
3. Liebrenz M, Schleifer R, Buadze A, et al. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. *Lancet Digit Health*. 2023;5:e105–e106.
4. How to use ChatGPT's advanced data analysis feature. https://mitsloanedtech.mit.edu/ai/tools/data-analysis/how-to-use-chatgpts-advanced-data-analysis-feature/. Accessed January 9, 2024.
5. Ghaffar Nia N, Kaplanoglu E, Nasab A. Evaluation of artificial intelligence techniques in disease diagnosis and prediction. *Discov Artif Intell*. 2023;3:5.
6. Pinto-Coelho L. How artificial intelligence is shaping medical imaging technology: a survey of innovations and applications. *Bioengineering (Basel)*. 2023;10:1435.
7. Dai L, Wu L, Li H, et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nat Commun*. 2021;12:3242.
8. Morton CE, Smith SF, Lwin T, et al. Computer programming: should medical students be learning it? *JMIR Med Educ*. 2019;5:e11940.
9. Touma S, Antaki F, Duval R. Development of a code-free machine learning model for the classification of cataract surgery phases. *Sci Rep*. 2022;12:2398.
10. Gong EJ, Bang CS, Lee JJ, et al. No-code platform-based deep-learning models for prediction of colorectal polyp histology from white-light endoscopy images: development and performance verification. *J Pers Med*. 2022;12:963.
11. Eisma JH, Dulle JE, Fort PE. Current knowledge on diabetic retinopathy from human donor tissues. *World J Diabetes*. 2015;6:312–320.
12. Hendrick AM, Gibson MV, Kulshreshtha A. Diabetic retinopathy. *Prim Care*. 2015;42:451–464.
13. Bresnick GH, Mukamel DB, Dickinson JC, Cole DR. A screening approach to the surveillance of patients with diabetes for the presence of vision-threatening retinopathy. *Ophthalmology*. 2000;107:19–24.
14. Kinyoun J, Martin D, Fujimoto W, Leonetti D. Ophthalmoscopy versus fundus photographs for detecting and grading diabetic retinopathy. *Invest Ophthalmol Vis Sci*. 1992;33:1888–1893.
15. Early photocoagulation for diabetic retinopathy: ETDRS report number 9. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology*. 1991;98:766–785.
16. Stratton IM, Kohner EM, Aldington SJ, et al. UKPDS 50: risk factors for incidence and progression of retinopathy in type II diabetes over 6 years from diagnosis. *Diabetologia*. 2001;44:156–163.
17. Effect of intensive diabetes treatment on the development and progression of long-term complications in adolescents with insulin-dependent diabetes mellitus: diabetes control and complications trial. Diabetes Control and Complications Trial Research Group. *J Pediatr*. 1994;125:177–188.
18. Tomita Y, Lee D, Tsubota K, et al. Updates on the current treatments for diabetic retinopathy and possibility of future oral therapy. *J Clin Med*. 2021;10:4666.
19. Messidor-ADCIS. http://wwwadcisnet/en/third-party/messidor2/. Accessed July 1, 2023.
20. Decencière E, Zhang X, Cazuguel G, et al. FEEDBACK ON A PUBLICLY DISTRIBUTED IMAGE DATABASE: THE MESSIDOR DATABASE. Diabetic retinopathy; image database; image processing; Messidor. *Image Anal Stereol*. 2014;33:4.
21. Krause J, Gulshan V, Rahimy E, et al. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. *Ophthalmology*. 2018;125:1264–1272.
22. Singh P, Mukundan R, De Ryke R. Feature enhancement in medical ultrasound videos using contrast-limited adaptive histogram equalization. *J Digit Imaging*. 2020;33:273–285.
23. Hitam MS, Awalludin EA, Yussof WNJHW, Bachok Z. *Mixture contrast limited adaptive histogram equalization for underwater image enhancement*. 2013 International Conference on Computer Applications Technology (ICCAT); 2013, 20-22 January. 2013.
24. Alwakid G, Gouda W, Humayun M. Deep learning-based prediction of diabetic retinopathy using CLAHE and ESRGAN for enhancement. *Healthcare (Basel)*. 2023;11:863.
25. Tondin BR, Barth AL, Sanches PRS, et al. Development of an Automatic Antibiogram Reader System Using Circular Hough Transform and Radial Profile Analysis. In: *XXVII Brazilian Congress on Biomedical Engineering*. Springer International Publishing; 2022:1837–1842.
26. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9:676–682.

27. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*. 2017;124: 962−969.
28. Abbas Q, Fondon I, Sarmiento A, et al. Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features. *Med Biol Eng Comput*. 2017;55: 1959−1974.
29. Nneji GU, Cai J, Deng J, et al. Identification of diabetic retinopathy using weighted fusion deep learning based on dual-channel fundus scans. *Diagnostics (Basel)*. 2022;12:540.
30. Budach L, Feuerpfeil M, Ihde N, et al. The effects of data quality on machine learning performance. *arXiv*. 2022. https://doi.org/10.48550/arXiv.2207.14529.
31. Saponara S, Elhanashi A. Impact of Image Resizing on Deep Learning Detectors for Training Time and Model Performance. In: *International Conference on Applications in Electronics Pervading Industry, Environment and Society*. Springer; 2021: 10−17.
32. Hao Y, Pei H, Lyu Y, et al. Understanding the Impact of Image Quality and Distance of Objects to Object Detection Performance. In: *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2023: 11436−11442.
33. Jeong Y, Hong YJ, Han JH. Review of machine learning applications using retinal fundus images. *Diagnostics (Basel)*. 2022;12:134.
34. Hayati M, Muchtar K, Maulina N, et al. Impact of CLAHE-based image enhancement for diabetic retinopathy classification through deep learning. *Procedia Comput Sci*. 2023;216: 57−66.
35. Faes L, Wagner SK, Fu DJ, et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. *Lancet Digit Health*. 2019;1:e232−e242.
36. Sánchez CI, Niemeijer M, Dumitrescu AV, et al. Evaluation of a computer-aided diagnosis system for diabetic retinopathy screening on public data. *Invest Ophthalmol Vis Sci*. 2011;52: 4866−4871.
37. Antal B, Hajdu A. An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. *IEEE Trans Biomed Eng*. 2012;59:1720−1726.
38. Seoud L, Hurtut T, Chelbi J, et al. Red lesion detection using dynamic shape features for diabetic retinopathy screening. *IEEE Trans Med Imaging*. 2016;35:1116−1126.
39. Buscemi A. A comparative study of code generation using ChatGPT 3.5 across 10 programming languages. *arXiv*. 2023. https://doi.org/10.48550/arXiv.2308.04477.
40. Nelli F. *Python Data Analytics: With Pandas, NumPy, and Matplotlib*. Second edition. Berkeley, CA: Apress; 2018.
41. Krügel S, Ostermaier A, Uhl M. ChatGPT's inconsistent moral advice influences users' judgment. *Sci Rep*. 2023;13:4569.
42. Strong E, DiGiammarino A, Weng Y, et al. Performance of ChatGPT on free-response, clinical reasoning exams. *medRxiv*. 2023. https://doi.org/10.1101/2023.03.24.23287731.
43. Rao A, Pang M, Kim J, et al. Assessing the utility of ChatGPT throughout the entire clinical workflow. *medRxiv*. 2023. https://doi.org/10.1101/2023.02.21.23285886.
44. Open AI. We've made ChatGPT Plus fresher and simpler to use. https://openai.com/blog/introducing-gpts. Accessed January 19, 2024.