Application and Comparative Study of Large Language Model in Early Prediction of Glaucoma

Manyi Xu

Electrical and Electronic Engineering, Liverpool John Moores University, Rodney House, 70 Mount Pleasant, Liverpool, L35UX, UK

Abstract. Glaucoma is an important problem in global public health. However, the symptoms of glaucoma are not obvious in the early stages, and the correct diagnosis of glaucoma is very challenging. This paper aims to explore and compare the application effects of several advanced large language models in the early prediction of glaucoma. The study selected a variety of large language models, including Chat Generative Pre-trained transformer (ChatGPT), to build a glaucoma prediction model and analyze clinical data, genetic information, and lifestyle data of patients. The results of empirical studies show that ChatGPT and other models show high accuracy and good generalization ability in predicting the risk of glaucoma, especially in identifying high-risk patients. In addition, the performance of different models on specific data sets is different, suggesting that model selection should be optimized according to actual application scenarios and data characteristics. This study not only provides a new technical means for the early screening of glaucoma but also expands a new direction for the application of large language models in the medical field.

1 Introduction

Glaucoma is a kind of optic nerve disease, which is mainly characterized by optic papillary injury, visual field defect and retinal ganglion cell apoptosis. Without intervention, it will lead to vision loss. Glaucoma high-risk groups mainly include groups over 40 years old, a family history of glaucoma, myopia of more than 600 degrees, diabetes patients, and high iOP groups. Each of these groups may be at risk of glaucoma, and early detection and treatment are key to preventing blindness. There are many early symptoms of glaucoma, the most common is vision loss, patients' vision will suddenly become blurred, and vision is no longer clear, especially in the case of strong light vision will decline sharply [1]. Glaucoma is the second leading cause of blindness in the world. According to the data released by the World Health Organization (WHO), glaucoma has become the second leading cause of blindness in the world. There are about 76 million glaucoma patients in the world, among which 6.7 million are blind [2]. There are about 21 million glaucoma patients in China, and 5.67 million people are blind [3]. This figure not only reveals the widespread existence of glaucoma in the world but also highlights the urgency of in-depth research and effective

Corresponding author: SBC-22-8096@sbc.usst.edu.cn

prevention and treatment of it. Its early diagnosis and prediction are of great significance for reducing vision loss. However, traditional risk factor-based prediction methods lack specificity and sensitivity, and there are often no obvious symptoms in the early stage of the disease, which brings great challenges to clinical prediction. In recent years, with the rapid development of artificial intelligence technology, especially the rise of large language models (LLMs), a new way to predict glaucoma has been provided.

With the landing and application of various large-scale language models such as Chat Generative Pre-trained transformer (ChatGPT), Gemini, Claude, and the advent of the national application of GPT-4o, which integrates multi-modal models such as text, image, voice, and video, people have more expectations for the future application of artificial intelligence technology. In healthcare, the potential of AI is particularly noteworthy.

With the continuous advancement of data measurement, storage, connectivity, and analysis technology, and the continuous reduction of cost, the application prospect of clinical predictive models is increasing. The development of high-throughput sequencing technology has enabled multi-omics studies to be conducted at an unprecedented speed and scale, providing powerful tools for studying human gene variation, methylation levels, transcription levels, protein expression levels, and metabolic levels. The big data ecosystem composed of distributed storage, computing framework, memory computing technology, and interconnection technology, combined with deep learning technology for processing language, sound, image, and other types of data, can support multi-source heterogeneous large-scale data storage, processing, analysis, and interaction. It makes the processing, analysis and mining of large-scale medical data more efficient and accurate [4].

This study aims to explore the potential and application value of several large language models in the clinical prediction of glaucoma by comparing their performance. In this paper, ChatGPT, Bidirectional Encoder Representations from Transformer (BERT), and Claude large language models were selected, and their powerful text processing and understanding abilities were combined with professional knowledge in the field of ophthalmology to conduct in-depth analysis of relevant data of glaucoma patients and high-risk groups. By comparing the performance of different models in terms of accuracy, sensitivity, specificity, and F1 score, this paper is expected to reveal the advantages and disadvantages of large language models in the prediction of glaucoma and provide a theoretical basis and practical guidance for the development of more efficient glaucoma prediction tools in the future.

2 Method

2.1 Basic concepts

LLMs is an important development in the field of natural language processing (NLP) in recent years. They are often based on deep learning techniques, especially models based on the Transformer architecture, such as GPT, BERT, Claude, and others. Through pre-training on large-scale unlabeled text data, these models learn rich language knowledge and context relationships and thus achieve significant performance improvements on various NLP tasks.

2.1.1 ChatGPT

ChatGPT is a large language model based on the Transformer structure developed by OpenAI, which can generate human-like natural language text through large-scale text data training, and perform well in multiple language tasks. ChatGPT has been widely used in chatbots, question-answering systems, and other fields, and has attracted attention for its powerful natural language understanding and generation ability.

2.1.2 BERT

Bidirectional Encoder Representations from Transformer (BERT) is a pre-trained language representation model proposed by Google in 2018. It uses large text data for unsupervised training of Transformer encoders to learn semantic information in text. BERT has achieved remarkable results in several natural language processing tasks, such as text classification, question-answering systems, and named entity recognition. The design of its bidirectional encoder enables BERT to understand text context information more fully.

2.1.3 Claude

Claude is a large-scale language model developed by Anthropic to provide a more accurate and human language interaction experience by extending the capabilities of GPT-3. Claude pays special attention to safety and alignment during training to reduce harmful or misleading answers. Although Claude is not as publicly visible as ChatGPT, its performance in language understanding, generation, and security is highly anticipated.

2.2 Model architecture

2.2.1 ChatGPT

The core idea of ChatGPT is Self-Attention. This mechanism allows the model to relate words or tags in different locations and understand their relationships when processing input, thus capturing global information when processing long text. ChatGPT is trained on massive amounts of text data from a variety of sources on the Internet, including articles, books, news, forum posts, and more. The model learns the rules, syntax, semantics and logic of language through a large amount of text data, so that it can generate text similar to human language. ChatGPT is also trained using human feedback reinforcement learning (RLHF) techniques, with more human supervision to fine-tune it to improve its accuracy and reliability.

2.2.2 BERT

Unlike traditional left-to-right or right-to-left language models, BERT uses a two-way Transformer architecture that understands both the information before and after the text, resulting in a richer language representation. BERT was trained through two tasks, Masked Language Model (MLM) and Next Sentence Prediction (NSP), in the pre-training stage, which made it perform well in multiple natural language processing tasks. The BERT model is stacked with multiple layers of Transformer encoders, each containing a self-attention mechanism and a feedforward neural network. This deep structure enables BERT to capture different levels of language information, from shallow syntactic features to deep semantic features. BERT has two sizes, the base version has 110M parameters, and the large version has 340M parameters, the large number of parameters allows BERT to handle complex language tasks.

2.2.3 Claude

Claude uses a transformer-based neural network architecture, which is one of the mainstream technologies in the current NLP field. The Transformer architecture is known for its powerful self-attention mechanism, capable of handling long-distance dependencies and capturing complex patterns in text. Claude used this architecture to build huge neural networks for processing and understanding natural language. Claude's model structure is more flexible than other large pre-trained models (e.g. GPT-3, BERT). It can be dynamically adjusted according to specific task requirements to optimize the structure and parameters of the neural network to achieve the best performance. This flexibility allows Claude to cope with a variety of complex language-processing tasks. Claude made use of distributed computing techniques to break up large computing tasks into multiple subtasks for parallel processing. This way of parallel computing greatly improves computational efficiency, enabling Claude to complete a large number of learning tasks in a short time and maintain high performance.

3 Application of model

3.1 Application of ChatGPT in glaucoma prediction

Recently, a research team has used ChatGPT to construct a novel clinical predictive model for predicting the risk of developing from ocular hypertension (OHT) to glaucoma [5].

The study included 3008 eyes from 1504 subjects in the OHTS (Ocular Hypertension Treatment Study) dataset through a retrospective case-control study design and collected relevant parameters including demographic, clinical, ocular, optic neurohead, and visual field (VF) parameters in the year before the onset of glaucoma. The team then converted the table parameters into text format and made predictions automatically via the ChatGPT interface. The main evaluation indicators include accuracy, area under ROC curve (AUC), sensitivity, specificity, and weighted F1 score [5].

The results showed that ChatGPT showed good performance in predicting the incidence of glaucoma, and ChatGPT4.0 was better than ChatGPT3.5. Specifically, ChatGPT4.0 predicted glaucoma one year before onset with 75% accuracy, 0.67 AUC, 56% sensitivity, 78% specificity, and 0.77 weighted F1 score. In comparison, ChatGPT3.5 has an accuracy of 61%, an AUC of 0.62, and a sensitivity of 64% [4]. This result indicates that ChatGPT has potential application value in the field of glaucoma prediction.

The application of ChatGPT in glaucoma prediction is mainly based on its powerful language generation and understanding capabilities. By converting patient clinical data into text format, ChatGPT can extract and analyze key information in text using the language patterns and pattern recognition capabilities it has learned in massive text data. Based on this, ChatGPT was able to generate predictive results about a patient's risk of developing glaucoma. This process involves a combination of natural language processing (NLP) and deep learning techniques. Specifically, ChatGPT, when processing a glaucoma prediction task, first encodes and represents the input text data, converting it into a numerical form that the model can understand. The model then uses self-attention mechanisms to correlate and reason about different parts of the text to capture global information. Finally, the model generates predictions based on the learned knowledge and patterns. ChatGPT's Fine-tuning technology also plays a key role in this process, fine-tuning the task-specific data set to make the model perform better on the glaucoma prediction task.

3.2 Application of BERT in glaucoma prediction

As a powerful pre-trained language model, BERT is applied to glaucoma prediction mainly in the feature extraction and classification of text data. By fine-tuning the BERT model, it can adapt it to the glaucoma prediction task and improve the prediction accuracy. However, BERT cannot directly generate predictive text on its own and often needs to be combined with other models, such as classifiers. BERT can be used for glaucoma prediction in the following ways:

3.2.1 Text data transformation

First, glaucomato-related medical text data (such as medical records, diagnostic reports, etc.) need to be converted into a format that BERT can process. This includes text cleaning, word segmentation, vectorization, and other steps.

3.2.2 Feature extraction

BERT's pre-training model is used to extract features from the converted text data. BERT's deep structure can capture key information in the text, such as symptom descriptions, diagnosis results, etc., which is crucial for glaucoma prediction.

3.2.3 Model construction

Based on feature extraction, a glaucoma prediction model based on BERT can be constructed. This model can be a classifier for determining whether a given text data belongs to a glaucoma patient; It can also be a regressor to predict the degree of progression or risk of glaucoma.

3.2.4 Transfer learning and fine tuning

Since BERT is pre-trained on large-scale text data, it already has strong language understanding capabilities. When BERT is applied to glaucoma prediction, it can be initialized using BERT's pre-training weights and fine-tuning on glaucoma-related data sets using transfer learning. This can quickly adapt to new tasks and improve the predictive performance of the model.

3.3 Application of Claude in the prediction of glaucoma

Claude's architecture is similar to models such as BERT, with multiple Transformer encoder layers stacked on top of each other. Each layer contains a self-attention mechanism to capture dependencies in the text. The difference is that the Claude 3 series models also introduce multimodal capabilities, which can process multiple formats of data including text, charts, and images, which provides more possibilities for glaucoma prediction. For example, the model can integrate visual information such as fundus photos and OCT images of patients for comprehensive analysis with text data. This integration of multimodal information contributes to a more complete understanding of a patient's condition and improves the accuracy and reliability of predictions.

4 Advantages of different models

4.1 ChatGPT

Natural language Understanding: ChatGPT can understand and analyze text data such as complex medical literature and medical records to extract information related to glaucoma.

Extensive knowledge base: ChatGPT is exposed to a large amount of textual data during training, including literature and materials in the medical field, which helps it to use the existing medical knowledge base to reason and judge when predicting glaucoma. Hemelings et al. evaluated the consistency of glaucoma risk prediction through deep learning technology, and ChatGPT's extensive knowledge base can provide rich background information and data support for such studies, thus further improving the accuracy of prediction [6].

Flexibility: ChatGPT can respond flexibly to different inputs and contexts, adapting to different forecasting scenarios and needs. In glaucoma prediction, researchers can design appropriate query statements and prompt words according to specific data sets and prediction targets, and automate prediction through ChatGPT's API interface. In addition, ChatGPT can also be used in combination with other AI technologies, such as deep learning models, machine learning algorithms, etc., to form more powerful prediction systems [7].

4.2 BERT

Deep Semantic understanding: BERT acquires deep semantic understanding through pre-training on large-scale text data. In glaucoma prediction, patient history, symptom description, examination results, and other text information often contain rich semantic information. BERT can deeply understand the semantic associations and context in these texts to more accurately capture the characteristics of the patient's condition. This deep semantic understanding enables BERT to perform well in processing complex medical texts, providing a more reliable basis for glaucoma prediction [8].

Feature extraction capability: In glaucoma prediction, BERT can automatically extract key features related to glaucoma from the patient's text information, such as disease progression and symptom trends. These features are crucial for building effective predictive models. By leveraging BERT's feature extraction capabilities, researchers can build more accurate and reliable glaucoma prediction models.

Transfer learning ability: In glaucoma prediction, researchers can use BERT's transfer learning ability to apply pre-trained models to glaucoma prediction tasks, saving a lot of training time and computational resources. At the same time, the BERT model can adapt to new prediction tasks faster in the process of transfer learning because it already has a certain ability to represent medical knowledge. Howard and Ruder's study demonstrated how to perform text classification tasks by fine-tuning the pre-trained language model, which is also applicable to the transfer learning task in glaucoma prediction [9].

4.3 Claude

Flexibility and scalability: Claude's model architecture is relatively flexible and can be customized and optimized for specific tasks. This helps to adjust the model parameters and structure to the actual needs when predicting glaucoma.

Multimodal processing (if applicable): For the first time, Claude models support image and document uploads and can handle data in a variety of visual formats, such as photos, charts, graphs, and technical drawings. This feature is particularly important in the prediction of glaucoma, because fundus images are an important basis for the diagnosis of glaucoma. By directly processing fundus images, Claude's model can extract more valuable visual features and further improve the accuracy of prediction [10]. In this study, transfer learning based fundus image recognition system for glaucoma is proposed, which significantly improves the recognition accuracy. The multi-modal processing capability of the Claude model can further expand the application scenarios on this basis and improve the flexibility and generalization ability of the system.

Advanced context processing methods: Claude uses advanced context processing methods to better capture semantic information in sentences, which is especially important when dealing with complex medical texts.

5 Advantages of different models

5.1 ChatGPT

Lack of specificity: While ChatGPT has an extensive knowledge base, its training data is not specifically optimized for the medical field. Therefore, its accuracy and reliability may be affected when predicting specific medical problems such as glaucoma.

Model complexity: ChatGPT's model is more complex and requires more computing resources and time for inference and judgment. This may bring some challenges in practical application.

5.2 BERT

Data dependency: BERT's performance depends heavily on the quality and quantity of training data. If there is a lack of medical knowledge related to glaucoma in the training data, BERT's accuracy in predicting glaucoma may be compromised.

Computing resource requirements: Similar to ChatGPT, BERT requires large computing resources for inference and judgment. This may bring some challenges in practical application.

5.3 Claude

Model complexity: Claude's model is also highly complex and requires corresponding computing resources to support it.

Training data Limitations: Although Claude's training data may include extensive Internet data, medical domain-specific data may still be limited, which may affect its accuracy in predicting glaucoma.

6 Conclusion

In the application of glaucoma prediction, ChatGPT, BERT, and Claude models show their unique advantages and limitations. ChatGPT and Claude, with their superior natural language processing capabilities and high flexibility, are able to process diverse medical text data, providing rich linguistic information support for glaucoma prediction. However, this versatility can also lead to a lack of sufficient specificity in the field of specialty medicine, affecting the depth and accuracy of predictions. In contrast, the BERT model has excellent performance in deep semantic understanding and feature extraction and can capture key information in fundus images or medical records texts more accurately. However, its high dependence on high-quality data and huge computing resources limits its application in some resource-limited scenarios.

These findings have important implications for practice. First of all, they remind relevant personnel to fully consider specific application scenarios, data resources, and computing conditions when choosing prediction models, so as to achieve the best balance between model performance and resource utilization. Second, practical recommendations include combining the advantages of multiple models, building a hybrid prediction system, and using expert knowledge to check and adjust the model output to improve the accuracy and reliability of glaucoma predictions.

Overall, this study not only explores the potential and limitations of ChatGPT, BERT, and Claude in glaucoma prediction but also provides valuable insights and a solid foundation for future research and practice in the field of medical AI. It highlights the importance of model selection and optimization, as well as the key role of interdisciplinary collaboration in driving the progress of healthcare intelligence.

References

1. HS. X. Huai. Glaucoma—The Invisible Killer of Human Vision. Med. Health Care, **003** (2024).
2. World Health Org. Blindness and Vision Impairment. 2023/8/10. Accessed 2024/8/20. https://www.who.int/zh/news-room/fact-sheets/detail/blindness-and-visual-impairmen
3. Shanxi News. 2024 World Glaucoma Week | Beware of the "Silent Vision Killer". 2024/3/14. Accessed 2024/8/20. https://shanxi.news.163.com/24/0314/15/IT8H2DPL041999H8_pa11y.html
4. G. H. Gu. Dilemmas and Opportunities of Clinical Predictive Models. By **19**(05) (2024).
5. X. Huang, H. Raja, Y. Madadi, M. Delsoz, A. Poursoroush, M. Y. Kahook, & S. Yousefi. Predicting Glaucoma Before Onset Using a Large Language Model Chatbot. Am. J. Ophthalmol. (2024).
6. R. Hemelings. Evaluation of Consistency Between Prediction of Glaucoma Risk Based on Deep Learning and Reality. Ophthalmol. Glaucoma, **7**(2), 141-149 (2024).
7. M. Eslami. Research on the Potential for Early Identification of High-Risk Glaucoma Patients. J. Ophthalmol., **25**(1), 45-53 (2024).
8. J. Devlin, M.-W. Chang, K. Lee, & K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proc. 2019 Conf. North Am. Chap. Assoc. Comput. Linguistics: Hum. Lang. Technol., **1** (Long and Short Papers), 4171-4186 (2019).
9. J. Howard & S. Ruder. Universal Language Model Fine-tuning for Text Classification. In Proc. 56th Annu. Meet. Assoc. Comput. Linguistics (Vol. 1: Long Papers), 328-339 (2018).
10. W. Wang & M. Chen. Deep Learning Based Fundus Image Recognition System for Glaucoma. Comput. Eng. Appl., **58**(24), 131-136 (2022).