Original Article

HEALTH SCIENCES MEDICINE

DOI: 10.32322/jhsm.1506378

J Health Sci Med. 2024;7(5):538-542

Diagnosing retinal disorders with artificial intelligence: the role of large language models in interpreting pattern electroretinography data

Aslan Aykut[1,2], Büşra Akgün[2], Almila Sarıgül Sezenöz[1,3], Mehmet Orkun Sevik[2], Özlem Şahin[2]

[1]*Department of Ophthalmology and Visual Sciences, Kellogg Eye Center, University of Michigan, Ann Arbor, US*

[2]*Department of Ophthalmology, Faculty of Medicine, Marmara University, İstanbul, Turkiye*

[3]*Department of Ophthalmology, Faculty of Medicine, Başkent University, Ankara, Turkiye*

Cite this article as: Aykut A, Akgün B, Sarıgül Sezenöz A, Sevik MO, Şahin Ö. Diagnosing retinal disorders with artificial intelligence: the role of large language models in interpreting pattern electroretinography data. *J Health Sci Med.* 2024;7(5):538-542.

Received: 28.06.2024 ◆ **Accepted**: 30.08.2024 ◆ **Published**: 27.09.2024

ABSTRACT

Aims: To evaluate the diagnostic accuracy of Claude-3, a large language model, in detecting pathological features and diagnosing retinitis pigmentosa and cone-rod dystrophy using pattern electroretinography data.

Methods: A subset of pattern electroretinography measurements from healthy individuals, patients with retinitis pigmentosa and cone-rod dystrophy was randomly selected from the PERG-IOBA dataset. The pattern electroretinography and clinical data, including age, gender, visual acuities, were provided to Claude-3 for analysis and diagnostic predictions. The model's accuracy was assessed in two scenarios: "first choice," evaluating the accuracy of the primary differential diagnosis and "top 3," evaluating whether the correct diagnosis was included within the top three differential diagnoses.

Results: A total of 46 subjects were included in the study: 20 healthy individuals, 13 patients with retinitis pigmentosa, 13 patients with cone-rod dystrophy. Claude-3 achieved 100% accuracy in detecting the presence or absence of pathology. In the "first choice" scenario, the model demonstrated moderate accuracy in diagnosing retinitis pigmentosa (61.5%) and cone-rod dystrophy (53.8%). However, in the "top 3" scenario, the model's performance significantly improved, with accuracies of 92.3% for retinitis pigmentosa and 76.9% for cone-rod dystrophy.

Conclusion: This is the first study to demonstrate the potential of large language models, specifically Claude-3, in analyzing pattern electroretinography data to diagnose retinal disorders. Despite some limitations, the model's high accuracy in detecting pathologies and distinguishing between specific diseases highlights the potential of large language models in ocular electrophysiology. Future research should focus on integrating multimodal data, and conducting comparative analyses with human experts.

Keywords: Retinitis pigmentosa, cone rod dystrophy, pattern electroretinography, large language models

INTRODUCTION

Pattern electroretinography (PERG) has been a valuable tool in ophthalmology for testing retinal ganglion cell function, photoreceptor health and diagnosing various retinal diseases by presenting alternating visual stimuli and recording the electrical responses from the retina.[1,2] PERG provides objective information regarding the health and integrity of the retinal pathway, with the macula and optic nerve in particular. Analysis of PERG waveforms, particularly the N95, P50, and N35 components, allows for the assessment of macular function and detection of abnormalities in the ganglion cell layer and inner retina.[1] The different PERG patterns observed in different diseases, such as the reduced amplitudes in retinitis pigmentosa (RP) and the delayed implicit times in glaucoma, aid in differential diagnosis and facilitate appropriate management strategies.[3,4] PERG offers a non-invasive and objective measure of retinal function, making it a valuable tool for monitoring disease progression and evaluating treatment efficacy.[5]

The emergence of artificial intelligence (AI) has revolutionized various aspects of medical diagnosis, offering promising approaches to interpreting complex medical data and assisting clinicians in making more informed decisions.[6] In the field of ophthalmology, AI algorithms have demonstrated remarkable capabilities in analyzing retinal images for diabetic retinopathy, age-related macular degeneration, and other retinal diseases.[7] Beyond analyzing retinal images, AI algorithms have been successfully used to predict glaucoma progression, automate visual field interpretation, and personalize treatment.[8,9] AI offers several advantages in medical diagnostics, including the ability to detect subtle patterns that human observers often miss, enable rapid and objective assessments, and improve diagnostic accuracy,

Corresponding Author: Aslan Aykut, aslanaykut81@gmail.com

leading to better patient outcomes.[10,11] However, despite these promises, implementing AI in healthcare still faces issues with privacy, algorithmic bias, and the requirement for thorough validation to ensure reliability and security.[12]

Large language models (LLMs) represent a unique and powerful subset of artificial intelligence that are trained on massive text and code datasets, enabling them to understand and generate human-like text, translate languages, and answer complex questions in an informative manner.[13] Unlike traditional AI models that excel at specific tasks such as image recognition or data classification, LLMs use their extensive training to process and understand information more holistic and contextual manner, mimicking the learning and reasoning processes of humans.[14] This ability to extract meaning, identify patterns, and draw conclusions from complex data sources makes them valuable tools for analyzing various medical data, including clinical notes, research articles and even genome sequences.[15-17]

Interpretation of pattern ERG data is a complex task that typically requires extensive training and expertise in ophthalmology and electrophysiology and often challenges even experienced clinicians.[18] The aim of this study is to evaluate the potential of Claude-3, a large language model accessible to a wider range of users, in analyzing pattern ERG data and providing diagnostic insights, potentially helping clinicians and researchers interpret this complex data.

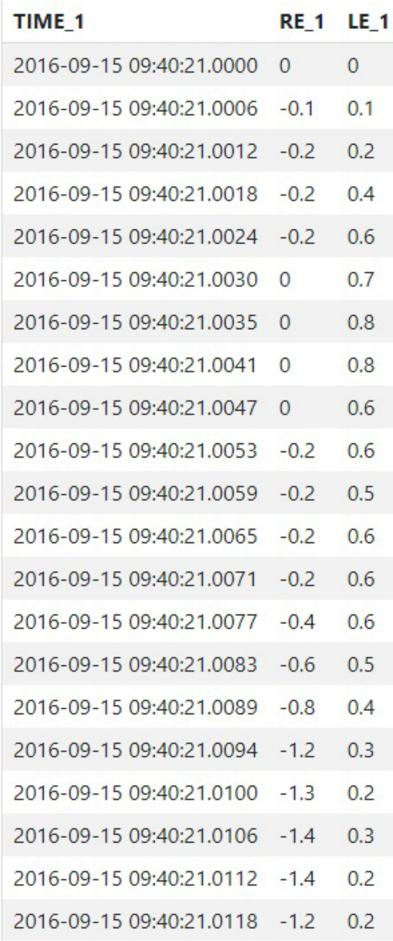
METHODS

Dataset

This study utilized the Pattern Electroretinogram-Institute of Applied Ophthalmobiology (PERG-IOBA) dataset available from PhysioNet, which serves as a research resource for complex physiologic signals.[19] Since this publicly available dataset from the was used in this study, ethical approval is not required. All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki. The terms of use of the database have been adhered to. The dataset includes 1,354 transient PERG responses from 304 subjects in 336 records, collected between 2003 and 2022.[20] It consists of 105 healthy subjects and 199 patients diagnosed with various retinal conditions. The most common diseases represented in the data set were RP with 48 patients, macular dystrophy with 32 patients, Stargardt disease with 16 patients and cone-rod dystrophy (CRD) with 14 patients. Clinical diagnosis, including age, gender, and visual acuity measurements in logMAR scale, was provided in CSV (comma separated values) format. The dataset had been anonymized, and data collection dates had been randomly date-shifted to maintain patient privacy and confidentiality.[20]

PERG Signal Acquisition

PERG signals in the dataset were captured by experienced technicians using the computerized Metrovision Optoelectronic Stimulator Vision Monitor MonPack 120 (Metrovision, Pérenchies, France). The acquisition protocol strictly adhered to the International Society for Clinical Electrophysiology of Vision (ISCEV) guidelines.[21] Signals were recorded at a high sampling rate of 1700 Hz over a duration of 150 milliseconds, producing 255 equally spaced observations per signal. Figure shows a portion of the raw data recorded from a patient.



TIME_1	RE_1	LE_1
2016-09-15 09:40:21.0000	0	0
2016-09-15 09:40:21.0006	-0.1	0.1
2016-09-15 09:40:21.0012	-0.2	0.2
2016-09-15 09:40:21.0018	-0.2	0.4
2016-09-15 09:40:21.0024	-0.2	0.6
2016-09-15 09:40:21.0030	0	0.7
2016-09-15 09:40:21.0035	0	0.8
2016-09-15 09:40:21.0041	0	0.8
2016-09-15 09:40:21.0047	0	0.6
2016-09-15 09:40:21.0053	-0.2	0.6
2016-09-15 09:40:21.0059	-0.2	0.5
2016-09-15 09:40:21.0065	-0.2	0.6
2016-09-15 09:40:21.0071	-0.2	0.6
2016-09-15 09:40:21.0077	-0.4	0.6
2016-09-15 09:40:21.0083	-0.6	0.5
2016-09-15 09:40:21.0089	-0.8	0.4
2016-09-15 09:40:21.0094	-1.2	0.3
2016-09-15 09:40:21.0100	-1.3	0.2
2016-09-15 09:40:21.0106	-1.4	0.3
2016-09-15 09:40:21.0112	-1.4	0.2
2016-09-15 09:40:21.0118	-1.2	0.2

Figure. A portion of the raw data recorded from a patient

Study Sample Selection

A specific subset of PERG measurements from healthy individuals, patients with RP and CRD were randomly selected from the larger dataset for this analysis. The PERG data for each subject in this subset were extracted from the dataset and provided to the Claude-3 language model for analysis and diagnostic predictions. The study was designed as a pilot study, hence the number of participants was kept limited.

Data Input to Claude-3 LLM

To evaluate the potential of large language models (LLMs) in analyzing PERG waveforms and providing diagnostic insights, we employed Claude-3, a commercially available LLM. We provided Claude-3 with a prompt that included the following instructions:

''Analyze the provided pattern ERG data for both eyes, identifying abnormalities in the N35, P50, and N95 waves, oscillatory potentials, and overall waveform morphology. Indicate the presence of pathology with a "Yes" or "No." If pathology is detected, select the top three differential diagnoses from a comprehensive list associated with pattern ERG features. Each diagnosis should include a detailed justification focusing on the bilateral ERG findings, particularly concerning both cone and rod functions, and consider the patient's current age, gender, and bilateral presentation. Rank these diagnoses by likelihood and outline the potential need for further tests or information to confirm these diagnoses.

Please include the following patient details:

- Age: [ ]
- Gender: [ ]
- Right Eye Visual Acuity (logMAR): [ ]
- Left Eye Visual Acuity (logMAR): [ ]

Note: The age at symptom onset is unknown. Given the complexity of diseases associated with pattern ERG abnormalities, additional clinical data, imaging, or tests may be necessary for a definitive diagnosis."

Important Outcomes

First scenerio;

First choice accuracy: This metric assesses whether Claude-3 identified the correct pathology as the first differential diagnosis. This is critical for determining the model's precision in diagnosing the most likely condition without additional input.

Second scenerio;

Top three accuracy: This broader metric evaluates whether the correct diagnosis was included in the model's top three differential diagnoses. This measure reflects the model's ability to detect and prioritize potential diagnoses, which is critical for clinical settings where multiple potential diagnoses may be considered before reaching a final conclusion.

Statistical Analysis

To determine whether the variables were normally distributed, the Shapiro-Wilk normality test was used. Demographic characteristics (age and sex) and visual acuity measurements were summarized using mean and standard deviation for continuous variables and frequencies and percentages for categorical variables. One-way analysis of variance (ANOVA) was performed to compare the mean age of the three groups, and a chi-square test was used to assess differences in gender distribution. To compare visual acuity between groups, we used the Kruskal-Wallis test with Bonferroni correction for multiple comparisons. Sensitivity, specificity, precision and F1 scores were calculated for each study group. Statistical analyzes were performed using Statistical Package for Social Sciences (SPSS) software, version 25.0 (IBM, Chicago, IL, USA). Values of $p<0.05$ were considered statistically significant in all tests.

RESULTS

Demographics and Visual Acuity

A total of 46 subjects were included in the study, with 20 healthy individuals, 13 patients with RP, and 13 patients with CRD. The demographic characteristics and visual acuity measurements for each group are summarized in Table 1.

Table 1. Demographics and visual acuity measurements for each group

Characteristic	Retinitis pigmentosa (n=13)	Cone-rod dystrophy (n=13)	Normal (n=20)	p
Age (years)				
Mean±SD	35.5±15.8	34.2±15.6	28.8±18.3	0.48
Range	(12-62)	(10-61)	(6-70)	
Gender				
Male	9 (69.2%)	8 (61.5%)	9 (45.0%)	0.38
Female	4 (30.8%)	5 (38.5%)	11 (55.0%)	
Mean visual acuity (LogMAR)				
Mean±SD	0.39±0.44	0.51±0.37	0.15±0.22	0.0015*

p*: significant, Post-hoc Dunn test (Bonferroni correction): RP vs. CRD: p>0.327 RP vs. Normal: p<0.023 CRD vs. Normal: p<0.011 SD: Standard deviation, LogMAR: Logarithm of the minimum angle of resolution, RP: Retinitis pigmentosa, CRD: Cone-rode dystrophy

Diagnostic Accuracy

The diagnostic accuracy of Claude-3 in detecting the presence or absence of pathology in all study groups was successful and all cases were correctly identified. In the normal group, the model confirmed no pathology in all 20 (100%) subjects. Similarly, in the pathologic group, the model confirmed pathology in all 26 (100%) subjects. The accuracy of the differential diagnosis showed variability, with RP and CRD in the first scenerio having an accuracy of 61.5% (8 of 13 cases) and 53.8% (7 of 13 cases), respectively. Notably, the model performed better when we used the second scenerio, with RP and CRD achieving higher success rates of 92.3% (12 of 13 cases) and 76.9% (10 of 13 cases), respectively. However there was no statistical difference when comparing model's accuracy of the differential diagnosis between RP and CRD in both scenarios (p=1 and 0.59), respectively.

Performance Metrics

We evauleted performance metrics acording to 2 scenarios. Results are given in Table 2.

Table 2. Performance metrics for Claude-3 diagnosis

Performance metric	Healthy vs. pathologic	RP	CRD
Sensitivity	100%	69.23%	53.85%
Specificity	100%	100%	84.62%
Precision	100%	100%	77.78%
F1 score	100%	81.82%	63.64%

RP: Retinitis pigmentosa, CRD: Cone-rode dystrophy

DISCUSSION

Our results suggest that Claude-3 achieves perfect performance, with 100% sensitivity, specificity, precision, and F1 score, demonstrating its ability to accurately identify all cases with pathology as well as all healthy cases without any misclassifications. It can effectively distinguish between healthy subjects and those with retinal diseases, achieving 100% accuracy in detecting the presence or absence of disease based on PERG data with minimal clinical data. The ability to accurately differentiate between healthy and pathologic cases is crucial in a clinical setting, as it can help prioritize patients who require further diagnostic evaluation and potential treatment.[22]

PERG is a highly valuable diagnostic tool in the evaluation of RP, a genetic disorder that causes progressive retinal degeneration.[4] In RP, PERG waveforms typically exhibit reduced amplitudes which is due to impaired function of photoreceptors and retinal ganglion cells.[2] These abnormalities can be detected even in early stages of the disease when visual acuity is still preserved, making PERG a sensitive tool for early diagnosis and monitoring disease progression.[23]

PERG is particularly informative in diagnosing and monitoring CRDs, a group of inherited retinal diseases characterized by deterioration of cone and rod photoreceptors.[24] The PERG can provide detailed assessments of cone function, which is crucial in cone-rod dystrophies where cone dysfunction typically presents before rod dysfunction.[24] For example, PERG can help distinguish different patterns of visual impairment in patients with cone-rod dystrophy, with some having more severe cone dysfunction than others.[24] This functional assessment is consistent with clinical observations and genetic findings, thereby supporting the diagnosis and understanding the disease progression in these patients.[25]

When considering the first scenario, Claude-3 demonstrated moderate accuracy in diagnosing RP (61.5%) and CRD (53.8%). However, when evaluating the second scenario, the model's performance significantly improved, with accuracies of 92.3% for RP and 76.9% for CRD. This suggests that Claude-3 is capable of identifying the correct diagnosis within the top three suggestions, even if it may not always be the first choice. These results are promising, indicating the potential of LLMs in analyzing PERG data for the diagnosis of retinal disorders. However, our literature search did not yield any studies done with artificial intelligence specifically using LLMs on this subject; therefore, we cannot directly compare our results to previous findings.

However, the model's performance metrics in identifying specific retinal disorders based on the first scenario varied between RP and CRD. While Claude-3 showed good performance in identifying RP cases, with high specificity and precision, its performance in identifying CRD cases was moderate, with lower sensitivity, specificity, and precision. This difference in performance may be attributed to the heterogeneity of CRD phenotypes and the overlap of PERG features with other retinal disorders, making it more challenging for the model to accurately identify CRD cases based solely on the first-choice diagnosis.[24]

Integrating AI into clinical practice offers several benefits, such as providing rapid, objective assessments of complex medical data and detecting subtle patterns that may be overlooked by human observers.[11] However despite its promise one of the major concern is the "black box" nature of these models, where the reasoning behind their predictions remains opaque.[26] In this models training data are often obscured or undocumented, and their methods opaque.[27] This lack of transparency can affect trust and acceptance among clinicians, particularly when dealing with complex medical decisions.[28]

Our study has several strengths and limitations. One of the strengths is the use of a large, well-characterized data set (PERG-IOBA) that conforms to the ISCEV guidelines for PERG collection, ensuring data reliability and consistency.[20] Another strength of our study is the use of a commercially available large language model, Claude-3, which is accessible to a wider range of users compared to specialized AI models that require extensive technical expertise. This accessibility enables greater potential in clinical settings, as healthcare professionals without strong AI knowledge can still benefit from the model's insights. However, our study has notable limitations, the most significant being the relatively small sample size, as it was designed as a pilot study. Additionally, the study focuses on a specific subset of retinal diseases, and while Claude-3 shows promising results in analyzing PERG data, its performance for other types of ocular electrophysiological tests and different retinal diseases remains to be investigated.

There are several important directions for future research in this area. First, the integration of PERG data with other diagnostic modalities such as optical coherence tomography and visual field testing may represent a significant advance toward a multimodal diagnostic approach. By combining data from these different sources, LLMs could provide a more comprehensive and nuanced understanding of retinal health and improve the ability to diagnose complex conditions that may not be detectable with a single diagnostic method. Conducting comparative analysis between the performances of LLMs and human experts is also crucial. Such studies would help delineate the strengths and limitations of each approach and provide insights into how best to use AI in clinical settings. By directly comparing AI with human diagnostics, researchers can identify specific scenarios where AI excels or lags behind, thereby refining AI applications to effectively support clinical decision making.[29]

CONCLUSION

This study is the first to demonstrate the potential of large language models, particularly Claude-3, in analyzing PERG data for the diagnosis of retinal diseases. Despite some limitations, the model's high accuracy in detecting pathologies and distinguishing between specific diseases highlights the potential of AI in ophthalmology. Future research should focus on addressing limitations.

ETHICAL DECLARATIONS

Ethics Committee Approval

Since the PERG IOBA dataset from the PhysioNet database was used in this study, ethical approval is not required. The terms of use of the database have been adhered to.

Informed Consent

Since the PERG IOBA dataset from the PhysioNet database was used in this study, informed consent is not required.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

Acknowledgement

Our research's data was presented in '15th Medical Informatics Congress' as 'Oral Presentation' on May 30, 2024.

REFERENCES

1. Thompson DA, Bach M, McAnany JJ, Šuštar Habjan M, Viswanathan S, Robson AG. ISCEV standard for clinical pattern electroretinography (2024 update). *Doc Ophthalmol*. 2024; 148(2):75-85. doi:10.1007/s10633-024-09970-1
2. Robson AG, El-Amir A, Bailey C, et al. Pattern ERG correlates of abnormal fundus autofluorescence in patients with retinitis pigmentosa and normal visual acuity. *Invest Ophthalmol Vis Sci*. 2003;44(8):3544-3550. doi:10.1167/iovs.02-1278
3. Gallo Afflitto G, Chou TH, Swaminathan SS, et al. Pattern electroretinogram in ocular hypertension, glaucoma suspect and early manifest glaucoma eyes: a systematic review and meta-analysis. *Ophthalmol Sci*. 2023;3(4):100322. doi:10.1016/j.xops.2023.100322
4. Janáky M, Pálffy A, Horváth G, Tuboly G, Benedek G. Pattern-reversal electroretinograms and visual evoked potentials in retinitis pigmentosa. *Doc Ophthalmol*. 2008;117(1):27-36. doi:10.1007/s10633-007-9099-0
5. Robson AG, Nilsson J, Li S, et al. ISCEV guide to visual electrodiagnostic procedures. *Doc Ophthalmol*. 2018;136(1):1-26. doi:10.1007/s10633-017-9621-y
6. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng*. 2018;2(10):719-731. doi:10.1038/s41551-018-0305-z
7. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2): 167-175. doi:10.1136/bjophthalmol-2018-313173
8. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*. 2016;316(22):2402-2410. doi:10.1001/jama.2016.17216
9. Asaoka R, Murata H, Hirasawa K, et al. Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images. *Am J Ophthalmol*. 2019;198:136-145. doi:10.1016/j.ajo.2018.10.007
10. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. 2019;1(6):e271-e297. doi:10.1016/S2589-7500(19)30123-2
11. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;577(7788):89-94. doi:10.1038/s41586-019-1799-6
12. Char DS, Abràmoff MD, Feudtner C. Identifying ethical considerations for machine learning healthcare applications. *Am J Bioeth*. 2020;20(11):7-17. doi:10.1080/15265161.2020.1819469
13. Raffel C, Shazeer NM, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *J Mach Learn Res*. 2019;21(140):1-67.
14. Head CB, Jasper P, McConnachie M, Raftree L, Higdon G. Large language model applications for evaluation: opportunities and ethical implications. *N Direct Evaluat*. 2023;2023(178-179):33-46. doi:10.1002/ev.20556
15. Meng X, Yan X, Zhang K, et al. The application of large language models in medicine: a scoping review. *iScience*. 2024;27(5): 109713. doi:10.1016/j.isci.2024.109713
16. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature*. 2023;620(7972):172-180. doi:10.1038/s41586-023-06291-2
17. Wu J, Ma Y, Wang J, Xiao M. The application of chatgpt in medicine: a scoping review and bibliometric analysis. *J Multidiscip Healthc*. 2024;17:1681-1692. doi:10.2147/JMDH.S463128
18. Yap GH, Chen LY, Png R, et al. Clinical value of electrophysiology in determining the diagnosis of visual dysfunction in neuro-ophthalmology patients. *Doc Ophthalmol*. 2015;131(3):189-96. doi:10.1007/s10633-015-9515-9
19. Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*. 2000; 101(23):E215-220. doi:10.1161/01.cir.101.23.e215
20. Fernández I, Cuadrado Asensio R, Larriba Y, Rueda C, Coco-Martin RM. A comprehensive dataset of pattern electroretinograms for ocular electrophysiology research: the PERG-IOBA dataset (version 1.0.0). *PhysioNet*. 2024. doi:10.13026/d24m-w054
21. Bach M, Brigell MG, Hawlina M, et al. ISCEV standard for clinical pattern electroretinography (PERG): 2012 update. *Doc Ophthalmol*. 2013;126(1):1-7. doi:10.1007/s10633-012-9353-y
22. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008;56(1):45-50. doi:10.4103/0301-4738.37595
23. Popović P, Jarc-Vidmar M, Hawlina M. Abnormal fundus autofluorescence in relation to retinal function in patients with retinitis pigmentosa. *Graefes Arch Clin Exp Ophthalmol*. 2005; 243(10):1018-1027. doi:10.1007/s00417-005-1186-x
24. Hamel CP. Cone rod dystrophies. *Orphanet J Rare Dis*. 2007;2:7. doi:10.1186/1750-1172-2-7
25. Downes SM, Payne AM, Kelsell RE, et al. Autosomal dominant cone-rod dystrophy with mutations in the guanylate cyclase 2D gene encoding retinal guanylate cyclase-1. *Arch Ophthalmol (Chicago, Ill : 1960)*. 2001;119(11):1667-1673. doi:10.1001/archopht.119.11.1667
26. Schwartz IS, Link KE, Daneshjou R, Cortés-Penfield N. Black box warning: large language models and the future of infectious diseases consultation. *Clin Infect Dis*. 2024;78(4):860-866. doi:10.1093/cid/ciad633
27. Harrer S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. *EBioMedicine*. 2023;90:104512. doi:10.1016/j.ebiom.2023.104512
28. Au Yeung J, Kraljevic Z, Luintel A, et al. AI chatbots not yet ready for clinical use. *Frontiers in digital health*. 2023;5:1161098. doi:10.3389/fdgth.2023.1161098
29. Rojas-Carabali W, Sen A, Agarwal A, et al. Chatbots Vs. Human experts: evaluating diagnostic performance of chatbots in uveitis and the perspectives on ai adoption in ophthalmology. *Ocul Immunol Inflamm*. 2023:1-8. doi:10.1080/09273948.2023.2266730