https://doi.org/10.31288/oftalmolzh202561317

Possibilities of artificial intelligence with the use of a structured algorithm in the diagnosis of keratoconus

Bezkorovaina I. M. [1], **Gontar A. R.**[1], **Nakonechnyi D. O.**[2], **Ivanchenko A. Iu.** [1]

[1] Poltava State Medical University
Poltava (Ukraine)

[2] Private Eye Clinic «Svitogliad»
Poltava (Ukraine)

Purpose: *To assess the efficacy of using artificial intelligence (AI) in the diagnosis of keratoconus based on corneal topography maps, with subsequent determination of disease stage and examining the impact of the level of standardization of the clinical algorithm on the accuracy and consistency of results.*

Material and Methods: *Fifty-nine corneal topography maps of 37 patients from the database of the ACE DIAGNOSTIC PLATFORM (Baush+Lomb) were retrospectively analyzed. The study was conducted in three stages which simulated various levels of AI integration, from a completely autonomous AI operation to the operation strictly according to a three-level structured algorithm that we have developed (Copyright for the Work No. 139012 issued 26.08.2025). The adapted language model of ChatGPT (OpenAI, 2023) was used for establishing the diagnosis and staging. Data were statistically analyzed using Microsoft Excel 2019 and Analysis ToolPak. Data analysis was conducted by solving binary and multi-class classification problems. Accuracy, sensitivity, specificity, precision and the F1-measure were assessed at each stage and compared across stages.*

Results: *The overall accuracy and accuracy of staging were 79.7% and 33.3%, respectively, at stage 1, and improved to 93.2% and 74.36%, respectively, at stage 2. The highest values of characteristics were seen at stage 3 (sensitivity, 100%; specificity, 95%; overall accuracy, 98%; and accuracy of staging, 89.7%).*

Conclusion: *AI demonstrated a potential in the automated diagnosis of keratoconus. However, the best efficacy was achieved only when AI was used in the presence of a clearly structured algorithm. Autonomous AI decision led to a high variation in results. Introduction of algorithmic logic enabled significant improvements in the accuracy and consistency of results.*

Keywords:

artificial intelligence, ChatGPT, keratoconus, corneal topography

Introduction

Keratoconus is a progressive degenerative corneal disorder that affects vision and can lead to visual disability if left untreated. It is often referred to as a silent disease because during its early stages, it can go undiagnosed for long periods or misdiagnosed as simple forms of myopia or astigmatism [1-4]. Keratoconus progression requires timely detection, because more severe stages often require surgical intervention, particularly cross-linking, implantation of segments or even keratoplasty [5-6].

Computerized corneal topography, a major technique for the diagnosis of keratoconus, is a corneal relief imaging technique that allows assessing the curvature, thickness and elevation of the anterior and posterior surfaces, and corneal symmetry [4, 6]. Despite the objectivity of the digital images obtained, their interpretation remains subjective and depends on the proficiency and experience of the doctor and the presence of clinical reasoning and algorithmic approach. In many clinics, especially primary care clinics, these factors are either absent or poorly implemented, which can result in the late diagnosis of keratoconus or excessive alertness.

Recently, there have been tremendous developments in artificial intelligence (AI) in medicine. In the field of ophthalmology, special attention has been given to studies aimed at diabetic retinopathy, glaucoma, macular degeneration, and keratoconus detection using deep learning and neuronal networks [7-10]. None of these developments, however, has been widely introduced into clinical practice. Previous studies [8-11] have demonstrated that large language models and neuronal networks have a high capacity for analysis of corneal topography images. However, findings of most of these studies have not been implemented into clinical practice due to difficulties in standardizing the decisions, lack of transparency of the logic and risk of erroneous conclusions. In this context, it is essential to combine AI with a clear logical analysis algorithm that takes into account all clinically relevant parameters and enables avoiding excessive system autonomy.

Therefore, developing a structural approach to the AI-based automated diagnosis of keratoconus is very important. Determining the fact of the presence and stage of the disease is of a key importance for the prognosis and treatment strategy.

The purpose of the study was to assess the efficacy of using AI in the diagnosis of keratoconus based on corneal topography maps, with subsequent determination of disease stage and examining the impact of the level of standardization of the clinical algorithm on the accuracy and consistency of results.

Material and Methods

This retrospective study was conducted at the Department of Otorhinolaryngology and Ophthalmology, Poltava State Medical University, and used a database of corneal topography data collected with ACE DIAGNOSTIC PLATFORM at the private eye clinic "Svitogliad", Poltava. Totally, 59 corneal topography maps meeting the digital processing quality criteria (image sharpness, correct alignment, and the absence of artifacts that could affect the results) were analyzed.

Thirty-seven patients (59 eyes) were included in the study. Of these, 27 were men and 10 were women. The keratoconus group comprised 39 eyes of patients with a clinically verified diagnosis of keratoconus (stage I to IV, based on the consensus) and the control group comprised 20 eyes of healthy individuals without any signs of corneal disease. Patients with pellucid marginal degeneration, cicatricial post-inflammatory changes, corneal injury or post-operative corneal changes were excluded. ChatGPT (OpenAI, 2023) language model with an interface access enabling presenting corneal topography maps as images directly for automatic identification and interpretation was used for the analysis. The analysis included three stages, each of which simulating a particular level of AI self-sufficiency in clinical diagnostics.

At stage 1, ChatGPT received only visual or numerical data from a corneal topography map (PachyMin, the thinnest corneal thickness) – shift of the thinnest point along the Y axis, Anterior Axial Curvature symmetry, anterior surface elevation, posterior surface elevation, Q-value (corneal asphericity), Kmax (maximal corneal curvature, astigmatism) – without any comment or instruction regarding the norms, diagnostics logic, or structured algorithm. The task was to determine to what extent the language model can independently classify keratoconus and its stage using its own knowledge base and logical reasoning only.

At stage 2, the model was given orienting instructions regarding major diagnostic parameters and risk ranges without direct formulation of the diagnostic algorithm. For example, the model knew that PachyMin < 470 µm is a potential marker of keratoconus, or that Kmax> 47.2 D may indicate the pathology. However, it was AI that had to interpret this data and form the final diagnosis.

At stage 3, the model operated strictly according to a three-level structured algorithm that we have developed (level 1, separate parameter analysis; level 2, determination of the pathology; level 3, staging) (Copyright for the Work No. 139012 issued 26.08.2025) [12], which provided a stage-by-stage analysis for each parameter, taking into account exact threshold values and level-to-level transition rules.

Data were calculated based on the comparison of results of ChatGPT and consensus-based clinical diagnosis. All calculations were made manually and subsequently cross-checked.

Data analysis was conducted by solving binary and multi-class classification problems. Classification quality was assessed using a set of metrics, including accuracy, sensitivity, specificity, precision and the F1-measure for binary classification, and macroaverage values of precision, sensitivity, F1-measure and the overall accuracy for multi-class classification. The study was conducted in three phases: training on unlabeled data; training on labeled data without staging; and training on labeled data with staging. The classification metrics obtained during all phases were compared. Decision consistency was assessed by F1-measure analysis at each stage of model work. Data were analyzed using Microsoft Excel 2019 and Analysis ToolPak.

Diagnostic efficacy metrics were calculated based on the quantitative characteristics of true positive, false positive, true negative and false negative results. 95% confidence intervals (CI) were calculated to assess the stability of the results obtained.

Results

For stage 1, AI correctly diagnosed 31 of 39 cases with keratoconus (sensitivity, 79.5%; 95% CI, 64.5-89.3), and the overall accuracy was 79.7%. Additionally, 16 of 20 control images were correctly found normal (specificity, 80.0%; 95% CI, 58.4-91.9). The accuracy of staging, however, was relatively low, with only 13 of 39 cases (33.3%) being correctly staged, with errors being chaotic. Table 1 presents the results of corneal topographic map processing by AI without instructions (stage 1). Repeat analysis by ChatGPT of these maps demonstrated a high variability: initial decision was changed in 17 cases, indicating the inconsistency of decisions.

At stage 2, with the introduction of orienting ranges for major diagnostic parameters, there was a significant

Table 1. Diagnostic accuracy of the artificial intelligence (AI) model at three stages of the study

	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)	Sensitivity	Specificity	Overall accuracy	Precision	F1-measure	Staging accuracy
Stage 1	31	16	4	8	79.5%	80.0%	79.7%	88.6%	83.8%	33.33%
Stage 2	36	19	1	3	92.3%	95.0%	93.2%	97.3%	94.7%	74.36%
Stage 3	39	19	1	0	100.0%	95.0%	98.0%	97.5%	98.7%	89.74%

improvement in all metrics. The model correctly identified 36 of 39 pathologic maps (sensitivity, 92.3%; 95% CI, 79.7-97.4) and 19 of 20 normal maps (specificity, 95.0%; 95% CI, 76.4-99.1). Overall accuracy and F1-measure increased to 93.2% and 94.7%, respectively, demonstrating the efficacy of even partial standardization of diagnostic criteria. The accuracy of staging increased almost threefold from 33.3% to 74.4% (32 of 39 cases were correctly staged). The results of the study for stage 2 are presented in table 1. Decision consistency, however, was still low: after all the maps were re-uploaded into the system for re-processing, the decision with respect to staging was changed in nine maps.

At this stage of the study, there were technical difficulties associated with the limitations of the free version of ChatGPT language model. Particularly, a limited number of queries, limited amount of context for analysis and impossibility of processing a large number of images during a session made full-value testing impossible. A transition to ChatGPT Plus (a paid subscription from OpenAI) provided priority access to GPT-4o, a more powerful model, allowed us to continue the study, remove the technical limitations, and enabled full-value implementation of the first and second stages of the analysis.

At stage 3 of the study, ChatGPT operated strictly according to our three-level algorithm, the model correctly diagnosed all 39 cases of keratoconus (sensitivity, 100.0%; 95%CI, 91.0-100.0). Additionally, 19 of 20 control images were correctly found normal (specificity, 95.0%). In the case of a healthy eye that was found by the algorithm to be subclinical, the numerical parameters of the corneal topographic map were as follows: PachyMin, 580 µm; shift of the thinnest point along the Y axis, 0.56; normal symmetry on the Anterior Axial Curvature map; Q-value, 0.24; Kmax, 43.64D, and astigmatism, 0.54D). This eye was classified as normal by clinical assessment and corneal topography. The overall accuracy was as high as 98.0% (95% CI: 91.0–99.7), and the F1-measure was 98.7%. The accuracy of staging substantially improved, with 35 of 39 cases (89.7%; 95% CI: 76.4–95.9) being correctly staged (Table 1).

Decision consistency was also assessed. Repeat processing by ChatGPT of all maps within a week (with the model operating strictly according to the three-level structured algorithm) did not change the results in all cases, which confirmed the absence of differences associated with the stochastic nature of the language model.

Discussion

Numerous studies reported that AI was highly effective and already about as good as physicians at making diagnoses of eye diseases, but further studies, particularly on the generalization, interpretability and adaptation to real-life conditions, are necessary for its wide implementation in clinical practice [8-10, 13]. In our study, ChatGPT demonstrated a significant potential in the analysis and interpretation of corneal topography data. It should be taken into account that AI as such did not provide a 100% diagnostic correctness. Only in the presence of clear logical frameworks the model can perform the functions close to the physician's clinical decisions. This is in agreement with findings of foreign studies that noted the effectiveness of deep learning-based models only in the presence of a limited and well-controlled impact of the human factor or, vice versa, in the presence of complete standardization of input data and logic [2, 4, 6-8].

It is interesting that the autonomous operation of AI at stage 1 showed acceptable sensitivity and specificity, but a low correct staging rate. This can be explained by the fact that, in the absence of algorithmic rigidity, ChatGPT tends to excessive interpret or miss important parameters, especially in questionable cases. Therefore, the introduction of a structured algorithm at the diagnostic decision level is not merely desirable, but absolutely necessary. Additionally, stage 3 the model operated strictly according to a three-level structured algorithm that we have developed, which allowed the model to identify key combinations of risks and separate subclinical, marginal and clinically apparent forms of keratoconus.

Consistency of analysis with AI was an important practical aspect. Consistency of interpretation has a critical value under clinical reality, especially in the dynamic monitoring of patients. The finding of inconsistency at initial stages confirmed that unaided AI should not be the only justification for establishing the diagnosis. The algorithm ensured complete consistency of results only after the introduction of fixed rules, which indicates a reduced likelihood of accidental errors and increased clinical reliability.

Of note that not all maps were unequivocal: a portion of patients had the characteristics that did not allow even the physician to establish a definite diagnosis without further observation. This is why a “patient under attention” category consistent with the concept of early risk stratification was introduced into the algorithm. This enables minimizing the number of false positive decisions that may result in too early treatment or psychological pressure on the patient.

Given an increasing role of AI in current medicine, findings of this study and the algorithm developed may contribute to subsequent developments in the area of hybrid expert systems, where the clinician-AI interaction will be based on the principles of complementation and mutual control, with a consistently key role in decision making played by the physician.

In conclusion, introducing AI in the diagnosis of keratoconus is important, because it improves the objectivity of medical decisions, reduces the impact of the human factor, and shortens the time required for making the diagnosis. This should be, however, done with the use of a correctly formed algorithmic approach, which enables a significant improvement in the accuracy of staging, consistency of diagnostic decisions made and clinical reliability of AI in practical ophthalmology. Further studies will include enlargement of the database, integration with other corneal examination techniques, and the development of hybrid expert systems in which the clinical decision is combined with algorithmic analytics of AI.

References

1. Santodomingo-Rubido J, Carracedo G, Suzaki A, Villa-Collar C, Vincent SJ, Wolffsohn JS. Cont Lens Anterior Eye. 2022 Jun;45(3):101559. doi: 10.1016/j.clae.2021.101559.
2. Lin SR, Ladas JG, Bahadur GG, Al-Hashimi S, Pineda R. A Review of Machine Learning Techniques for Keratoconus Detection and Refractive Surgery Screening. Semin Ophthalmol. 2019;34(4):317–326. doi: 10.1080/08820538.2019.162081.
3. Ferdi AC, Kandel H, Nguyen V, Tan J, Arnalich-Montiel F, Abbondanza M, Watson SL. Five-year corneal cross-linking outcomes: A Save Sight Keratoconus Registry Study. Clin Exp Ophthalmol. 2023 Jan;51(1):9-18. doi: 10.1111/ceo.14177.
4. Bui AD, Truong A, Pasricha ND, Indaram M. Keratoconus Diagnosis and Treatment: Recent Advances and Future Directions. Clin Ophthalmol. 2023;16(17):2705-2718. doi: 10.2147/OPTH.S 392665
5. Niazi S, Gatzioufas Z, Doroodgar F, Findl O, Baradaran-Rafii A, Liechty J, Moshirfar M. Keratoconus: exploring fundamentals and future perspectives – a comprehensive systematic review. Ther Adv Ophthalmol. 2024 Mar 20:16:25158414241232258.
6. Li Z, Wang L, Wu X, Jiang J, Qiang W, Xie H, et al. Artificial intelligence in ophthalmology: The path to the real-world clinic. Cell Rep Med. 2023;4(7):101095. doi: 10.1016/j.xcrm.2023.101095
7. Afifah A, Syafira F, Afladhanti PM, Dharmawidiarini D. Artificial intelligence as diagnostic modality for keratoconus: A systematic review and meta-analysis. J Taibah Univ Med Sci. 2023;19(2): 296-303.
8. Wan Q, Wei R, Ma K, Yin H, Deng YP, Tang J. Deep Learning-Based Automatic Diagnosis of Keratoconus with Corneal Endothelium Image. Ophthalmol Ther. 2023 Dec;12(6):3047-3065. doi: 10.3390/diagnostics 13162715.
9. Muhsin ZJ, Qahwaji R, Al Shawabkeh M, et al. Smart decision support system for keratoconus severity staging using corneal curvature and thinnest pachymetry indices. Eye Vis (Lond). 2024;11(1):28. doi: 10.1186/s 40662-024-00394-1.
10. Yousefi S, Yousefi E, Takahashi H, Hayashi T, Tampo H, Inoda S, et al. Keratoconus severity identification using unsupervised machine learning. PLoS One. 2018;13(11) e 0205998. doi: 10.1371/journal.pone.0205998.
11. Zhamardiy VO, Shkola OM, Okhrimenko IM, Strelchenko OG, Aloshyna AI, Opanasiuk FH, et al. Checking of the methodical system efficiency of fitness technologies application in students' physical education. Wiad Lek. 2020;73(2):332–341. doi: 10.36740/WLek202002125
12. https://surl.li/ztvper
13. Nevska AO, Pohosian OA, Goncharuk KO, Chernenko OO, Hymanyk IV, Korol AR. Assessing the possibility of using portable and stationary non-mydriatic fundus cameras for diabetic retinopathy screening assisted by an artificial intelligence-based software platform in primary care. J Ophthalmol (Ukraine). 2024;6(521): 22–26. https://doi.org/10.31288/oftalmolzh202462226.

Disclosures

Received: 28.07.2025

Accepted: 31.10.2025

Author’s contribution. *Bezkorovaina I.M. - design, methodology, writing, reviewing, and editing; Gontar A.R. - data collection, writing the algorithm, writing the article; Nakonechny D.O. - data collection, analysis of the algorithm application; Ivanchenko A.Iu. - editing, statistical analysis. All authors analyzed the results and agreed on the final version of the manuscript.*

Disclaimer: *The opinions expressed in this article are those of the authors and do not necessarily reflect the official position of the institution.*

Sources of support. *There are no external sources of funding.*

Conflict of interest. *The authors declare that there are no conflicts of interest that could influence their* *opinion on the subject or materials described and discussed in this manuscript.*

Data availability statement. *All data generated or analyzed during this study are included in this published article.*

List of abbreviations: *AI – artificial intelligence, ChatGPT – Generative Pre-trained Transformer, AI – artificial intelligence, CI – confidence intervals.*