Original Article

Role of artificial intelligence (Google bard) in morphological, histopathological, and radiological image identifications

Objective Structured Practical Examination (OSPE) type-based performance

Sultan A. Meo, *MBBS, PhD,* ***Abdulelah A. AbuKhalaf,*** *MBBS,* ***Muhammad Zain S. Meo,*** *MBBS,* ***Muhammad Omair S. Meo,*** *MBBS,* ***Rashid Ayub,*** *PhD,* ***Riham A. ElToukhy,*** *MD,* ***Adnan M. Usmani,*** *MSc,* ***Waseem M. Hajjar,*** *MD, FRCS Ed.*

ABSTRACT

الأهداف: تقييم معرفة (Google Bard) في الأرقام والمسح الضوئي وتحديد الصور وتفسيراتها في التعليم الطبي وعلوم الرعاية الصحية من خلال أداء نوع الامتحان العملي الموضوعي (OSPE).

المنهجية: تم إنشاء بنك الأسئلة من نوع OSPE باستخدام مجموعة من أرقام العلوم الطبية وعمليات المسح والصور. للتقييم، تم اختيار ٦٠ شكلاً ومسحًا ضوئيًا وصورة وإدخالها في المنطقة المحددة في Google Bard لتقييم مستوى المعرفة. قمنا تحديد النتيجة على درجة من ٠ إلى ١، حيث يمثل الصفر الإجابة الخاطئة والواحد يوضح الإجابة الصحيحة.

النتائج: العلامات التي حصل عليها جوجل بارد في هياكل الدماغ والصور المورفولوجية والإشعاعية 7/10 (70%)؛ الهياكل العظمية، الصور الإشعاعية 9/10 (90%)؛ بنية الكبد والصور المورفولوجية والمرضية 4/10 (40%)؛ بنية الكلى والصور المورفولوجية 2/7 (28.57%)؛ الصور الإشعاعية العصبية 4/7 (57.14%)؛ والغدد الصماء بما في ذلك الغدة الدرقية والبنكرياس والثدي والصور المورفولوجية والإشعاعية 8/16 (50%). كان إجمالي العلامات الإجمالية التي حصل عليها Google Bard في مختلف أرقام OSPE وعمليات المسح الضوئي وأسئلة تعريف الصورة 34/60 (56.7%).

الخلاصة: سجل Google Bard درجة مرضية في التعرف على الصور المورفولوجية والنسيجية والإشعاعية وتفسيراتها. تشير النتائج إلى أنه في المستقبل، قد يساعد Google Bard طلاب الطب وأعضاء هيئة التدريس في التعليم الطبي والأطباء في أماكن الرعاية الصحية نظرًا لميزته بالمعرفة في سياق التعليم الطبي والعلوم الصحية.

Objectives: To evaluate the role of artificial intelligence (Google Bard) in figures, scans, and image identifications and interpretations in medical education and healthcare sciences through an Objective Structured Practical Examination (OSPE) type of performance.

Methods: The OSPE type of question bank was created with a pool of medical sciences figures, scans, and images. For assessment, 60 figures, scans and images were selected and entered into the given area of the Google Bard to evaluate the knowledge level.

Results: The marks obtained by Google Bard in brain structures, morphological and radiological images 7/10 (70%); bone structures, radiological images 9/10 (90%); liver structure and morphological, pathological images 4/10 (40%); kidneys structure and morphological images 2/7 (28.57%); neuro-radiological images 4/7 (57.14%); and endocrine glands including the thyroid, pancreas, breast morphological and radiological images 8/16 (50%). The overall total marks obtained by Google Bard in various OSPE figures, scans, and image identification questions were 34/60 (56.7%).

Conclusion: Google Bard scored satisfactorily in morphological, histopathological, and radiological image identifications and their interpretations. Google Bard may assist medical students, faculty in medical education and physicians in healthcare settings.

Keywords: Google Bard, diagnostic role, knowledge, image identifications

Saudi Med J 2024; Vol. 45 (5): 531-536
doi: 10.15537/smj.2024.45.5.20240141

From the Department of Physiology, College of Medicine, King Saud University (S. Meo); from the College of Medicine, Alfaisal University (AbuKhalaf, M.Z. Meo, M.O. Meo); from the Department of Science, College of Science (Ayub); from the Department of Family Medicine (ElToukhy); from the Department of Medicine (Usmani); from the Department of Surgery (Hajjar), College of Medicine, King Saud University. Riyadh, Kingdom of Saudi Arabia.

Received 18th February 2024. Accepted 21st April 2024.

Address correspondence and reprint request to: Dr. Sultan Ayoub Meo, Department of Physiology, College of Medicine, King Saud University, Riyadh, Kingdom of Saudi Arabia. Email: smeo@ksu.edu.sa
ORCID ID: https://orcid.org/0000-0001-9820-1852

Medical education is a dynamic field that demands a robust assessment process to ensure that medical students acquire essential knowledge and skills for their clinical practices. Medical education examinations depend on various assessment tools including multiple-choice questions (MCQ), Objective Structured Clinical Examination (OSCE), and Objective Structured Practical Examination (OSPE) while assessing knowledge and clinical skills in various disciplines.[1] The OSPE is a well-acknowledged assessment tool that has garnered global popularity among scholars. It supports the students in problem-solving practical practices. These types of assessment tools are incredibly important for the intellectual assessment levels of the learners.[1]

The OSPE is a highly promising assessment instrument to evaluate the professional performance of pupils in various academic settings across the world.[2] The OSPE is derived from the OSCE and used as an estimation tool both at undergraduate and postgraduate medical examination levels. Worldwide, many medical universities have adopted the OSPE for assessing the students' performance for various practical and clinical examinations.[3,4]

The OSPE is a highly structured and standardized assessment tool, which offers a holistic evaluation approach with practical skills with a diverse range of clinical competencies.[4] It enables the assessment of knowledge and skills in morphological, pathological, and radiological features identification, procedural proficiency, and problem-solving abilities. These comprehensive approaches ensure that medical students have abilities in their basic and clinical sciences abilities.[5]

The objectivity and reliability of OSPE provide a standard approach and construct validity which makes it a superior alternative tool to traditional assessment methods. OSPE contributes significantly to the quality of medical education and patient care benefits.[6,7]

At the beginning of the year 2023, the artificial intelligence (AI), Google Bard started acquiring fame among scholars, faculty, and researchers. The AI tools, Bard can help in preparing assorted articles, summarize the allied evidence, and provide ideas on any academic or professional assignments.[8] However, it poses threats to the traditional framework of medical education and research. These tools may develop new options for cheating on online examinations and minimize the critical thinking approach among students.[9]

Disclosure. Authors have no conflict of interests, and the work was not supported or funded by any drug company.

The assimilation of artificial intelligence tools in medical education and healthcare sciences has revolutionized various aspects of academia, research, and clinical settings in healthcare systems.[10] However, various viewpoints exist about knowledge, appropriateness, and allied services. The literature is lacking to investigate the Google Bard knowledge in figures, scans, and image identifications and interpretations. Therefore, this study evaluates the OSPE type of knowledge of Google Bard in figures, scans, and image identifications and interpretations in medical education and healthcare settings through Objective Structured Practical Examination (OSPE) pattern-based performance. This study may postulate the understanding and benefits to students, faculty and policymakers in medical education, and clinical settings in health sciences.

Methods. This cross-sectional study was performed in the Department of Physiology, College of Medicine, King Saud University, Riyadh, Saudi Arabia between September and October 2023. The study did not directly include any humans or animals, hence ethical approval was not obligatory.

Objective Structured Practical Examination (OSPE) bank. The 2 members of the research team prepared the contents-wise OSPE questions bank, figures, scans, and images from various textbooks, web sources, medical journals, and examination pools. The OSPE figures, scans, and images were precisely reviewed by another team member, and it was made sure that the OSPE questions were related to the contents. Each OSPE question was established on the figure, scan, or image identifications, with morphological, pathophysiological, and diagnostic characteristics. The OSPE questions were amended where required without changing the exact identification and characteristics of the OSPE questions and answers. There was no labelling on the OSPE questions, and the OSPE questions language was easy to understand (**Figure 1**). The research team rechecked the OSPE questions for any mistakes, unclear questions, or discrepancies. It was also checked that the figures, scans, or images were well-constructed and clear without any indications or hints. A pilot test was piloted on Google Bard with figures, scans, and images to examine the technicality of the selected OSPE questions. Once the research team was satisfied with the OSPE questions and their quality, all figures, scans, and images were compiled for the OSPE.

Selection of OSPE questions. A total of 60 OSPE questions based on morphological, histopathological, and radiological images from various disciplines of medical sciences were selected from the pool of OSPE

questions. These questions were based on various figures, scans, and images in medical sciences without any specific labelling that may demonstrate their identity. The OSPE questions were based on morphological, histological, histopathological, and or radiological figures, scans, and images. The OSPE questions were prepared as per medical education standards. For example, Identify the given object, Which body organ can be examined by this object?, Identify the given scan and provide the most appropriate diagnosis?, Identify the given image and the most probable diagnosis. (Figures 1&2). The OSPE questions were marked based on the number of wrong or accurate responses.

While preparing the OSPE questions, the questions which were not allied to the given subject areas or ambiguous questions, repetition of the questions, and biased questions were deleted from the OSPE pool questions. Moreover, the OSPE questions were excluded from the pool that had no visual components like figures, scans, or images.

Entry of OSPE questions into the Google Bard. The OSPE questions, with figures, scans, or images were manually entered, one at a time, and a new session was generated for each entry to minimize the chances of memory memorizing prejudice (Figure 1). The assignment was given to Bard to evaluate the knowledge and interpreting skill of Bard in each OSPE question. The OSPE figures, scans, and images were entered into Google Bard given space, and responses were recorded. The "regenerate response," was not used instead we considered the original response as a final one. Scoring was done on a scale of 0 to 1, where 0 represented an erroneous answer and one represented a correct answer.

Statistical analysis. A variety of descriptive statistical tests were used to assess the data, which were presented as numbers (n) and percentages (%).

Results. The knowledge of Google Bard was assessed, and its ability to respond to individual OSPE questions in medical sciences was explored. The OSPE questions were based on figures, scans and image identifications and interpretations which covered distinct areas in health sciences. The number of OSPE questions from brain structures, morphological and radiological images were 10, bone structures, fractures with 3-D images (10), liver structure and morphological images (10), kidneys structure and morphological, histopathological images (7), interventional neuro-radiological images (7), and endocrine glands including thyroid, pancreas, breast images.[16] The total number of OSPE questions was 60 (Table 1).

Tables 1&2 and Figure 2 demonstrated that out of 60 OSPE questions in various health sciences disciplines, Google Bard attempted all the OSPE questions. The marks obtained by Google Bard in brain structures, morphological and radiological images were 7/10 (70%); bone structures, radiological images 9/10 (90%); liver

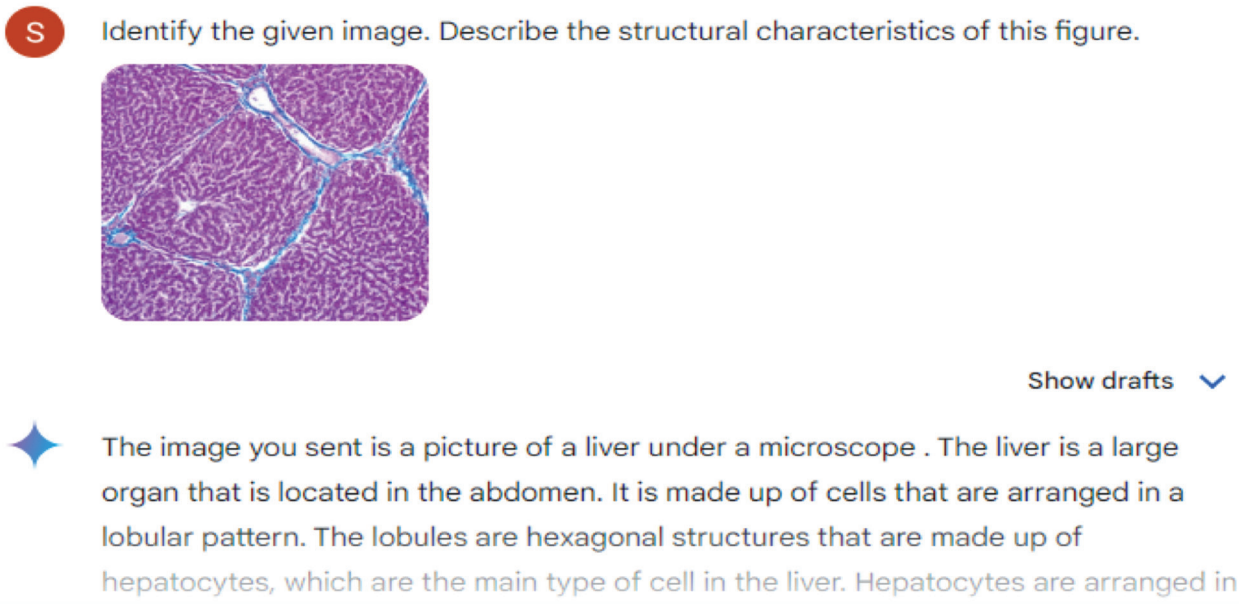

Figure 1 - Objective Structured Practical Examination type of question used to evaluate Google Bard's knowledge in image identifications and interpretations.

Table 1 - Distribution of OSPE questions based on figures, scans, and images of different body organs.

Figures and Images of body organs	Number of figures and images
Brain: Structures, morphological and radiological images	10
Bone: Structure, fractures with 3-D images	10
Liver: Structure and morphological images	10
Kidney: Structure and morphological images	7
Interventional Neuro =-radiological images	7
Endocrine glands: Thyroid, pancreas, breast images	16
Total	60

Table 2 - Marks obtained by Google Bard in OSPE questions on various figure scans and image identification and interpretation.

Figures, scans, and images	Number of OSPE questions and Marks obtained
Brain: Structures and morphological and radiological images	7/10 (70%)
Bone: Structures, and radiological images	9/10 (90%)
Liver: Structure and morphological images	4/10 (40%)
Kidneys: Structure and morphological images	2/7 (28.6%)
Interventional neuro-radiological images	4/7 (57.1%)
Endocrine glands: Thyroid, pancreas, breast, morphological and radiological images	8/16 (50.0%)
Total marks obtained	34/60 (56.7%)

OSPE: Objective Structures Practical Examination

structure and morphological, histo-pathological images 4/10 (40%); kidneys structure and morphological images 2/7 (28.6%); interventional neuro-radiological images 4/7 (57.1%); and endocrine glands including thyroid, pancreas, breast morphological and radiological images 8/16 (50%). The overall total marks obtained by Google Bard in various OSPE figures, scans and image identification questions were 34/60 (56.7%) (Table 2, Figure 1).

Discussion. In recent years, there has been rapid advancement in artificial intelligence (AI). Its tools are transforming the landscape of academia, medical education, and scientific research.[11] Artificial intelligence has made significant strides in revolutionizing medical sciences and healthcare settings. It has been proven to an instrumental in disease diagnosis, treatment protocol development, medicine, and healthcare analytics. This technology holds immense promise for improving patient care outcomes and enhancing the medical education and healthcare delivery system. The AI tools including Google Bard achieved enormous interest from the public, academicians, students, and research scholars. Google Bard rapidly responds with questions, identifications, and interpretations across the OSPE figures, scans, and images. It is a valuable tool to augment methodical information with evidence-based explanations. The present study investigated the Google Bard scientific knowledge in medical sciences through OSPE settings. The results reveal that Google Bard obtained a 56.7% score in basic and clinical sciences allied figures, scans and image identifications and interpretations through the OSPE.

The literature highlights the assessment knowledge level of AI including Chat GPT and Google Bard in MCQs type of examination, but no study is available in OSPE types of examination settings based on the figures, scans, and images with morphological, histopathological, and radiological identifications and interpretations.[12-18] This study highlights the role of Google Bard in figures, scans, and image identifications and interpretations to be added to the medical and healthcare science literature and highlights the knowledge and role of Google Bard in figures, scans, and image identifications and interpretations. It will support medical educators, students, faculty members, and physicians while providing medical education and healthcare services through online lectures, practical, training sessions and treating patients through telemedicine.

Over the last 3 decades, the approaches of evaluation in medical education have been markedly modified. The assessment tools are transformed from pen-and-paper tests of expertise toward a more complex system of assessment. The OSPE-based judgement of proficiency has become extensive in the field of medical education and healthcare sciences. The OSPE offers a reasonable objectivity-based assessment of the applicant's practical skills and capabilities. The OSPE is a trustworthy structured approach to knowledge and skill-based assessment of the applicant's performance based on laboratory-based practical stations.[19,20]

Fijačko et al[21] highlighted that Google Bard can correctly interpret the ECG images with a score of (13/27; 48.2%). The authors also mentioned that the primary work serves as proof of concept, and ECG explanation might be assisted by more AI tools shortly.[21] Hamanaka et al[22] investigated the role of AI in the diagnosis of lung cancer. The authors found no significant difference between the histological reports detected by the AI or the physicians. Moreover, it was also stated that in future there may be a chance that AI may substitute humans. Sarangi et al[13] assessed

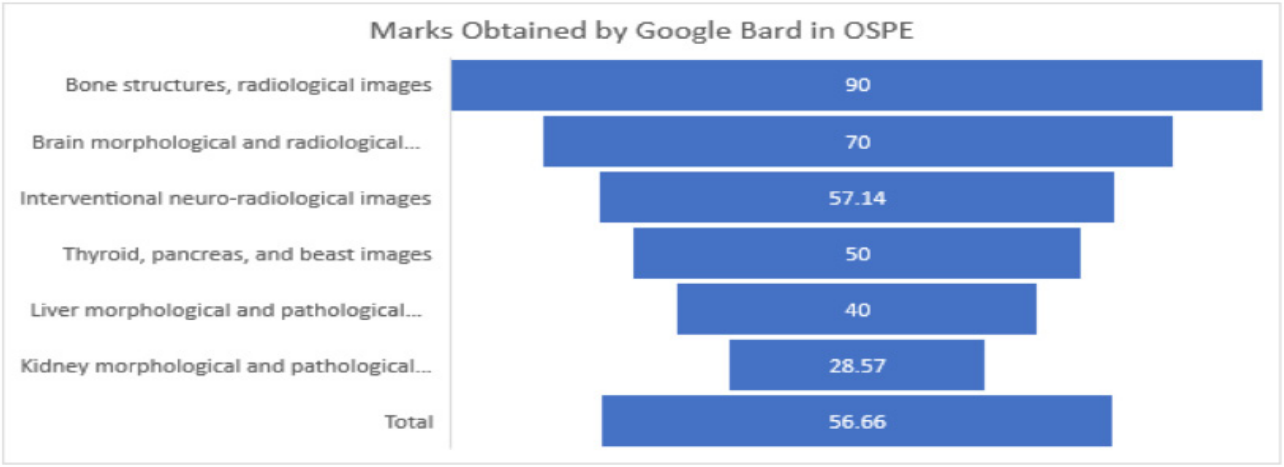

Figure 2 - Marks obtained by Google Bard in Objective Structured Practical Examination (OSPE) in figures and image identifications.

the ChatGPT, Google Bard, and Bing performance in answering the radiological case pieces by assessing the answers to those provided by radiology residents. The two residents (63.33 and 57.5%) outperformed the AI models: Bard (44.17%), Bing (53.33%), and ChatGPT (45%). It was found that the resident showed more correct answers than the AI. There is a great need for such studies from various corners of the globe to highlight the role of AI-based tools in medical education and healthcare settings before implementing these tools in medical sciences. This is the fact that medical sciences are extremely sensitive especially while providing diagnostic and patient management healthcare services.

In the present study, the OSPE type of pattern questions were given to the Google Bard with a wide range of contents within the medical sciences figures, scans and image identifications and interpretations with a thorough judgment. The grades achieved by Google Bard reveal that it has borderline satisfactory analytical thinking capabilities. The OSPE can assess reasonable expertise in complicated settings while analyzing the information, and applying rational, reasoning, and decisions while identifying and interpreting figures, scans, and images. It is crucial to remember that, even though OSPE is an effective assessment instrument, it might not fully capture all the information required for a thorough knowledge of the practical applications of OSPE or performance-based appraisals. Future research may be required to evaluate the role of Google Bard in the assessment of practical, clinical skills and knowledge, as we did not compare the practical basis skill abilities of Google Bard in real world settings.

Google Bard identifies numerous theme contents with broad identifying and reasoning abilities of the diverse subject contents in medical sciences. Google Bard may be a useful source of knowledge for the medical community; however, applied applications must be cautiously measured since the scores on examination questions are limited and there is a need for further updates in the technology tools. This is also a fact that Google Bard made noteworthy progress not just in understanding and generating text but establishing applications beyond its original design. The noteworthy expansion in the realm of the Google Bard is image identifications and interpretations. It may tackle the complex task of identifying, interpreting, and describing medical images. In educational and healthcare settings, Google Bard's image identification and interpretation can enhance the learning experiences by providing detailed descriptions of figures, scans and images used in textbooks, presentations, or online courses and healthcare settings. This can benefit students and faculty in diverse learning styles and cultural contexts. It can assist healthcare professionals by providing detailed descriptions of medical images, diagnosis, and enhancing health care services between medical practitioners and patients. It must be noted that Google Bard's expedition into figures, scans and image identification is promising, but it comes with some set of challenges such as managing large datasets, ethical use, and accuracy are areas which require great attention.

Study strengths and limitations. This is a novel study that evaluated the level of knowledge and understanding of Google Bard in morphological, histopathological, and radiological figures, scans, and image identifications and interpretations in medical sciences and healthcare settings. In the future, such studies on figures, scans

and image identifications and interpretations should be performed from various corners of the globe to reach better conclusions. However, shortly after progress in technology, the Google Bard could be further developed and therefore results may be further improved. Rigorously evaluating these AI-based tools' utility in the medical sciences and healthcare settings is highly essential across the knowledge spectrum and must be used in both controlled and real-world assessments. The limitation of this study is that we were unable to compare it with other AI tools and perform in real-world settings as we routinely perform in the labs.

In conclusion, the Google Bard obtained a borderline satisfactory score in the OSPE questions based on the figures, scans and image identifications and interpretations in medical education and healthcare settings. The findings based on the Google Bard knowledge assessment demonstrated a certain level of understanding and explanation of figures, scans and image identifications and interpretations. The study findings advocate that in future, Google Bard may assist the medical faculty in medical education and physicians in healthcare settings since it has knowledge and potential in the framework of medical education and healthcare sciences.

Acknowledgment. *The authors extend their appreciation to the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia for funding this research work through project no (IFKSUOR3-4-7). The authors gratefully acknowledge Sofia Fields Author Services (www.sofiafields.com) for the English language editing.*

References

1. Meo SA. Evaluating learning among undergraduate medical students in schools with traditional and problem-based curricula. *Adv Physiol Educ* 2013; 37: 249-253.
2. Lakum NR, Maru AM, Algotar CG, Makwana HH. Objective Structured Practical Examination (OSPE) as a tool for the formative assessment of practical skills in the subject of physiology: a comparative cross-sectional analysis. *Cureus* 2023; 15: e46104.
3. Harden RM. What is an OSCE? *Med Teach* 1988; 10: 19-22.
4. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979; 13: 41–54.
5. Kumar N, Rahman E. OSPE for the non-surgical aesthetic training programme. *Clin Teach 2022*; 19: e13515.
6. Prasad HLK, Prasad HVK, Sajitha K, Bhat S, Shetty KJ. Comparison of Objective Structured Practical Examination (OSPE) versus Conventional Pathology Practical Examination Methods among the second-year medical students-a cross-sectional study. *Med Sci Educ* 2020; 30: 1131-1135.
7. Hultgren C, Lindkvist A, Curbo S, Heverin M. Students' performance of and perspective on an objective structured practical examination for the assessment of preclinical and practical skills in biomedical laboratory science students in Sweden: a 5-year longitudinal study. *J Educ Eval Health Prof* 2023; 20: 13.
8. Hutson M. Could AI help you to write your next paper? *Nature* 2022; 611: 192–193.
9. Rahman M.M. Watanobe Y. ChatGPT for education and research: opportunities, threats, and strategies. *Appl Sci* 2023, 13: 5783.
10. Narayanan S, Ramakrishnan R, Durairaj E, Das A. Artificial intelligence revolutionizing the field of medical education. *Cureus* 2023; 15: e49604.
11. Ray PP. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *IOTCPS* 2023; 3: 121-154.
12. Meo SA, Al-Masri AA, Alotaibi M, Meo MZS, Meo MOS. ChatGPT knowledge evaluation in basic and clinical medical sciences: multiple choice question examination-based performance. *Healthcare (Basel)* 2023; 11: 2046.
13. Sarangi PK, Narayan RK, Mohakud S, Vats A, Sahani D, Mondal H. Assessing the capability of ChatGPT, Google Bard, and Microsoft Bing in solving radiology case vignettes. *Indian J Radiol Imaging* 2023; 34: 276-282.
14. Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the performance of ChatGPT in ophthalmology: an analysis of its successes and shortcomings. *Ophthalmol Sci* 2023; 3: 100324.
15. Friederichs H, Friederichs WJ, März M. ChatGPT in medical school: how successful is AI in progress testing? *Med Educ Online* 2023; 28: 2220920.
16. Patil NS, Huang RS, van der Pol CB, Larocque N. Comparative Performance of ChatGPT and Bard in a Text-Based Radiology Knowledge Assessment. *Can Assoc Radiol J* 2023; 8465371231193716.
17. Fowler T, Pullen S, Birkett L. Performance of ChatGPT and Bard on the official part 1 FRCOphth practice questions. *Br J Ophthalmol* 2023: bjo-2023-324091.
18. Meo SA, Al-Khlaiwi T, AbuKhalaf AA, Meo AS, Klonoff DC. The Scientific Knowledge of Bard and ChatGPT in endocrinology, diabetes, and diabetes technology: multiple-choice questions examination-based performance. *J Diabetes Sci Technol* 2023: 19322968231203987.
19. Abraham RR, Raghavendra R, Surekha K, Asha K. A trial of the objective structured practical examination in physiology at Melaka Manipal Medical College, India. *Adv Physiol Educ* 2009; 33: 21-23.
20. Singh G, Singh R. Practical skills evaluation of undergraduate medical students by an objective structured practical examination (OSPE): An effective tool for formative assessment in microbiology. *MGM J Med Sci (Online)* 2020; 7: 74-78.
21. Fijačko N, Prosen G, Abella BS, Metličar Š, Štiglic G. Can novel multimodal chatbots such as Bing Chat Enterprise, ChatGPT-4 Pro, and Google Bard correctly interpret electrocardiogram images? *Resuscitation* 2023; 193: 110009.
22. Hamanaka R, Oda M. Can artificial intelligence replace humans for detecting lung tumors on radiographs? An examination of resected malignant lung tumors. *J Pers Med* 2024 31: 14: 164.