RESEARCH

Open Access

Enhancing puncture skills training with generative AI and digital technologies: a parallel cohort study

Zhe Ji[1], Yuliang Jiang[1], Haitao Sun[1], Bin Qiu[1], Yi Chen[1], Mao Li[1], Jinghong Fan[1] and Junjie Wang[1*]

Abstract

Background Traditional puncture skills training for refresher doctors faces limitations in effectiveness and efficiency. This study explored the application of generative AI (ChatGPT), templates, and digital imaging to enhance puncture skills training.

Methods 90 refresher doctors were enrolled sequentially into 3 groups: traditional training; template and digital imaging training; and ChatGPT, template and digital imaging training. Outcomes included theoretical knowledge, technical skills, and trainee satisfaction measured at baseline, post-training, and 3-month follow-up.

Results The ChatGPT group increased theoretical knowledge scores by 17–21% over traditional training at post-training (81.6 ± 4.56 vs. 69.6 ± 4.58, $p < 0.001$) and follow-up (86.5 ± 4.08 vs. 71.3 ± 4.83, $p < 0.001$). It also outperformed template training by 4–5% at post-training (81.6 ± 4.56 vs. 78.5 ± 4.65, $p = 0.032$) and follow-up (86.5 ± 4.08 vs. 82.7 ± 4.68, $p = 0.004$). For technical skills, the ChatGPT (4.0 ± 0.32) and template (4.0 ± 0.18) groups showed similar scores at post-training, outperforming traditional training (3.6 ± 0.50) by 11% ($p < 0.001$). At follow-up, ChatGPT (4.0 ± 0.18) and template (4.0 ± 0.32) still exceeded traditional training (3.8 ± 0.43) by 5% ($p = 0.071$, $p = 0.026$). Learning curve analysis revealed fastest knowledge (slope 13.02) and skill (slope 0.62) acquisition for ChatGPT group over template (slope 11.28, 0.38) and traditional (slope 5.17, 0.53). ChatGPT responses showed 100% relevance, 50% completeness, 60% accuracy, with 15.9 s response time. For training satisfaction, ChatGPT group had highest scores (4.2 ± 0.73), over template (3.8 ± 0.68) and traditional groups (2.6 ± 0.94) ($p < 0.01$).

Conclusion Integrating AI, templates and digital imaging significantly improved puncture knowledge and skills over traditional training. Combining technological innovations and AI shows promise for streamlining complex medical competency mastery.

Keywords Puncture, Training, Refresher doctor, Template, Digital technology, ChatGPT

*Correspondence:
Junjie Wang
junjiewang_edu@sina.cn
[1]Department of Radiation Oncology, Peking University Third Hospital,
49th North Garden Road, Haidian District, Beijing 100191, P.R. China

Introduction

Puncture procedures are fundamental skills for medical practitioners, especially in fields such as radiology, anesthesiology, and oncology. The precise execution of puncture techniques is crucial for patient safety and the success of further medical interventions. However, traditional teaching methods have some limitations in training complex invasive procedures like puncture. Students often find it difficult to understand three-dimensional (3D) anatomical structures from 2D images and lack opportunities to practice on real patients under supervision [1].

In recent years, emerging technologies like artificial intelligence (AI), 3D printing, and digital imaging have been increasingly applied in medical education to enhance training efficiency and outcomes [2–4]. AI can provide intelligent training suggestions and improve problem-solving efficiency. 3D printing offers realistic models and customized templates. Digital imaging improves visualization and measurements. The integration of these technologies has shown great potential in puncture training.

However, the application of AI in this field is still in the early stage. Chat generative pre-trained transformer (ChatGPT) as a conversational AI has shown capabilities in medical knowledge delivery [5]. But few studies have evaluated its roles in assisting puncture skills training.

Therefore, this study aims to explore the application of ChatGPT combined with templates and digital imaging in puncture training for refresher doctors. We hypothesize that the integration of these technologies can improve learning outcomes and training quality compared to traditional methods. The findings will provide insights and evidence for integrating innovations in medical education.

Materials and methods

Study design and participants

This parallel observational study was conducted at our teaching hospital from January 2023 to September 2023. A total of 90 refresher doctors were recruited on a voluntary basis from the pool of doctors enrolled in a standardized 12-week puncture skills training program. The training program primarily focused on CT-guided puncture procedures. Participants were sequentially grouped into three distinct groups (30 per group) in order of registration.

Interventions

Traditional training group (group 1, n = 30)

Participants in this group received conventional pedagogical training which included:

24 h of interactive classroom lectures utilizing PowerPoint presentations covering anatomy, case demonstrations, puncture techniques, radiation safety, and complications management.

Reading assignments from two standard textbooks, 'Interventional Radiology Procedures in Biopsy and Drainage' and 'CT- and MR-Guided Interventions in Radiology', and review of the 20 latest pieces of literature on CT-guided puncture techniques.

Observation of 8 live CT-guided biopsy procedures performed by experienced faculty members, including those of the head and neck, lung, pelvis, retroperitoneum, and bone.

16 h of supervised hands-on practice on artificial human torso models and 16 h of supervised procedures on real patients. During the practice sessions on the artificial models, trainees focused on developing needling techniques and building muscle memory. The supervised procedures on real patients, under the guidance of experienced instructors, provided the trainees with an additional 16 h of clinical experience by targeting various anatomical regions, including the thyroid, lungs, liver, lymph nodes, and other relevant anatomical structures. This comprehensive approach, combining equal hours of practice on artificial models and real patients, ensured that the trainees gained both technical proficiency and practical experience, while prioritizing patient safety throughout the training process.

Template and digital imaging group (group 2, n = 30)

Employment of Patient-specific 3D-printed Puncture Templates:

Utilizing advanced technology, we employed universally applicable coplanar templates capable of mass production for puncture practices. These were not individualized 3D printed models but standardized tools, featuring guiding holes evenly spaced at 2.5 mm intervals. The templates are designed to be affixed to a versatile navigation frame, rather than the frame being attached to the template. This navigation frame is equipped with multiple articulating joints, enabling it to be adjusted to any angle for precise alignment. This combination of a universally applicable template and an adjustable navigation frame enables precise guidance of the puncture needle towards the imaging target. This system offers significant precision in needle trajectory alignment. This approach bridges the gap between theoretical knowledge and practical application, enhancing the accuracy of puncture practices with a tangible guide for needle insertion and alignment. It also facilitates a deeper understanding and mastery of puncture techniques among trainees, promoting accurate and efficient learning (Fig. 1).

Access to Advanced Imaging Software:

Participants were provided access to the radioactive seed brachytherapy treatment planning system (BTPS), developed by Beihang University. This system
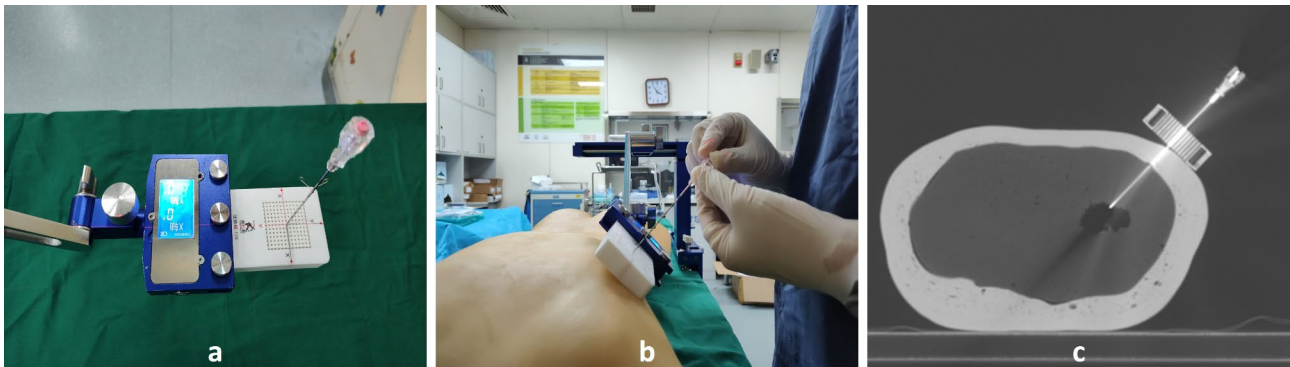
Fig. 1 Template-Assisted Puncture Training: (**a**) 3D-printed puncture template and fixed bracket. (**b**) Trainees perform punctures on the mold with the assistance of a template. (**c**) CT images taken during puncture operation
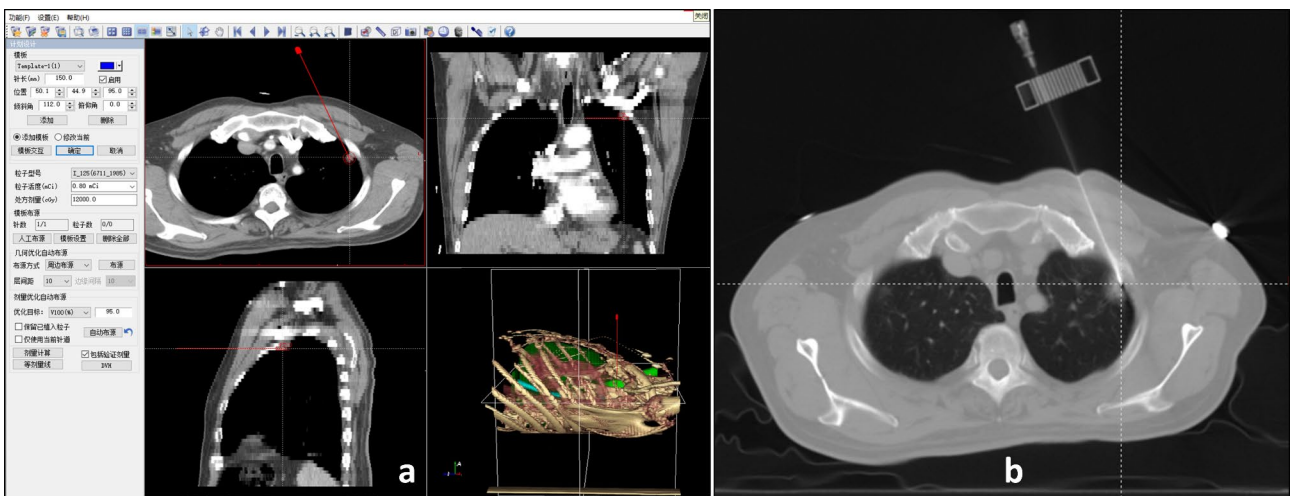

Fig. 2 Puncture Biopsy under the Guidance of BTPS: (**a**) Design of puncture pathways and digital reconstruction within the BTPS. (**b**) The actual puncture status aligns with the preoperative plan

is specifically tailored for radioactive seed implantation and multi-needle puncture, featuring functionalities such as multimodal image fusion, three-dimensional image reconstruction, and multiple needle track planning. The system allows for the integration and comprehensive visualization of multiplanar CT/MRI data, facilitating precise procedural guidance. Additionally, the system provides critical quantitative measurements like needle depth, angle, and proximity to vital structures, playing a key role in helping trainees understand and refine their puncture techniques. Integrating the Radioactive Seed Brachytherapy Treatment Planning System into the training program leverages digital technology to offer a more interactive, feedback-rich, and data-driven learning experience (Fig. 2).

ChatGPT, template and digital imaging group (group 3, n = 30)

This group received a comprehensive training approach by incorporating ChatGPT (Open AI) alongside traditional and digital augmented training:

Personalized Teaching Content Generation:

Utilizing ChatGPT, a tailored teaching content was generated to bridge the individual knowledge gaps of each trainee. These gaps were meticulously identified through a comprehensive 15-question quiz, encompassing critical areas such as anatomy (focusing on organ localization and understanding surrounding structures) (prompt cue words: "Identify," "Describe," "Compare," "Locate", etc.), puncture techniques (including selection of appropriate needle size, approach strategies, insertion depth, and angle adjustment) (prompt cue words: "Select," "Determine," "Approach," "Insert," "Adjust angle", etc.), along with patient communication and care (encompassing obtaining consent, continuous monitoring during procedures, and post-procedure care

protocols) (prompt cue words: "Obtain consent," "Monitor," "Post-procedure care," "Communicate", etc.).

Real-time Conversational Guidance:

During hands-on practice sessions, ChatGPT provided real-time conversational guidance, aiding in various aspects like providing orientation cues to accurately locate target sites, suggesting technique and posture adjustments to enhance procedural accuracy and comfort, troubleshooting procedural issues as they arose, and answering inquiries regarding organ relationships, needle trajectory optimization, radiation protection measures, and patient interaction skills. This real-time feedback mechanism fostered a conducive learning environment for trainees to rectify mistakes instantaneously and internalize correct techniques (The prompts include: 'Provide guidance on locating the target site for a [specific organ] biopsy,' 'Suggest posture adjustments for optimal ergonomics during a [specific procedure],' 'What should I do if I encounter resistance during needle insertion?,' 'Explain the relationship between [specific organ] and adjacent structures to avoid during puncture,' and 'How can I effectively communicate the steps of the procedure to the patient for reassurance?' among others.).

Common Trainee Questions Resolution:

An extensive repository of over 100 common trainee questions was curated by experienced faculty members, anticipating potential queries and challenges faced during training. ChatGPT was equipped to provide precise answers and clarifications to these questions, encompassing a wide spectrum of topics such as procedural steps, anatomy identification, correct equipment settings, radiation safety precautions, handling complications and emergency management, effective communication with patients, error correction techniques, and sharing expert tips and tricks for successful punctures.

ChatGPT Integration in Training Process:

ChatGPT was strategically integrated into the training process to enhance learning without compromising hands-on practice time. Trainees utilized ChatGPT for pre-practice preparation, reviewing relevant anatomy and procedural steps to reinforce theoretical knowledge. During practical sessions, a dedicated tablet with ChatGPT was available at each training station, allowing trainees to briefly consult for immediate clarifications or guidance with minimal interruption to their hands-on work. Post-practice, trainees engaged with ChatGPT to analyze their performance and identify areas for improvement, promoting self-directed learning. To maintain an optimal balance between AI assistance and practical training, ChatGPT interactions were limited to an average of 5 min per 30-minute practice session. This structured integration of ChatGPT complemented the existing use of templates and digital imaging, creating a comprehensive, technology-enhanced learning environment. The

digital repository of common trainee questions, curated by experienced faculty members, remained an invaluable resource, enabling trainees to access expert knowledge and solutions instantaneously.

Outcomes
Theoretical knowledge assessment
Theoretical knowledge was appraised through a series of three 60-item validated multiple-choice tests encompassing key areas including anatomy, puncture techniques, equipment usage, principles of puncture techniques, etc. (Appendix A). These tests were meticulously curated and validated by a panel of experienced faculty members to ensure content accuracy and relevance. The assessments were administered at three distinct intervals: before the initiation of the training program, immediately post-training, and at a 3-month follow-up. This longitudinal assessment approach aimed to evaluate the retention and application of theoretical knowledge over time.

Technical skills assessment
Prepare the necessary equipment and materials, such as puncture needles, human models or simulators, and imaging devices. Set up a simulated clinical environment with essential medical equipment and safety measures. Under the guidance of examiners or senior physicians, demonstrate the steps of the puncture procedure on human models. Use a streamlined checklist to assess the procedure, which includes key elements like patient identification and consent, equipment sterilization, patient positioning, aseptic technique, accurate needle guidance using imaging, technical execution of the puncture, post-procedure care, and proper documentation and communication (Appendix B). This checklist ensures a thorough evaluation of the operation's correctness, safety, and effectiveness.

Assessments were conducted at four critical junctures: baseline (pre-training), post each weekly training session, immediately post-training, and at a 3-month follow-up. All performances were meticulously video recorded to facilitate a blinded rating by the evaluating faculty members, thus minimizing bias. The checklists focused on the satisfactory completion of essential procedural steps, all rated on a nuanced 5-point scale.

Cumulative learning curves
Cumulative learning curves were delineated by plotting trainees' aggregate checklist scores against the number of practice sessions. These curves were meticulously analyzed by calculating the slope (representing the rate of learning), time to plateau (indicating when a learner reaches a steady state of performance), and goodness of fit (quantifying the accuracy of the curve fit), which are

Ji *et al. BMC Medical Education*   (2024) 24:1328

Page 5 of 9

essential metrics reflecting the efficacy and pace of skill acquisition.

Trainee feedback assessment

Post-training and 3-month follow-up questionnaires were administered to gauge trainees' self-reported training experiences. A 5-point Likert scale (1–5 points) was used for self-assessment of overall satisfaction, goal achievement and skill application across the 3 groups, where 1=very dissatisfied, 2=dissatisfied, 3=neutral, 4=satisfied, and 5=very satisfied (Appendix C). Additionally, free-text feedback was solicited to capture open-ended responses for a deeper insight into the training program's strengths and areas for improvement.

ChatGPT assessment

Subjective evaluations of ChatGPT responses are conducted by experienced instructors. These evaluations include relevance judgment (determining whether the answer is directly related to the question) and accuracy judgment (identifying any errors in the answer content). Completeness judgment (defined as fully addressing queries without further prompting) is also performed to evaluate whether ChatGPT's responses address all parts of the question.

Statistical analysis

Quantitative data were presented as mean±standard deviation (SD). Differences between groups were explored using One-Way analysis of variance (ANOVA) test. Qualitative data were analyzed using the Chi-square test to compare categorical variables. All statistical analyses were performed using statistical package for the social sciences (SPSS) version 25 software, with a p-value of <0.05 considered as statistically significant.

Results

Baseline characteristics of trainees

The gender distribution of the trainees was 81.1% male and 18.9% female. The average age was 36.5±4.63 years. The distribution of education levels was 42.2% Bachelor's, 36.7% Master's, and 21.1% Doctorate. The specialty distribution was 16.7% Interventional Medicine, 23.3% Radiology, 18.9% Nuclear Medicine, 20.0% Oncology, and 21.1% Radiotherapy. The title distribution was 32.2% Junior, 51.1% Intermediate, and 16.7% Senior. "The distribution of working experience was 47.8% for those with less than 5 years and 52.2% for those with 5 to 10 years. The baseline characteristics were balanced across the 3 groups with no significant differences, indicating comparability (Table 1).

Comparison of theoretical knowledge test scores

The trainees' theoretical knowledge scores at baseline, post-training, and 3-month follow-up were 60.5±5.02, 76.6±6.86, and 80.2±7.88, respectively. At baseline, there were no significant differences among the 3 groups ($p>0.05$). At post-training and 3-month follow-up, Group 3 was better than Group 2, and Group 2 was better than Group 1, with significant differences between groups ($p<0.05$) (Table 2).

Comparison of skills test scores

The trainees' skill scores at baseline, post-training, and 3-month follow-up were 3.0±0.26, 3.9±0.42, and 3.9±0.34, respectively. At baseline, there were no significant score differences among groups ($p>0.05$). At post-training, Group 2 and 3 were better than Group 1 with significant differences ($p<0.05$), and there were no significant differences between Groups 2 and 3 ($p>0.05$). At 3-month follow-up, both Group 2 and Group 3 were superior to Group 1, with statistically significant differences between Group 1 and Group 2 ($p<0.05$). However, there were no statistically significant differences between Group 1 and Group 3, or between Group 2 and Group 3 ($p>0.05$) (Table 2).

Learning curve analysis

Among the 3 groups' learning curves (Fig. 3), Group 3 had the steepest slope. For theoretical knowledge, Group 3's learning curve slope was 13.02, higher than Group 1 (5.17) and Group 2 (11.28). For skills, Group 3's learning curve slope was 0.62, higher than Group 1 (0.53) and

Table 1 Baseline characteristics of trainees

Characteristics	Group 1 (n, %)	Group 2 (n, %)	Group 3 (n, %)	P
Gender				0.930
Male	24(80%)	25(83.3%)	24(80%)	
Female	6(20%)	5(16.7%)	6(20%)	
Age	36.9±4.16	35.7±4.65	37.1±5.06	0.442
Education				0.522
Bachelor	12(40%)	10(33.3%)	16 (53.3%)	
Master	11(36.7%)	14(46.7%)	8 (26.7%)	
Doctor	7(23.3%)	6(20%)	6 (20%)	
Specialty				0.500
Interventional medicine	4(13.3%)	6(20%)	5(16.7%)	
Radiology	8(26.7%)	7(23.3%)	6(20%)	
Nuclear medicine	3(10%)	6(20%)	8(26.7%)	
Oncology	10(33.3%)	4(13.3%)	4(13.3%)	
Radiotherapy	5(16.7%)	7(23.3%)	7(23.3%)	
Title				0.940
Junior	11(36.7%)	10(33.3%)	8(26.7%)	
Intermediate	14(46.7%)	15(50%)	17(56.7%)	
Senior	5(16.7%)	5(16.7%)	5(16.7%)	
Experience				0.956
<5 years	14(46.7%)	15(50%)	14(46.7%)	
5–10 years	16(53.3%)	15(50%)	16(53.3%)	

Table 2 Scores from the theoretical knowledge test and skills test

| Grouping	Pre-training		Post-training		3 Months After Training	
	Theory Score	Skill Score	Theory Score	Skill Score	Theory Score	Skill Score
Group 1	61.0±5.73	3.0±0.26	69.6±4.58	3.6±0.50	71.3±4.83	3.8±0.43
Group 2	60.1±4.35	3.0±0.26	78.5±4.65	4.0±0.18	82.7±4.68	4.0±0.32
Group 3	60.4±5.00	2.9±2.54	81.6±4.56	4.0±0.32	86.5±4.08	4.0±0.18
P	0.795	0.52	<0.001	<0.001	<0.001	<0.006
P (G1 vs. G2)	0.880	1	<0.001	<0.001	<0.001	0.026
P (G2 vs. G3)	0.991	0.688	0.032	1	0.004	0.695
P (G1 vs. G3)	0.971	0.688	<0.001	<0.001	<0.001	0.071
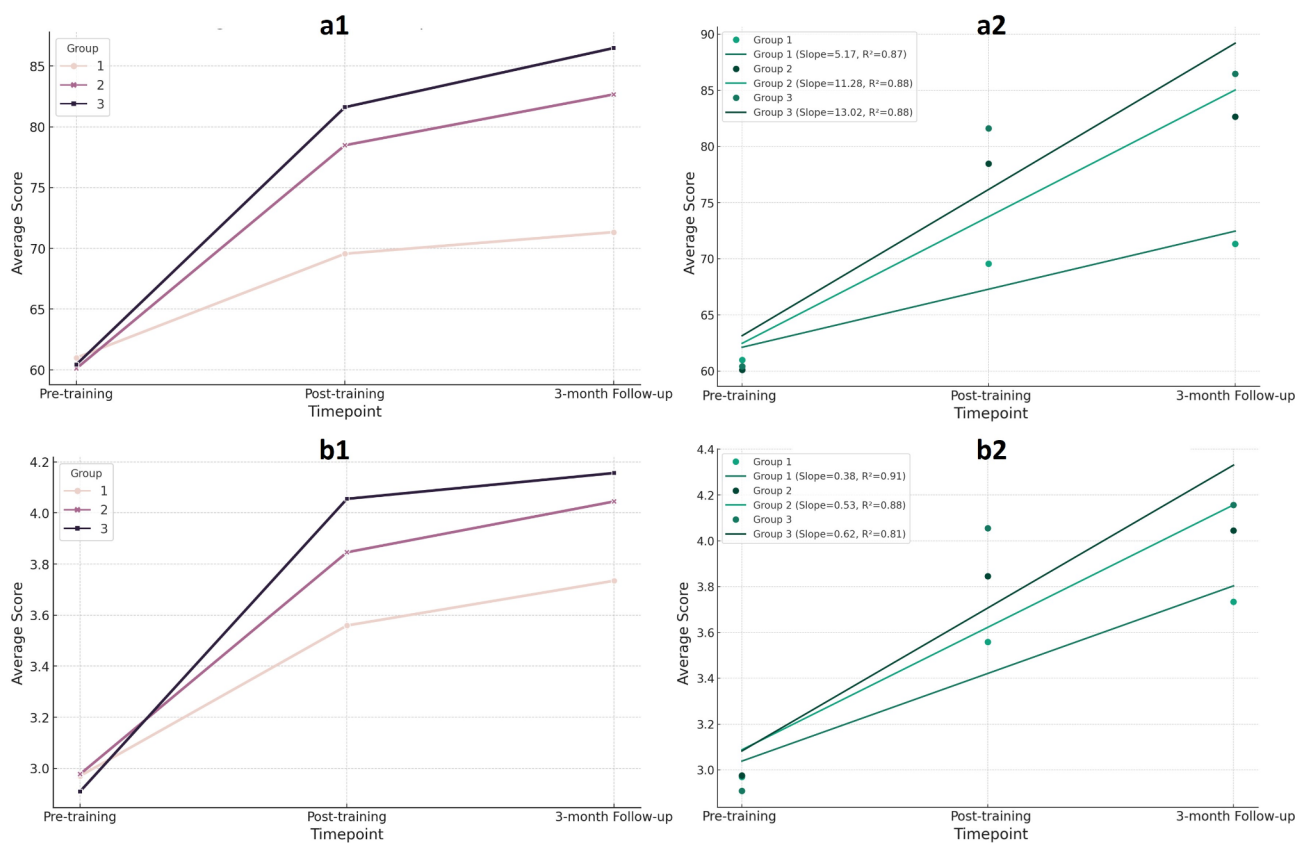
Fig. 3 Learning Curves of the Three Groups: a1. For theoretical knowledge, the learning curve of Group 3 was better than that of Group 2, and Group 2's was better than Group 1's. a2. For theoretical knowledge, the learning curve slope of Group 3 was higher than Group 2's, and Group 2's was higher than Group 1's. b1. For skills, the learning curve of Group 3 was better than that of Group 2, and Group 2's was better than Group 1's. b2. For skills, the learning curve slope of Group 3 was higher than Group 2's, and Group 2's was higher than Group 1's. (Group 1: Traditional Training Group; Group 2: Template and Digital Imaging Group; Group 3: ChatGPT, Template, and Digital Imaging Group)

Table 3 Evaluation of ChatGPT responses

Indicators	Value
Relevance	100%
Completeness	50%
Accuracy	60%
Response Time	15.9±2.51
Number of Interactions	3.0±0.79

Group 2 (0.38). All groups had good curve fitting (R^2: 0.81–0.91).

ChatGPT response analysis

ChatGPT responses achieved 100% relevance and 50% completeness. The average number of interactions per response was 3 times. The accuracy rate was 60%. The average response time was 15.9 s (Table 3).

Training satisfaction survey

The mean overall satisfaction score was 3.5±1.05. The mean goal achievement score was 3.7±1.02. The mean practical application score was 3.4±1.11. All scores showed that Group 3 performed better than Group 2, and Group 2 performed better than Group 1, with significant differences between groups ($p < 0.05$) (Table 4).

Discussion

The aim of this study was to explore the effects of integrating ChatGPT, 3D printing templates, and digital imaging technologies for training doctors' puncture skills. Results showed that all three training methods were effective, but compared to traditional training, the two digitally-assisted methods notably improved theoretical knowledge and technical skills in puncture. Specifically, the ChatGPT group increased theoretical scores by 17–21% over the traditional group and 4–5% over the template group. This may be attributed to ChatGPT's ability to identify learning difficulties in real-time and provide guidance, reducing cognitive load and facilitating active learning, as described in Lee's study [6]. The provision of tangible simulation training with templates and digital imaging can also enhance understanding and concretize abstract concepts [7, 8]. For skill scores, the ChatGPT and template groups performed similarly, outperforming the traditional group by 11% post-training and 5% at the 3-month follow-up. A. Haleem et al. [9] reviewed that digital assistance has emerged as an essential tool to achieve quality education and lead to skill enhancement. This is largely owing to instant feedback and procedural guidance from digital technologies, which helps correct mistakes and refine technical deviations [10], enabling more rapid and accurate skill acquisition.

The discrepancy between the increased theoretical scores and the unchanged skill scores in the Group 3 (template+digital+ChatGPT) compared to the Group 2 (template+digital) can be attributed to ChatGPT's primary enhancement of theoretical understanding rather than manual dexterity [11], as well as the inherently slower improvement rate of practical skills compared to theoretical knowledge. At the 3-month follow-up, the skill scores of Group 1 (Traditional) and 3 were also statistically similar. This implies that while ChatGPT bolsters short-term teaching, its effects on reinforcing long-term retention may be limited, warranting further investigation.

Analysis of the learning curves revealed the fastest knowledge and skill acquisition in the ChatGPT group, with curve slopes of 13.02 and 0.62, respectively. ChatGPT's real-time interactivity enabled a more effective and smoother pathway to skill proficiency, making it worthy of broader implementation. However, due to limited research and data, determining whether intelligent algorithms can optimize training and expedite mastery of complex skills requires more evidence from comparative studies across various learning contexts.

Our evaluation metrics for ChatGPT responses indicated a 100% relevance but only 50% completeness. The 60% accuracy observed is lower than the 76% reported in W. Choi et al.'s study [12], which is likely due to stricter grading criteria and limitations in the depth of medical knowledge. Despite these metrics, the ChatGPT group showed significant improvements in theoretical knowledge scores. This can be attributed to the interactive nature allowing immediate error correction, increased trainee engagement, and its use as a complementary tool alongside validated resource. The average response time of 15.9 s under our experimental conditions is promising, though more benchmarking is needed to compare it with existing visualization tools across various scenarios. To enhance ChatGPT's performance, we propose: fine-tuning on medical literature, developing effective prompting strategies, integrating with verified databases, implementing expert review, and establishing a continuous feedback loop. These findings underscore both the promise of AI-assisted medical education and the need for careful implementation and refinement. Future research should focus on optimizing these systems while maintaining the engaging elements that contribute to improved learning outcomes, as well as conducting more comprehensive comparisons with existing educational tools.

Additionally, we believe that Group 3's high satisfaction among trainees can be attributed to several factors: the novelty of AI integration, which increased engagement; ChatGPT's personalized responses that addressed individual learning needs; immediate feedback, which allowed for quick problem-solving; and the empowerment trainees felt from having instant access to information, which potentially boosted their confidence and sense of autonomy.

Table 4 Results of the satisfaction survey on training

Indicators	score			P	P (G1 vs. G2)	P (G2 vs. G3)	P (G1 vs. G3)
	Group 1	Group 2	Group 3				
Satisfaction	2.6±0.94	3.8±0.68	4.2±0.73	<0.001	<0.001	0.038	<0.001
Objective Achievement	2.7±0.79	3.8±0.73	4.5±0.57	<0.001	<0.001	<0.001	<0.001
Technical Application	2.4±0.77	3.5±0.78	4.4±0.72	<0.001	<0.001	<0.001	<0.001

This study is subject to several constraints, chiefly, these include: a small sample size possibly impacting difference detection; few measurements of learning trajectory prohibiting observations of skill plateauing; dependence on subjective appraisals of ChatGPT which await more objective evaluations; a focus solely on ChatGPT's capabilities without comparisons to other artificial intelligence systems or language models; and the voluntary nature of participant enrollment, which may have introduced selection bias, potentially affecting the generalizability of our results. The innovations of this work encompass: (1) Novel systematic consolidation of 3D navigational templates with digital imaging technologies to create multifaceted immersive simulation environments; (2) Deploying ChatGPT's personalized content generation and real-time guidance capabilities to actualize customized intelligent augmentation; (3) Constructing a comprehensive evaluation framework spanning theoretical tests, technical assessments, learning curves, and satisfaction, thereby validating training efficacy across multiple facets and elevating appraisal validity.

Conclusion

Digitally-enhanced modalities significantly improved the acquisition of puncture knowledge and skills. Combining advancements in both pedagogical and technological domains can synergistically catalyze progress in medical education. Incorporating generative AI and digital tools serves as an effective auxiliary for traditional simulation training to enable safer and more streamlined mastery of multifaceted competencies. Our evidence endorses expanded implementation of these emerging methodologies. We anticipate witnessing their extensive integration, which aims to sharpen wide-ranging medical capabilities.

Abbreviations

AI	Artificial Intelligence
ChatGPT	Chat Generative Pre-trained Transformer
3D	Three-dimensional
ANOVA	Analysis of Variance
SPSS	Statistical Package for the Social Sciences
SD	Standard Deviation
BTPS	Brachytherapy Treatment Planning System

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12909-024-06217-0.

Supplementary Material 1

Data availability
The data supporting the findings of this study are available from the corresponding author, J.W., upon reasonable request. Restrictions apply to the availability of these data, which were used under license for the current study.

Declarations

Ethics approval and consent to participate
This study was reviewed by the Peking University Third Hospital Medical Science Research Ethics Committee. As this was an observational study involving only educational interventions and did not involve any patient data or interventions, the committee determined that formal ethical approval was not required. The study was conducted in accordance with the ethical standards of our institution and with the 1964 Helsinki Declaration and its later amendments. The Ethics Committee waived the need for formal ethics approval and individual informed consent for this educational research.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Alaagib NA, Musa OA, Saeed AM. Comparison of the effectiveness of lectures based on problems and traditional lectures in physiology teaching in Sudan. BMC Med Educ. 2019;19(1):365.
2. Banerjee M, Chiew D, Patel KT, Johns I, Chappell D, Linton N, et al. The impact of artificial intelligence on clinical education: perceptions of postgraduate trainee doctors in London (UK) and recommendations for trainers. BMC Med Educ. 2021;21(1):429.
3. Ye Z, Dun A, Jiang H, Nie C, Zhao S, Wang T, et al. The role of 3D printed models in the teaching of human anatomy: a systematic review and meta-analysis. BMC Med Educ. 2020;20(1):335.
4. Vrillon A, Gonzales-Marabal L, Ceccaldi P-F, Plaisance P, Desrentes E, Paquet C, et al. Using virtual reality in lumbar puncture training improves students learning experience. BMC Med Educ. 2022;22(1):244.
5. Sedaghat S. Early applications of ChatGPT in medical practice, education and research. Clin Med (Lond). 2023;23(3):278–9.
6. Lee H. The rise of ChatGPT: Exploring its potential in medical education. Anat Sci Educ. 2024;17(5):926–31.
7. Ji Z, Wang G, Chen B, Zhang Y, Zhang L, Gao F, et al. Clinical application of planar puncture template-assisted computed tomography-guided percutaneous biopsy for small pulmonary nodules. J Cancer Res Ther. 2018;14(7):1632–7.
8. Jensen J, Graumann O, Jensen RO, Gade SKK, Thielsen MG, Most W, et al. Using virtual reality simulation for training practical skills in musculoskeletal wrist X-ray - A pilot study. J Clin Imaging Sci. 2023;13:20.
9. Haleem A, Javaid M, Qadri MA, Suman R. Understanding the role of digital technologies in education: a review. Sustainable Oper Computers. 2022;3:275–85.

10. Donkin R, Askew E, Stevenson H. Video feedback and e-Learning enhances laboratory skills and engagement in medical laboratory science students. BMC Med Educ. 2019;19(1):310.
11. Dempere J, Modugu K, Hesham A, Ramasamy LK. The impact of ChatGPT on higher education. Front Educ. 2023;8:1206936.
12. Choi W. Assessment of the capacity of ChatGPT as a self-learning tool in medical pharmacology: a study using MCQs. BMC Med Educ. 2023;23(1):864.

Publisher's note