Journal of Robotic Surgery (2025) 19:221
https://doi.org/10.1007/s11701-025-02309-1

RESEARCH

Multi-perspective analysis of daVinci surgical virtual reality training: a prospective randomized controlled study

Rong Wang[1] · Nan Zheng[2,3] · Yujing Liang[1] · Haoxin Cui[2,3] · Tong Ren[5] · Wenhui Xing[4] · Yanghui Li[1]

Received: 3 December 2024 / Accepted: 1 April 2025

Abstract

This study explored the impact of virtual reality (VR) training on improving the acquisition of basic robotic surgical skills by analyzing the effects of training in the daVinci Surgical System (dVSS) simulator from multiple perspectives. 27 subjects were randomly divided into a VR-training group and a control group, with the VR-training group ($n = 12$) receiving training on the dVSS simulator (XI) to achieve expert-specific proficiency status, and the control group ($n = 15$) receiving no training. All the subjects subsequently wore electroencephalography (EEG) equipment to perform 6 tasks and repeated 3 times on the dVSS simulator (SI). The Global Evaluative Assessment of Robotic Skills (GEARS) scores, dVSS scores, the National Aeronautics and Space Administration Task Load Index (NASA TLX) scores, and EEG data of all the subjects were collected to conduct a comprehensive and multi-perspective analysis for dVSS training. Learning curve analysis revealed that all trainees improved their basic robotic surgical skills and reached a steady state after training. Compared with the control group, the VR-training group received higher the GEARS scores (24.91 ± 3.36 vs. 19.68 ± 3.07; $p < 0.01$) and dVSS scores and lower the NASA TLX scores (40.04 ± 10.55 vs. 48.2 ± 9.88; $p < 0.01$). In the EEG analysis, the VR-training group had higher scores in the Beta band and the Low-gamma band in the brain regions and had greater energy activation than did the control group. This randomized controlled trial combined subjective and objective evaluations to comprehensively analyze subjects' technical and nontechnical skills. It demonstrated that training on the dVSS simulator significantly improved trainees' basic robotic surgical skills and that they could achieve better basic robotic surgical skills at lower workloads. *Trial registration*: The study was retrospectively registered at the Chinese Clinical Trial Center. The trial registration number (TRN) was ChiCTR2400088465, and the registration day was August 20, 2024.

Keywords Skills assessment · Skills training · Virtual reality · Basic robotic surgical skills · Electroencephalography

✉ Rong Wang
wangrongd@126.com

✉ Nan Zheng
nan.zheng@ia.ac.cn

[1] Department of Adult Cardiac Surgery, The Sixth Medical Center of PLA General Hospital of Beijing, Beijing 100048, China

[2] State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[3] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

[4] Faculty of Nursing, Hebei University of Chinese Medicine, Shijiazhuang 050200, China

[5] Department of Adult Cardiac Surgery, Chinese PLA Medical School, Beijing 100089, China

Abbreviations

VR	Virtual reality
dVSS	DaVinci Surgical System
EEG	Electroencephalography
GEARS	Global Evaluative Assessment of Robotic Skills
NASA TLX	National Aeronautics and Space Administration Task Load Index
MD	Mental demand
PD	Physical demand
TD	Temporal demand
P	Performance
E	Effort
CL	Cognitive load
TRN	Trial registration number
PSD	Power spectral density

Introduction

Robotic-assisted surgery has rapidly become commonplace in clinical practice. As a new technology, surgeons need to learn new skills to overcome its limitations compared with traditional surgery, such as reduced depth perception, reduced haptic feedback, the master–slave console mode, and specialized hand–eye–foot coordination [1]. Surgical simulation allows surgeons to practice surgical skills in a controlled and safe environment repeatedly [2], but the consensus [3] on training curricula and related evaluation and certification is still in the beginning stage [4].

It has been shown that robotic surgical learning has a steep learning curve [5] that requires surgeons to acquire appropriate technical and nontechnical skills [6]. Technical skills include understanding instruments, manual dexterity, camera control, clutch control, and spatial awareness; nontechnical skills include communication, cognitive skills, leadership, decision-making, and workload capacity [7]. The evaluation of surgical skills is changing from time- and procedure-volume-based evaluation [8] to proficiency-based training and certification; however, the current evaluation scales used for skill evaluation in robotic surgery are subjective, whereas objective feedback on technical skills is essential for the structured learning of surgical skills, and nontechnical skills related to the surgeon workload can also directly affect surgical outcomes [9].

The daVinci Surgical System (dVSS) simulator allows subjects to train in basic robotic surgical operations and provides a scoring system for evaluation. Previous studies [10] have shown that surgeons' skills in surgical robotic manipulation improve after training on the dVSS simulator, and most of the methods for evaluating skill improvement have been the Global Evaluative Assessment of Robotic Skills (GEARS) [11], time-based evaluations or dVSS simulator scores. Recently, more multidimensional evaluations have emerged, such as error-based operational analysis [12], anastomosis competency-based analysis [13], and physiological signal-based analysis [14], but reports on the predictive validity or benchmarking of these evaluation modalities are lacking.

The process of learning can be reflected in the learning curve. Nontechnical skills, especially mental load, directly affect surgeon performance and surgical outcomes. Cognitive Load (CL) is often characterized as the amount of finite working memory resources an individual must allocate to meet the cognitive demands of a task [15]. Howie [16] demonstrated that cognitive overload can have a negative impact on surgical training and performance. Therefore, measuring CL and employing CL-based education techniques can reduce CL overall to enhance trainees' surgical learning and encourage psychological safety. In robotic surgery training, researchers often use self-assessment methods such as the NASA-TLX [17] to measure task workload. In addition, physiologic signals and other objectively measurable data are collected to establish correlations with nontechnical skills for evaluation. EEG [18], heart rate [19], and eye movements [20] are the most studied signals that indicate complex behaviors and brain load. EEG measures electrical activity in the brain and can assess cognitive processing. It has been used to continuously assess brain activity during surgical performance [21]. While existing literature has utilized EEG for skill assessment, its focus has predominantly been on classifying experienced and novice surgeons through EEG spectral analysis [22], without considering the changes in EEG and their significance in different scenarios, such as surgical training and different surgical procedures. However, the correlations between physiological signals and nontechnical skills still need to be explored in further research.

Therefore, a single evaluation in existing studies may not be able to comprehensively and effectively evaluate the level of surgeons' technical and nontechnical skills in robotic surgical operations [23]. To validate the effectiveness of the dVSS simulator for training surgeons in basic robotic surgical skills, we conducted a prospective randomized controlled trial in which, first, the minimum number of training sessions required for trainees to become proficient in basic robotic surgical skills was determined through post-training learning curve plotting. Second, differences in technical skills between the training group and the control group on the dVSS simulator were compared through a multi-perspective evaluation analysis. Finally, the differences in cognitive activities between the training group and the control group during the experimental operation were explored through EEG analysis [24], which analyzed and assessed the subjects' nontechnical skills such as brain work and cognitive load. The effects of the dVSS simulator on the training of basic robotic surgical skills and the influence of nontechnical skills on their ability to learn skills were comprehensively assessed.

Methods

Subject selection and randomization

All the subjects were informed about the experiment and signed an informed consent form. The experimental data were collected from January 2024 to April 2024. The study was registered at the Chinese Clinical Trial Center (ChiCTR2400088465).

All the subjects were divided into Residents (≤ 5 years of surgical experience) and Fellows (> 5 years of surgical experience) with 5 years of surgical experience, and then randomly divided into the control group and the VR-training group via random number generation within the two tiers of the population, which ensured the balance of the number of cases between the groups and the reliability of the results of the experiment via hierarchical randomization grouping. The control group was taught the basic operations of the dVSS simulator (XI), such as the introduction of robot components, interpretation of operational tasks, and adaptation to robotic arm operation. The VR-training group was required to complete operational tasks on the dVSS simulator (XI) and achieve proficiency. Before randomization into groups, all the subjects completed a basic demographic questionnaire. All the subjects were randomized into groups according to 1:1 stratification (Fig. 1).

Design of the experiment

The training was conducted on the dVSS simulator (XI), a commercially available robotic surgical simulation platform specifically designed for training basic robotic surgical skills and widely accepted by colleagues worldwide. We surveyed two robotic surgery specialists and selected 6 tasks from the dVSS simulator (XI) and the other 6 tasks from the dVSS simulator (SI) to form a set of training task lists and experimental task lists respectively (Appendix 1 in supplementary material). Proficiency in the training tasks was achieved by two robotic surgery specialists operating them until they were satisfied with their performance and felt that they could demonstrate their competence. Therefore, we defined training to proficiency as each trainee operating each task 10 or

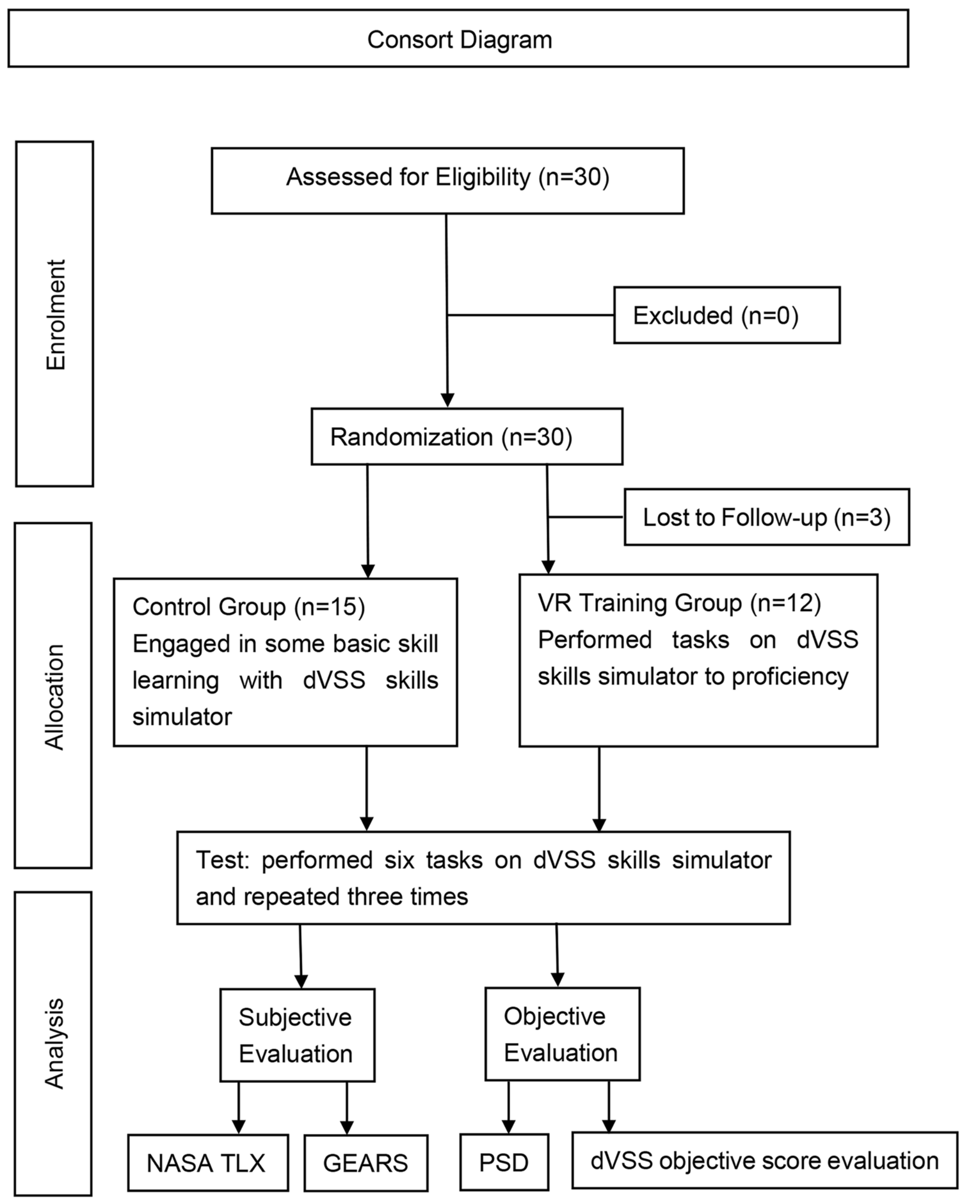


Fig. 1 Flowchart of the randomized grouping used in the experiment. *NASA TLX* National Aeronautics and Space Administration Task Load Index; *GEARS* Global Evaluative Assessment of Robotic Skills. *dVSS* daVinci Surgical SystemTM (Intuitive Surgical Inc); *PSD* power spectral density

more times and scoring 85 or more consecutively three times during training.

After all the trainees achieved proficiency in each task, all the subjects wore EEG equipment (Siesta Wireless EEG Amplifier, Compumedics, Australia) to experiment on the dVSS simulator with 3 repeating rounds of 6 experimental tasks.

Evaluation methods

Learning curve

The learning curve was used to intuitively compare training results [25], learning curves were drawn for both the Fellows and Residents who participated in training and reached a qualified status, and the evaluation metrics shared in the 6 training tasks were analyzed. Learning curve evaluation is limited to the first 10 practice attempts in training. To allow for comparisons between tasks of varying difficulty, all 6 task scores were converted to a standardized *Z*-score.

GEARS

The GEARS score [26] was used to assess the robotic basic surgical skill level, which consists of 6 evaluation dimensions: depth perception, hand coordination, operating efficiency, autonomous operation, force perception, and mechanical arm control (Appendix 2 in supplementary material). The score is based on a five-point Likert scale, with higher scores indicating better performance. In this study, the GEARS score was evaluated by 10 graduate students with non-medical backgrounds, and they watched the operation video and scored according to the six evaluation dimensions of the subject's ability of the spatial depth of the moving camera, the coordination of the two-handed moving, the time to complete the task, the fluency of the operation, the ability to complete the task autonomously, the perception and control of the object's compression deformation. Since the GEARS score requires the evaluator to spend a lot of time watching the operation video, the senior robotic surgeon has limited time and energy. Study [27] shows that there was a high degree of correlation between senior robotic surgeons and the general public when they watched videos of robotic operations for GEARS rating. After the general public was explained and screened by the GEARS evaluation index, the evaluation scores of a large number of people were averaged, and it was found that a larger amount of data could cover up their shortcomings of without a professional background in clinical surgery, so there was no statistically significant difference between the GEARS scores of the general public and experts. Therefore, in this experiment, 10 graduate students with non-medical backgrounds were invited to perform blind analysis of the robot operation videos of all subjects, 10 evaluators were trained to evaluate before scoring, and then a relatively objective GEARS score was obtained, and all data analysis used the original scoring values.

NASA TLX

The NASA TLX score [28] was used to assess the level of workload after completing a robotic task. This subjective skill evaluation survey has six indices: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (P), Effort (E), and Frustration (F). Based on the definition of the NASA TLX scales, the score of PD reflects the workload of the motor component, the scores of TD, F, P, and E represent the overall cognitive workload, and the score of MD is assessed by the perceptual and cognitive demand. In this experiment, between each of the 2 tasks, the subjects were asked to self-assess their workload according to their status, and at the end of the rating, they were asked to rank the six indices according to the impact of workload. The score range of each evaluation index is 0–100 points, and the total score is calculated according to the weight distribution of the six evaluation dimensions according to the ranking of the subjects. The NASA TLX score is the average of the total score weighted by 9 weighted calculations.

dVSS score

The dVSS score is an objective scoring system in the dVSS simulator. After each task, the system will be rated directly according to performance, and its scoring dimensions vary according to the task. We select 5 scores for evaluation and analysis, including time, economy of exercise, instrument collisions, master workspace range, and overall score.

PSD

Power Spectral Density (PSD) [29] reflects the power distributions of different frequency components in EEGs and provides insight into brain activation. To accurately estimate the PSD, especially in the case of nonstationary EEG signals, we used Welch's method for Fourier transform [30], which helps to reduce the variance in the estimation process. The frequency range analyzed was 0–40 Hz: delta (0–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and low-gamma (30–40 Hz). PSDs were calculated for each channel and averaged for the whole brain.

Analysis of data

The performance data were selected from the dVSS simulator scoring system and *Z*-Scores were calculated. Learning curves and box plots were plotted via GraphPad (Prism version 9.1.0, GraphPad Software, La Jolla California USA). Parametric analyses were performed via SPSS (IBM SPSS Statistics for Macintosh, Version 25.0, Armonk, NY: IBM Corp) on the GEARS score, NASA TLX score, and dVSS simulator score data. MATLAB R2013b and the EEGLAB toolkit were used to preprocess the EEG data to eliminate artifacts and noise, and the Python-based MNE toolkit was used for PSD analysis and visualization. MANOVA was performed to compare PSD between groups. All analyses were two-sided tests and had a significance level of 0.05.

Results

Demographic characteristics

In the VR-training group, 3 subjects voluntarily withdrew from the experiment for personal reasons. Among the 27 subjects, 12 Residents and 15 Fellows, there were no significant differences between the baseline characteristics of the control group and the VR-training group in terms of basic demographic characteristics, surgical experience, or dVSS experience (Table 1).

Learning curve assessment

Training effects were compared directly through learning curve analysis. The learning curve (Fig. 2) revealed significant improvements in the performance of Fellows and Residents after training, with no significant differences in the total time ($p = 0.675$), the economy of the exercise ($p = 0.649$), or the overall score ($p = 0.802$). Particularly during the first 4 training periods, the trainees' performance improvements were most noticeable in terms of total time, performance of the activity, and total score, after the 5th training, the learning curves reached the platform period, and the trainees' performance began to stabilize.

dVSS score assessment

In the analysis of the dVSS scores, there was no significant difference between the Residents and the Fellows in terms

Table 1 Demographic characteristics of the participants

Participant characteristic	Control ($n = 15$)	VR-training ($n = 12$)
Demographics		
Median Age (range) (years)	35(25–53)	33(24–44)
Dominant hand, *n*		
Left	0	0
Right	15	12
Three-dimensional vertigo, *n*		
Yes	0	0
No	15	12
Specialty		
Urology	0	4
Cardiology	10	8
General surgery	5	0
Surgical experience		
Formal endoscopic surgery, *n*		
Yes	9	9
No	6	3
Conventional surgery, *n*		
Yes	15	11
No	0	1
Median years of surgical working (range) (years)	10(1–31)	8(1–21)
dVSS experience		
Previous Robotic console Experience, *n*		
Yes	0	3
No	15	9
Previous simulator training for the dVSS, *n*		
Yes	0	0
No	15	12

dVSS daVinci Surgical System

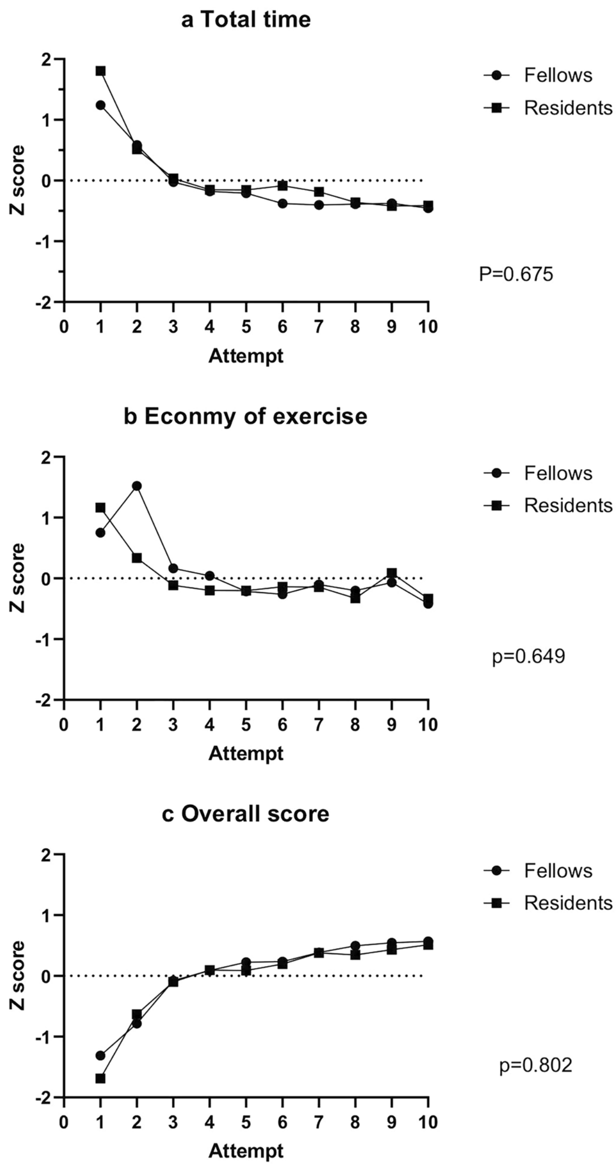


Fig. 2 Trainees' learning curves regarding total time (**a**), economy of exercise (**b**), and overall score (**c**). The *Z*-score calculation was used to compare scores between tasks of varying difficulty. The independent samples t-tests were used to calculate the difference in Z scores between Fellows ($n=6$) and Residents ($n=6$) in the first 10 attempts during training. **a** Total time($p=0.675$). **b** Economy of exercise ($p=0.649$). **c** Overall score ($p=0.802$)

of performance and scores on all tasks. In contrast, the VR-training group significantly outperformed the control group in terms of performance and scores for most tasks (Table 2). In particular, the VR-training group outperformed the control group in the composite performance of task 6 ($p<0.01$).

PSD assessment

In distance difference visualization, different bands under the same task are normalized to highlight the brain regions that can reflect different operation levels under the same task. The red color indicates that the PSD of this brain region in the training group is higher than that in the control group, and the blue color indicates that the PSD of this brain region in the training group is lower than that in the control group. The darker color of the PSD difference of the brain region in the groups with different operation levels is greater. In tasks 1, 2, 3, and 6, the Beta band is darker, and in tasks 1, 4, and 5, the Low-gamma band is darker (Fig. 3). Most of the brain regions in the training group had greater PSDs than did those in the control group.

NASA TLX assessment

There was no significant difference in the NASA TLX scores for each task between the Residents and the fellows, but the Residents had a lower overall task load. Compared with the control group, the VR-training group had significantly lower Physical demand (3.82 ± 2.46 vs. 8.42 ± 6.04; $p<0.01$) and NASA TLX overall score (40.04 ± 10.55 vs. 48.2 ± 9.88; $p<0.01$) and significantly higher Performance (18.72 ± 8.09 vs. 12.55 ± 5.62; $p<0.01$) (Table 3). Therefore, the physical demand and overall workload of the VR-training group were lower than those of the control group, and the satisfaction with the performance of the task operation was higher than that of the control group.

GEARS assessment

The Residents and the Fellows had no significant difference in the overall score of the GEARS (21.08 ± 4.31 vs. 22.74 ± 3.92; $p=0.31$), whereas the VR-training group had a significant superiority over the control group in the total score of the GEARS (24.91 ± 3.36 vs. 19.68 ± 3.07; $p<0.01$) (Fig. 4).

Discussion

The effectiveness of simulation training is increasingly recognized in surgical education [31]. Studies [32] have been conducted to validate the efficacy of dVSS training through animal experiments and surgery-based simulations. Nevertheless, there is a paucity of clarity and consistency in the literature regarding the quantifiable metrics and temporal parameters of dVSS training that precede the validation process. Studies of VR simulator training in traditional surgery and laparoscopic surgery have even confirmed the effectiveness of the transfer of training skills to clinical operations.

Table 2 dVSS objective score evaluation of 6 tasks

Skill	Task	*p*	
		VR Training($n = 12$) VS Control ($n = 15$)	Fellows($n = 15$) VS Residents($n = 12$)
Time/s	Task1	< 0.01	0.39
	Task2	0.11	0.05
	Task3	< 0.01	0.58
	Task4	0.04	0.19
	Task5	0.02	0.50
	Task6	0.01	0.51
Economy of exercise/cm	Task1	< 0.01	0.89
	Task2	0.11	0.25
	Task3	< 0.01	0.89
	Task4	< 0.01	0.30
	Task5	0.02	0.84
	Task6	< 0.01	0.83
Instrument collisions	Task1	0.04	0.47
	Task2	0.42	0.05
	Task3	0.08	0.45
	Task4	0.38	0.30
	Task5	0.52	0.18
	Task6	< 0.01	0.83
Master workspace range/cm	Task1	0.05	< 0.01
	Task2	0.52	0.33
	Task3	0.07	0.44
	Task4	0.95	0.23
	Task5	0.65	0.96
	Task6	0.22	0.07
Overall score	Task1	< 0.01	0.54
	Task2	0.19	0.13
	Task3	< 0.01	0.87
	Task4	< 0.01	0.37
	Task5	0.42	0.33
	Task6	< 0.01	0.39

Clinical benefit extrapolation of training outcomes is one of the most important aspects of validating the effectiveness of training, whereas the impact of training on real surgical outcomes is characterized by a certain degree of hysteresis [33].

Unlike conventional and laparoscopic surgery, the learning curve for robot-assisted surgery is steeper [34], and the increased incidence of adverse clinical outcomes at the beginning of the learning curve remains a concern. A growing number of studies have extensively validated various VR simulators [35]. Abinaya [36] demonstrated that training with a VR simulator that provides force feedback improves haptic realism and compensates for the limitations of measuring tissue damage only through virtual collision-based visual feedback. Raison [32] demonstrated training with a RobotiX Mentor simulator to complete a simulation-based training program for robot-assisted radical prostatectomy and concluded that surgery-based VR training was more effective than VR simulation of basic skills. The learning and training of nontechnical skills, including decision-making and judgment, has also emerged as an important part of robotic simulation training. Zhang [37] demonstrated a significant improvement in arthroscopy skills after Residents were trained in 3 arthroscopy skills on a simulator. The learning curves revealed that all 3 arthroscopy skills decreased significantly after 7 days of no training, and skill retention decreased with difficulty level. The higher the difficulty level of the skill is, the lower the skill retention rate. Specifically, comprehensive training and timely effective evaluation feedback are beneficial for shortening the learning curve [38]; however, the ideal training model and comprehensive evaluation methods are still in the exploratory stage of development [3]. In this study, we combined the mainstream VR simulator on the market and adopted the opinions of robot-assisted surgery experts to select

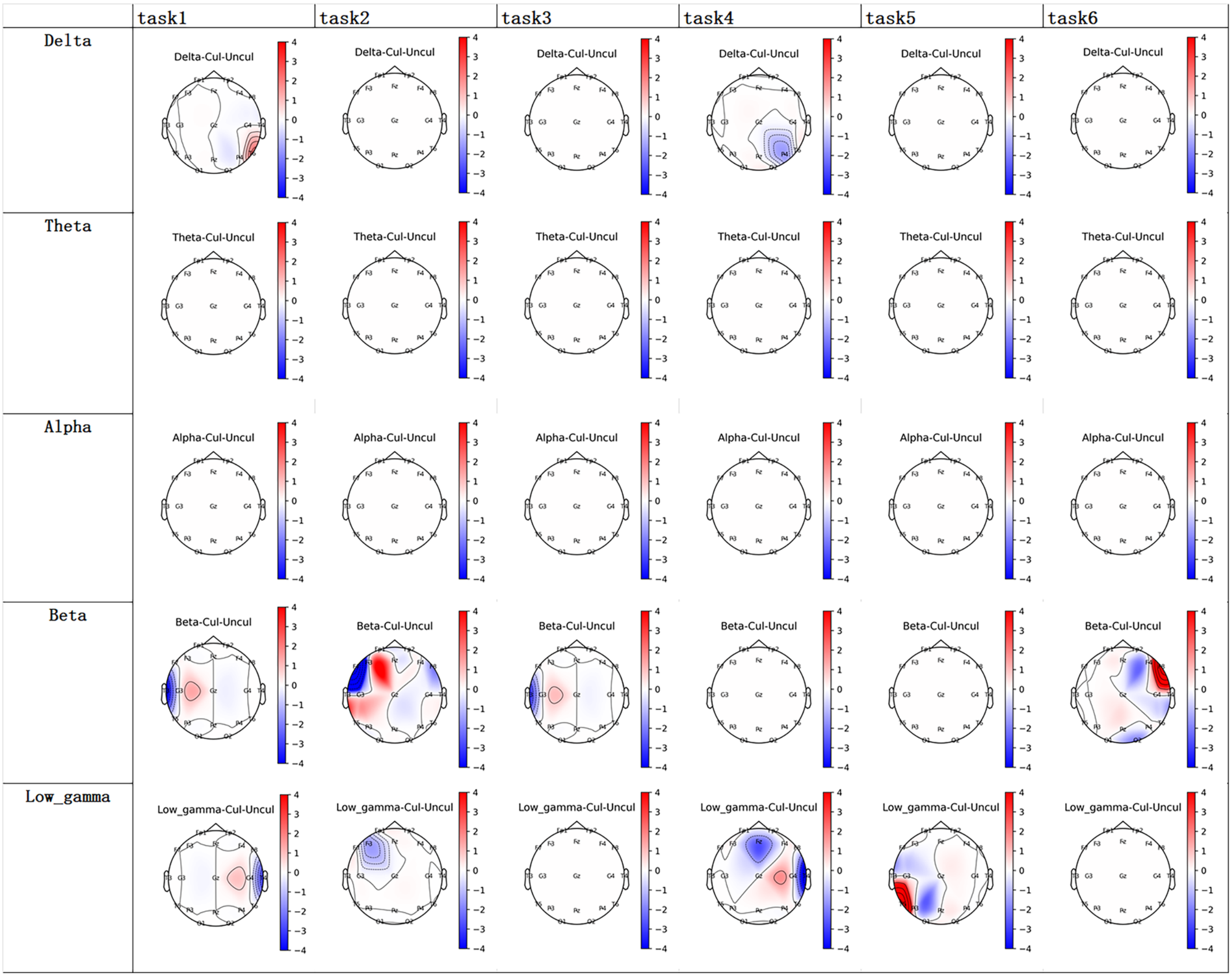


Fig. 3 Brain topography of channel differences under different Band-Duty between VR-training group versus control group. The PSD values of the EEG were calculated by Continuous Time Fourier Transform and the results were assigned positive or negative by the difference in the means of the two groups and mapped back to the brain topography

appropriate tasks to cover basic robotic surgical skills to form a list of training tasks. The experimental data were analyzed through learning curves to show the effectiveness of VR training and then combined with subjective and objective evaluation methods for basic robotic surgical skill assessment to provide a more accurate assessment of dVSS training effectiveness.

The learning curves visualize the process of achieving proficiency in VR training [39], and in this study, all the subjects, regardless of years of surgical experience, achieved proficiency in 10 attempts. The learning curves were steep in the first 4 attempts, which belonged to the stage of acquiring new surgical knowledge, and this stage was prone to errors due to unfamiliarity with the operation, thus resulting in lower scores. We found that from the 5th attempt onward, the learning curve showed a stabilizing trend in 3 metrics and entered the process of repeating and deepening the learning process of surgical knowledge, which led to a stable state of proficiency. In our experiments, the learning curves revealed that 5 repetitions were the average number required for trainees to reach proficiency, and each repetition subsequently increased the level of proficiency.

The GEARS score is a validated scoring system for the assessment of robot-assisted surgery, focusing on general performance during surgical robot manipulation, independent of the specialized proficiency in surgical manipulation required for a particular procedure. In most studies [40], scoring was performed by experienced robot-assisted surgery surgeons to avoid bias in scoring, given that the scoring was subjective in six dimensions on the basis of video recordings of the operation. Therefore, 10 postgraduate students were selected to rate the videos of the trials in a blinded manner. We found that surgeons with more experience had higher GEARS scores, but the difference was

Table 3 National Aeronautics and Space Administration Task Load Index assessment

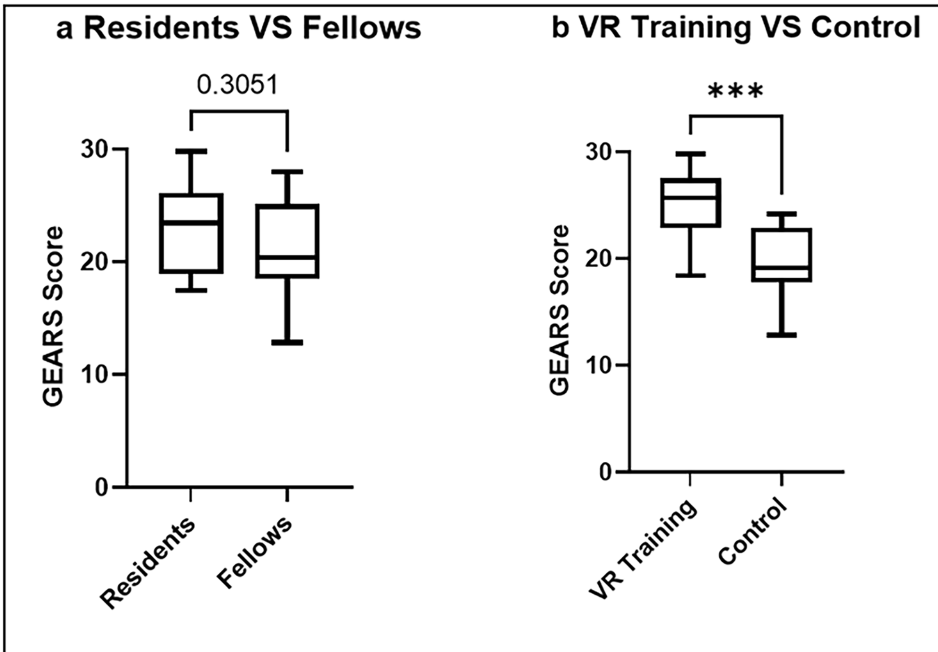
Residents VS Fellows group			
Load type	Residents group, $n = 12$ Mean (SD) NASA TLX score	Fellows group, $n = 15$ Mean (SD) NASA TLX score	p
Mental demand	6.89(6.14)	6.12(4.45)	0.71
Physical demand	5.57(4.40)	5.02(5.40)	0.78
Temporal demand	6.13(4.42)	4.31(3.47)	0.24
Performance	13.30(8.13)	15.29(6.12)	0.48
Effort	7.98(5.36)	6.17(3.92)	0.32
Frustration	6.34(6.49)	3.39(3.74)	0.15
Overall score	46.15(9.95)	40.31(10.18)	0.15
VR-training VS control group			
Load type	VR group, $n = 12$ Mean (SD) NASA TLX score	Control group, $n = 15$ Mean (SD) NASA TLX score	p
Mental demand	4.63(2.84)	7.93(6.10)	0.14
Physical demand	3.82(2.46)	8.42(6.04)	< 0.01
Temporal demand	4.62(3.90)	5.52(4.07)	0.57
Performance	18.72(8.10)	12.55(5.62)	< 0.01
Effort	6.64(6.02)	7.24(3.28)	0.75
Frustration	3.60(4.22)	5.59(5.94)	0.34
Overall score	40.04(10.55)	48.20(9.88)	< 0.01

NASA TLX National Aeronautics and Space Administration Task Load Index

not significant. Therefore, basic robotic surgical skills differ from conventional skills, and increased experience in general surgical operations may not have a significant correlation with the learning basic robotic surgical skills. Although expertise and performance frequently have been used interchangeably, they do not convey the same meaning in the field of robot-assisted surgery. Even the performance of expert surgeons may be poor in certain surgical situations, i.e., cases with intra-operative challenges. In contrast, the VR-training group achieved a significant improvement in the GEARS score compared with the control group, demonstrating the effectiveness of training.

Using the NASA TLX score to evaluate the task load of the subjects [17], in this experiment, the VR-training group had a significantly lower task load in terms of physical demands, performance satisfaction, and total score, and a lower task load in terms of mental demands, frustration, and effort and was more satisfied with their performance than the control group. In contrast, experienced surgeons had lower physical demands, mental demands, frustration, effort, and total scores and were more satisfied with their performance. Whether there is a relationship between task load and skill acquisition in the early stages of learning new surgical skills has been controversial [28], and it may also be possible that experienced surgeons have a lower task load in all areas [41].

In the EEG evaluation, the Beta band and the Low-gamma band energy activation was higher in the training group than in the control group. One study [42] showed that the Beta band correlates with cognitive load, which reflects the activation of brain regions required for higher-order cognitive and motor processes. The brain works harder to maintain control and focus and concentrates more on processing complex information and executing precise movements [43, 44]. The EEG analysis also echoed the results of the subjective task load score of the NASA TLX score, i.e., in VR simulator operations, the mental workload decreases as proficiency increases [45], and the training group was able to perform the VR task effectively with a lower cognitive load [46]. The low-gamma band is related to cognitive functions, including attention, memory, and information processing. The PSD of the Low-gamma band was higher in the training group than in the control group, which could be attributed to the increased synchronization and engagement of neural circuits involved in the motor task. This finding suggests that the enhancement of PSD in the low-gamma band of the trained operators may be due to more efficient neural circuits in the motor region as a result of repeated practice. In addition, the dynamic response of low-gamma band power to complex stimuli [47], especially in visual and motor tasks, highlights the role of low-gamma oscillations in integrating sensory information and motor planning, further explaining the differences between the training and control groups. Kabbara [48] reported that specialists in a particular skill show more efficient neural processing, characterized by



	Fellows group, n=15 Mean (SD)	Residents group, n=12 Mean (SD)	p
GEARS Score	22.74 (3.92)	21.08 (4.31)	0.305
	VR group, n=12 Mean (SD)	control group, n=15 Mean (SD)	p
GEARS Score	24.91 (3.36)	19.68 (3.07)	< 0.01

Fig. 4 Comparison of GEARS scores in groups. **a** GEARS score between Residents and Fellows groups (21.08 ± 4.31 vs. 22.74 ± 3.92; $p = 0.305$). **b** GEARS score between VR-training and control groups (24.91 ± 3.36 vs. 19.68 ± 3.07; $p < 0.01$)

less localized activity but more integrated brain activity. It means that the training group achieved a balance between local specialization and global integration, thus facilitating complex problem-solving and decision-making. This provides a guiding direction for future training in the field of robot-assisted surgery, targeting effective training to enable surgeons to have better performance under moderate cognitive load conditions.

There are also several limitations in this study. First, the dVSS simulator score may not be a valid system for assessing the performance of the subjects, which is scored according to each task and assesses the skills without involving the specific surgical procedures, which fails to simulate the real surgical environment more realistically for operation, so the objective score may not be a true response to the acquisition of robotic surgical skills. Second, the NASA TLX score and the GEARS score are subjective scoring methods, which are the subjective self-evaluation and the subjective evaluation of others, respectively, so the scores used may not fully represent the subjects' task load and the degree of mastery of basic robotic surgical skills. Nonetheless, we provide a multi-perspective evaluation of robotic surgical skill acquisition, and the conclusion of the effectiveness of simulation training in basic robotic surgical skills drawn from the combination of the three scores is more reliable than that of a single score. In addition, more in-depth objective EEG analyses provide mental-level evidence for evaluating the effectiveness of operational training.

Conclusion

This multi-perspective evaluation analysis proved that dVSS simulator training could significantly improve the acquisition of basic robotic surgical skills and that the trainees were able to achieve higher performance under a lower workload. In particular, the way of evaluating nontechnical skills such as cognitive load and brain work by combining EEG analysis further scientifically explains the differences in the activation and connectivity of brain regions in different populations when operating the surgical robot in the master–slave mode, and reveals the connection between physiological signals and non-technical skills. In this study, non-technical skills such as EEG and cognitive load were evaluated. This result provides a guiding direction for future training in the field of robot-assisted surgery and enables surgeons to make better robotic surgery operation performance under a certain pressure load through targeted and standardized training. Future research could explore the factors that may influence the performance of basic robotic surgical skills in more realistic simulation environments [49], and explore more comprehensive and rational ways to evaluate mastery of the skill [50, 51].

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11701-025-02309-1.

Author contributions Y.J. L. and H.X. C. contributed equally to this work and are joint senior authors. R. W. obtained funding. Y.J. L. designed the study. Y.H. L., T. R. and W.H.X. collected the data. Y.J. L. and H.X. C. were involved in data analysis. N. Z. and R. W. contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. All authors have read and approved the final manuscript. N. Z. and R. W. are joint corresponding authors.

Funding This work was supported by the National Key Research and Development Program of China [grant number 2022YFB4700800].

Availability of data and materials No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval and consent to participate The research performed in accordance with the Declaration of Helsinki. Ethical approval and consent to participate Ethical approval for this study was granted by The Six Medical Center of PLA General Hospital's ethics committee (HZKY-PJ-2023-52). The participants in this study gave their written informed consent to take part in this study and for anonymized findings of this study to be published. All authors have read, approved and given their consent for the manuscript to be published should it be accepted by the journal.

Consent for publication Not applicable.

Clinical trial number The study was registered at the Chinese Clinical Trial Center (ChiCTR2400088465).

References

1. Alvarez-Lopez F, Maina MF, Arango F et al (2020) Use of a low-cost portable 3D virtual reality simulator for psychomotor skill training in minimally invasive surgery: task metrics and score validity. JMIR Serious Games 8(4):e19723
2. James HK, Chapman AW, Pattison GTR et al (2019) Systematic review of the current status of cadaveric simulation for surgical training. Br J Surg 106(13):1726–1734
3. Artsen AM, Burkett SL, Duvvuri U et al (2022) Surgeon satisfaction and outcomes of tele-proctoring for robotic gynecologic surgery. J Robot Surg 16(3):563–568
4. Lu J, Cuff RF, Mansour MA (2021) Simulation in surgical education. Am J Surg 221(3):509–514
5. Louridas M, Iancu AM, Grantcharov T et al (2023) Modeling technical skills learning curves of incoming surgical trainees. J Surg Educ 80(1):51–61
6. Schmidt MW, Köppinger KF, Fan C et al (2021) Virtual reality simulation in robot-assisted surgery: meta-analysis of skill transfer and predictability of skill. BJS Open 5(2):zraa066
7. Sinha A, West A, Vasdev N et al (2023) Current practises and the future of robotic surgical training. Surgeon 21(5):314–322
8. Mazzon G, Sridhar A, Busuttil G et al (2017) Learning curves for robotic surgery: a review of the recent literature. Curr Urol Rep 18(11):89
9. Cha JS, Yu D (2022) Objective measures of surgeon non-technical skills in surgery: a scoping review. Hum Factors 64(1):42–73
10. Abreu AA, Rail B, Farah E et al (2023) Baseline performance in a robotic virtual reality platform predicts rate of skill acquisition in a proficiency-based curriculum: a cohort study of surgical trainees. Surg Endosc 37(11):8804–8809
11. Goh AC, Goldfarb DW, Sander JC et al (2012) Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. J Urol 187(1):247–252
12. Goldenberg MG, Goldenberg L, Grantcharov TP (2017) Surgeon performance predicts early continence after robot-assisted radical prostatectomy. J Endourol 31(9):858–863
13. Schmidt MW, Haney CM, Kowalewski KF et al (2022) Development and validity evidence of an objective structured assessment of technical skills score for minimally invasive linear-stapled, hand-sewn intestinal anastomoses: the A-OSATS score. Surg Endosc 36(6):4529–4541
14. Cowan A, Chen J, Mingo S et al (2021) Virtual reality vs dry laboratory models: comparing automated performance metrics and cognitive workload during robotic simulation training. J Endourol 35(10):1571–1576
15. Szulewski A, Howes D, van Merriënboer JJG et al (2021) From theory to practice: the application of cognitive load theory to the practice of medicine. Acad Med 96(1):24–30
16. Howie EE, Dharanikota H, Gunn E et al (2023) Cognitive load management: an invaluable tool for safe and effective surgical training. J Surg Educ 80(3):311–322
17. Lowndes BR, Forsyth KL, Blocker RC et al (2020) NASA-TLX assessment of surgeon workload variation across specialties. Ann Surg 271(4):686–692
18. Khosla A, Khandnor P, Chand T (2020) A comparative analysis of signal processing and classification methods for different applications based on EEG signal. Biocybern Biomed Eng 40(2):649–690
19. Cha JS, Athanasiadis DI, Peng Y et al (2024) Objective nontechnical skills measurement using sensor-based behavior metrics in surgical teams. Hum Factors 66(3):729–743
20. Narasimha S, Obuseh M, Anton NE et al (2024) Eye tracking and audio sensors to evaluate surgeon's non-technical skills: an empirical study. Appl Ergon 119:104320

21. Shafiei SB, Shadpour S, Intes X et al (2023) Performance and learning rate prediction models development in FLS and RAS surgical tasks using electroencephalogram and eye gaze data and machine learning. Surg Endosc 37(11):8447–8463
22. Natheir S, Christie S, Yilmaz R et al (2023) Utilizing artificial intelligence and electroencephalography to assess expertise on a simulated neurosurgical task. Comput Biol Med 152:106286
23. Huffman EM, Rosen SA, Levy JS et al (2021) Are current credentialing requirements for robotic surgery adequate to ensure surgeon proficiency? Surg Endosc 35(5):2104–2109
24. Mumtaz W, Xia L, Ali SSA et al (2017) Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD). Biomed Signal Process Control 31:108–115
25. Bartlett JD, Lawrence JE, Yan M et al (2020) The learning curves of a validated virtual reality hip arthroscopy simulator. Arch Orthop Trauma Surg 140(6):761–767
26. Sánchez R, Rodríguez O, Rosciano J et al (2016) Robotic surgery training: construct validity of Global Evaluative Assessment of Robotic Skills (GEARS). J Robot Surg 10(3):227–231
27. Polin MR, Siddiqui NY et al (2016) Crowdsourcing: a valid alternative to expert evaluation of robotic surgery skills. Am J Obstet Gynecol 215(5):644.e1-644.e7
28. Bell SW, Kong JCH, Clark DA et al (2022) The National Aeronautics and Space Administration-task load index: NASA-TLX: evaluation of its use in surgery. ANZ J Surg 92(11):3022–3028
29. Yue K (2018) The correlation between visual fatigue and duration of viewing as assessed by brain monitoring. Soc Inf Display 26(7):427–437
30. Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134(1):9–21
31. Pongeluppi RI, Coelho G, Ballestero MFM et al (2024) Development and evaluation of a mixed reality model for training the retrosigmoid approach. World Neurosurg 189:e459–e466
32. Raison N, Harrison P, Abe T et al (2021) Procedural virtual reality simulation training for robotic surgery: a randomised controlled trial. Surg Endosc 35(12):6897–6902
33. Hung AJ, Chen J, Che Z et al (2018) Utilizing machine learning and automated performance metrics to evaluate robot-assisted radical prostatectomy performance and predict outcomes. J Endourol 32(5):438–444
34. Soomro NA, Hashimoto DA, Porteous AJ et al (2020) Systematic review of learning curves in robot-assisted surgery. BJS open 4(1):27–44
35. Bjerrum F, Collins JW, Butterworth J et al (2023) Competency assessment for the Versius surgical robot: a validity investigation study of a virtual reality simulator-based test. Surg Endosc 37(10):7464–7471
36. Abinaya P, Manivannan M (2024) Haptic based fundamentals of laparoscopic surgery simulation for training with objective assessments. Front Robot AI 11:1363952
37. Zhang KJ, Zhou H et al (2023) Learning and short-term retention of simulation-based arthroscopic skills. J Surg Educ 80(1):119–126
38. Tom CM, Maciel JD et al (2019) A survey of robotic surgery training curricula in general surgery residency programs: how close are we to a standardized curriculum? Am J Surg 217(2):256–260
39. Feldman LS, Cao J et al (2009) A method to characterize the learning curve for performance of a fundamental laparoscopic simulator task: defining "learning plateau" and "learning rate." Surgery 146(2):381–386
40. Hu Y, Jiang B et al (2018) Vessel ligation fundamentals: a comparison of technical evaluations by crowdsourced nonclinical personnel and surgical faculty. J Surg Educ 75(3):664–670
41. Li S, Liu Y, Li K et al (2023) Validation and effect of the NASA-TLX score on the assessment of the workload of pediatric robotic operations. Surg Endosc 37(7):5077–5085
42. Chikhi S, Nadine M et al (2022) EEG power spectral measures of cognitive workload: a meta-analysis. Psychophysiology 59(6):e14009
43. Keles HO, Cengiz C, Demiral I et al (2021) High density optical neuroimaging predicts surgeons's subjective experience and skill level. PLoS ONE 16(2):e0247117
44. Modi HN, Singh H, Fiorentino F et al (2019) Association of residents' neural signatures with stress resilience during surgery. JAMA Surg 154(10):e192552–e192552
45. Modi HN, Singh H, Darzi A et al (2020) Multitasking and time pressure in the operating room: impact on surgeons' brain function. Ann Surg 272(4):648–657
46. Nemani A, Kamat A, Gao Y et al (2021) Functional brain connectivity related to surgical skill dexterity in physical and virtual simulation environments. Neurophotonics 8(1):015008–015008
47. Pi YL, Wu XH, Wang FJ et al (2019) Motor skill learning induces brain network plasticity: a diffusion-tensor imaging stud. PLoS ONE 14(2):e0210015
48. Kabbara A et al (2016) Graph analysis of spontaneous brain network using EEG source connectivity. arXiv preprint. https://doi.org/10.48550/arXiv.1607.00952
49. Birrenbach T, Stuber R, Müller CE et al (2024) Virtual reality simulation to enhance advanced trauma life support trainings - a randomized controlled trial. BMC Med Educ 24(1):666
50. Boal MWE, Anastasiou D, Tesfai F et al (2024) Evaluation of objective tools and artificial intelligence in robotic surgery technical skills assessment: a systematic review. Br J Surg 111(1):znad331
51. Wilcox Vanden Berg RN, Vertosick EA, Sjoberg DD et al (2024) Implementation and validation of an automated, longitudinal robotic surgical evaluation and feedback program at a high-volume center and impact on training. Eur Urol Open Sci 62:81–90