Mesopotamian Journal of Artificial Intelligence in Healthcare
Vol.2026, **pp.** 1–17
DOI: https://doi.org/10.58496/MJAIH/2026/001; ISSN: 3005-365x
https://mesopotamian.press/journals/index.php/MJAIH

Research Article

An Augmented Reality-Based Simulator for Enhancing Surgical Training and Skill Acquisition

Mohammed Hazim Alkawaz [1,*], Younis Al-Arbo [1], Meaad Mohammed Salih [1]

[1] *Department of Computer Science, College of Education for Pure Science, University of Mosul, Mosul, Iraq.*

ARTICLE INFO

Article History
Received 2 Nov. 2025
Revised 25 Dec. 2025
Accepted 16 Jan. 2026
Published 19 Feb. 2026

Keywords
Augmented Reality (AR),
Surgical Simulation,
Skill Acquisition,
Medical Education,
Surgical Training,
Spatial Cognition,
Simulator-Based
Learning.

ABSTRACT

Conventional surgical training systems are hampered by restricted case exposure, expense, and patient safety, whereas VR simulators often lack contextual and tactile realism. In this paper, an optical see-through AR simulator that overlays 3D anatomical guidance and instrument trajectory cues onto a physical phantom and provides real-time multimodal feedback and automated scoring are developed. A prospective controlled study (n=30 residents) compared AR training with a VR simulator and conventional model-based practice for a standardized orthopedic drilling task. The results were scores on completion time, placement deviation, error count, Objective Structured Assessment of Technical Skill (OSATS) global rating, confidence, decision-making accuracy, and 1-month skill retention.AR training achieved more efficient task completion (5.3±0.8 min vs 6.4±1.0 VR and 7.9±1.3 traditional; $p<0.01$), reduced placement error (1.5±0.3 mm vs 2.2±0.5 and 4.1±0.8; $p<0.001$), fewer errors, and higher OSATS scores (30.5±2.1 vs 27.8±3.0 and 23.4±3.5). Retention remained highest for AR at one month (28.9±2.5). Participants reported higher confidence (4.7/5). Therefore, AR can combine the physical fidelity of hands-on practice with in-situ guidance to accelerate skill acquisition and improve retention, supporting integration of AR simulation into surgical curricula.

1. INTRODUCTION

For most generations, surgical education has had an apprenticeship model, during which the trainee works their way up through observation and supervised participation. Despite its enduring importance, this approach is limited by standards of patient safety, a lack of opportunities for the repeated application of these practices, as well as an increasingly technical challenge with recent modern procedures. As a consequence, simulation-based training is now a cornerstone supplemental resource, which enables active training, uniform evaluation and error-tolerant rehabilitation beyond the operating room [1]. There really is so much evidence in favor of simulation. This fundamental work shows that trained individuals using structured simulation and VR practice convert these into skills that positively impact operating-room performance [2]. There are countless randomized trials within orthopaedics have reported that immersion in VR training enhances technical skill, procedural knowledge, and skill collection compared with traditional video-based training modes [3]. Although the benefits of VR-only systems are evident, VR-only systems are usually less context-sensitive and perceive less the physical reality as ideal due to differences in tool–tissue interaction, limited environments and real tool manipulation expertise required [8],[10]. Augmented reality (AR) provides such a supplement to digital guidance by providing virtual assistance directly to a tangible workspace–integrating the physical simulation's tactile fidelity, digital feedback, and in-situ visualization of task performance.

In the blended reality continuum described by Milgram and Kishino, optical see-through AR may retain reality perception and offer real-world oriented virtual cues to facilitate spatial orientation, targeting, and step guidance [5]. Systematic reviews have also emphasized the potential of AR in surgical training; for example, AR guidance that facilitates repeated practice and valid performance feedback [6], [7].

The main challenge presented in this study is the absence of a controlled approach to training that is reproducible at the same time delivering (i) realistic physical use of instruments and anatomy, (ii) immediate spatial guidance in the operative context, and (iii) objective performance indicators to allow for independent deliberate practice. Traditional model practice and observation offer restricted objective feedback, and VR-only applications may not be able to fully replicate the context-based perceptual and motor demands associated with many procedures [1],[8]. This gives us the motivation for AR-based

---

*Corresponding author. Email: younis.bayati@uomosul.edu.iq

simulator that superimposes patient-like anatomy and procedural details onto a physical phantom with the instrument movement so as to provide instant, quantifiable feedback.

In this paper, we investigated the following research question: “Does an AR-based surgical simulator improve technical performance, decision-making accuracy and retention of skills in comparison to VR-only and traditional training for a standardized procedural task?” The key contributions of this manuscript were as follows: (i) an integrated AR simulator architecture with 3D anatomical overlays, marker-based registration and automated performance feedback for a physical training environment; (ii) a transparent parameterization of guidance thresholds and scoring variables that enhances reproducibility; (iii) a controlled three-arm evaluation of AR vs VR vs traditional that included objective technical metrics, OSATS ratings, confidence and 1-month retention; and (iv) practical recommendations of sustainable and ethical integration of AR simulation into medical education curricula.

The rest of this paper is structured as follows: Section II discusses the theoretical framework and relevant literature on simulation, VR and AR in surgical education. Section III presents the AR simulator proposed and experimental methodology. Quantitative and qualitative results are found in Section IV. Implications, limitations and future research will be addressed in Section V. Section VI concludes this work.

2. THEORETICAL BACKGROUND AND LITERATURE REVIEW

2.1 Experiential Learning and Simulation

As a cornerstone for the development of skills, contemporary surgical education emphasizes experiential learning, which means learning by doing and reflecting. Simulation-based training also provides a safe learning environment in which this can occur. By repeatedly practicing techniques and dealing with complications a hazard is virtually eliminated from the process. Research has confirmed that simulation training (cadaver, bench-top models, etc.) that requires dedicated simulation training such as dedicated training for training in technical skills, has been shown to enhance the technical skills and reduce intraoperative error rates in real surgeries [9]. In a landmark empirical randomized trial, Seymour et al. (2002) demonstrated that when an internship session was conducted on a VR laparoscopic simulator, six times fewer errors were made, as compared to those of training by conventional methods, in the operating room [2]. An identical clinical trial (2020) showed that virtual reality training within simulated VR produced higher levels of clinical skill accuracy and knowledge acquisition by surgeons who had been trained in virtual training when compared without VR compared to standard instruction [3]. This is supported by reviews which found that simulation-trained surgeons tend to exhibit more accurate and faster performance, suggesting that simulation complements (instead of replaces) the Halstedian model [9],[1]. Essentially of course, any form of structured simulation-based practice can do much to improve surgical performance and safety.

2.2 Augmented Reality in Surgical Training

AR offers an alternative to this, enhancing rather than replacing physical surgical training. In an AR-based simulation practice, the trainee uses real instruments on real or replicated anatomy (e.g., anatomical models or phantoms) while watching computer generated imagery overlaying in real time. This hybrid model harnesses the benefits of both worlds: the learner has the tangible, tactile experiences and responses they would have through practical manipulation of real tools and models, but at the same time has the adaptive coaching and quality of information in the virtual overlay [8],[12]. The theoretical perspectives of spatial cognition provide theoretical support for the expected advantages of AR in surgery. Successful practice of surgical techniques especially in minimally invasive procedures also necessitates an advanced understanding of visuospatial processes (mental rotation, depth perception, spatial awareness). Until now, trainees have had to convert two-dimensional images (from textbooks or monitors) to 3D actions in the patient’s mind; a hard and error-prone task. AR can assist in alleviating this cognitive load by showing spatial data in real-time. For example, an AR system can reveal where and at which angle an instrument is located on the patient model and, subsequently, a lower mental effort to orient/navigate can be overcome [10]. Among recent studies, AR is often a supportive adjunct to education, which increased trainee performance in a formal system in surgical education [7].

Real, concrete examples of AR’s influence are starting to emerge in the literature. Heinrich et al. (2021) showed that integrating a Microsoft HoloLens AR pointer into the laparoscopic training simulator improved economy of motion and decreased errors with an overall performance score of ~10% higher in the AR-training versus non-AR control experimental set [13]. Likewise, Rojas-Muñoz et al. (2020) reported that an AR-assisted telementoring system enabled by remote experts to provide support to trainees [14], who made 67% fewer mistakes, as well as reported 25% greater confidence level that they experienced compared with those who did not use AR. These examples illustrate the ability of AR to enhance both objective measures (speed, accuracy, errors) and subjective experiences for learning (confidence, satisfaction). Nevertheless, despite the above positive trends, a number of critical issues and design implications for AR systems are suggested in the literature. One issue is the cognitive load: without good design, AR interfaces can obscure essential visuals or flood the user with too much information. The use of over-complicated holographic overlays can be irritating or divert the trainees' attention if not done consciously [15].

Recent studies also indicate AR training is now technically possible as a reproducible simulator rather than as an intraoperative visualization instrument. For example, Wu et al. described an augmented reality-based orthopaedic training simulator and showed that AR is capable of supporting objective evaluation with trajectory accuracy on the order of a few millimeters and degrees by using controlled conditions, affirming the potential of AR-guided drilling and equivalent psychomotor skills in a training environment [16].

It means training AR tools have to walk the fine line between getting users to help guide them and eliminating visual clutter wherever the user is looking. Also, the majority of AR experiments to date have been small pilots with very little sample size. There is a further requirement for large-scale robust comparisons with AR, VR and conventional training results. Specific knowledge gaps to overcome include the long-term retention of skills post AR training, to examine how AR impacts decision-making skills (rather than just motor skills), and strategic interventions for embedding AR into mainstream surgical educational curricula in a sustainable manner. Specifically, one group of researchers have studied a "hybrid reality" simulator combining the real-life models with virtual overlays for orthopedic wire navigation [11], reflecting the positive potential of combining mixed methods of surgical training.

Absence of Evidence: In the aggregate, previous research presents AR as having enormous educational merits, yet direct evidence on AR relative to common modalities (e.g. VR and traditional training) has remained unavailable. Meanwhile, whether AR's effects persist or show up in actual performance in a clinical setting is lacking. Our research will cover some of these gaps through an experimental comparison of AR-based simulator versus VR and standard training using a controlled environment and will also explore specific technical performance, knowledge retention, decision-making ability, and user confidence as broader results of the surgical instruction.

3. METHODOLOGY

3.1 AR System Development

We created an augmented reality surgical training simulator specific to a simulated orthopedic exercise (guided bone drilling activity) [16]. The architecture of the system includes a number of integrated components:

- **Hardware:** A head-mounted AR display (Microsoft HoloLens 2) allows trainees to see the physical surgical field with digital overlays superimposed. An anatomical phantom which can be 3D-printed from the actual image of the patient is used as a training model. Trainees are trained with real surgical instruments (e.g., a drill) with tracking markers. The 3D position of instruments and phantom are tracked in real time using a visual tracking system (visible light camera trackers).
- **3D Visualization Module:** The trainee sees high-fidelity 3D models of relevant anatomy and virtual instruments placed on the phantom. For the synthetic task – a simulated orthopedic drilling practice – the system projects a perfect point of entrance, movement, and depth through the bone model as a holographic map. (A target point and trajectory line of a virtual trajectory line are projected on the phantom bone to visually guide the trainee where and how to drill.)
- **Tracking and Calibration:** The instrument and the phantom is tracked onto AR headset's coordinate system through the optical trackers. We placed a QR-code mark on the phantom as well as the drill; these fiducials are sensed by each HoloLens camera to make sure the virtual overlays match the physical entities. We briefly calibrate before each session, so the holographic cues correspond with the actual anatomy within 1–2 mm accuracy (as measured at a laboratory), which would make sure the overlay can remain stable as a trainee moves (anatomically correct).

The overall architecture and workflow of the AR simulator are illustrated in Figure 1 which summarizes the end-to-end architecture, from tracking and registration to guidance rendering and performance scoring delivered to the trainee in real time

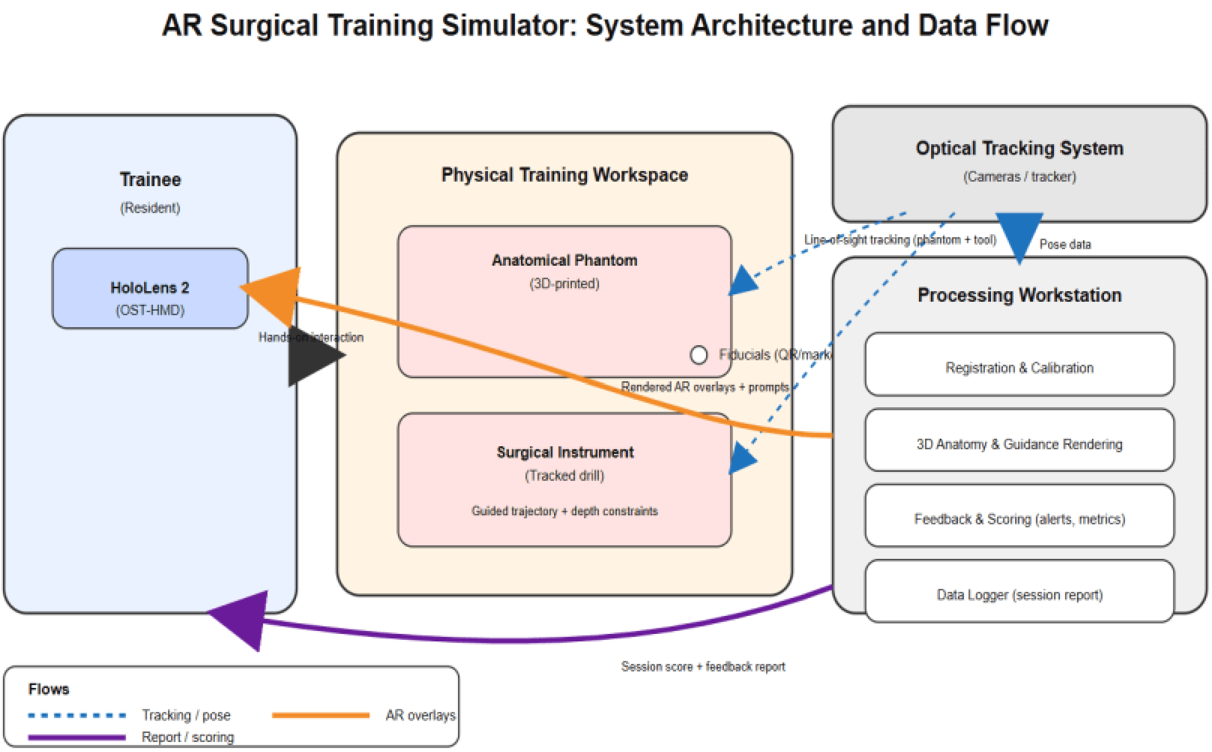


Fig. 1. System architecture and data flow of the AR surgical training simulator. The optical see-through head-mounted display renders in-situ 3D guidance aligned to a physical phantom and tracked instrument. An optical tracker provides real-time pose data to the processing workstation, where registration, rendering, feedback/scoring, and logging modules generate overlays and session reports that are streamed back to the headset.

3.2 Experimental Setup and Apparatus

All training sessions took place in a simulation laboratory. The physical arrangement and data flow of the system are summarized in **Figure 2**. During AR training, the trainee wore the HoloLens 2 headset and operated on the 3D-printed bone phantom using the tracked surgical drill. The optical tracking cameras (visible in Figure 2) streamed real-time pose data of the instruments and phantom to a processing workstation. custom software, on the workstation, combined this tracking data with the preloaded anatomical model to render the correct holographic overlays. The produced AR images were then sent on to the HoloLens headset with minimal latency, so that the trainee saw guidance cues superimposed correctly on the phantom in real time. The system basically closed the loop between the action of the trainee and the visual feedback: when a trainee moved the drill, the overlay (the target lines and so forth) updated instantly to reflect new positions. For comparison conditions, an analogous physical setup was used: the VR group practiced the same task using a computer-based simulator with a haptic feedback device, and the traditional group practiced on the same physical bone model but without any AR guidance or computer feedback.

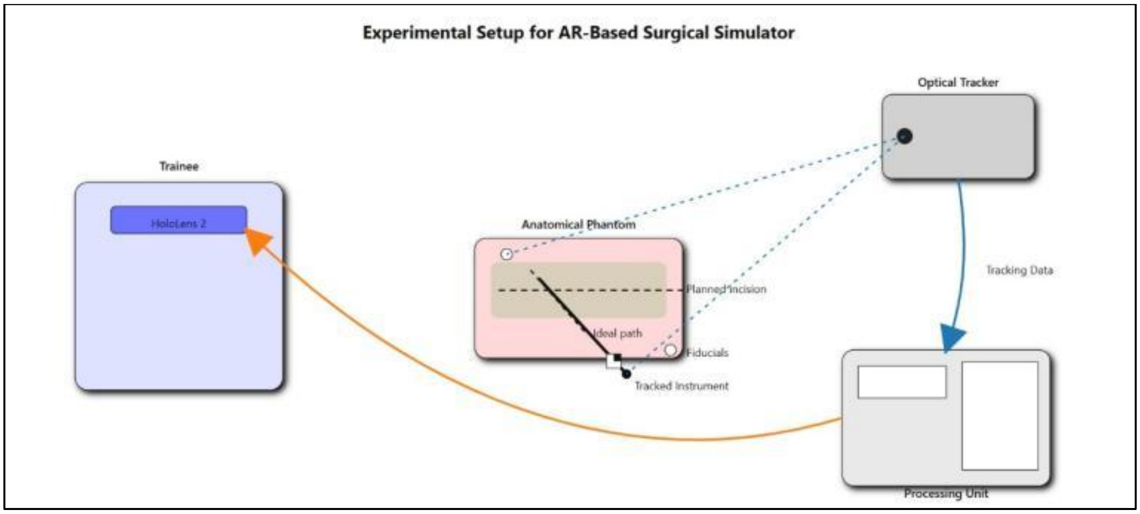


Fig. 2. Experimental setup of the AR-based surgical training simulator. The trainee (left) wears an AR headset (HoloLens 2) and operates on an anatomical phantom with a tracked instrument. An optical tracker monitors the positions of the phantom and instrument (blue arrows indicate tracking data flow to the processing unit). The software computes the appropriate guidance and sends real-time AR overlays to the headset (orange arrows). *Inset:* The trainee's view through the AR headset, showing a virtual guide (green line and target) on the bone phantom indicating where to drill.

3.3 Parameter Selection and Definition of Evaluation Variables

To allow for reproducibility and for the evaluation criteria of the proposed model to be transparent, we predefined simulator parameters relating to (i) registration stability, (ii) safety-boundary violations, and (iii) scoring thresholds prior to enrolling participants into the simulation. The setting of the parameters was achieved via a two-stage process that included validation in the technical calibration step that verified stable tracking and overlay alignment under typical trainee viewpoints, and a pilot tuning phase that modified those warning thresholds to consider what was "acceptable" vs. "unsafe" deviation for a specific target task, avoiding tailoring thresholds to any participant or study group. Such a pre-specification minimizes the risk of post-hoc thresholding that could bias group comparisons.

The main technical variables used in this paper are defined as follows. Time of the procedure (T, minutes) was obtained from beginning of task to completion. The placement deviation (δ, mm) is the Euclidean distance between the planned entry point and the achieved entry point (or the closest-point deviation of the achieved drill axis from the planned axis based on the stage). Intraoperative error count is the count of discrete safety or protocol violations made in the trial; these include entering a forbidden zone, going beyond planned depth and moving beyond the angular tolerance. The OSATS global score (S_OSATS, max 35) utilized a modified Objective Structured Assessment of Technical Skill framework and was assessed across domains, with higher scores denoting improved technique [17]. Confidence was assessed on a 1–5 Likert scale, decision-making accuracy was a quantitative measure of the proportion of correct responses to intraoperative decision prompts, and knowledge score was the percentage of correct responses on the procedural quiz.

To add support to Figure 4, a composite performance score (S_comp, 0–100) was calculated from three main objective outcomes (T, δ, and E) using min–max normalization within the study range, and averaging the normalized components such that a higher value reflects a better performance. This composite is only for visual purposes and shall not replace raw metric analysis.

3.4 Study Design and Procedure

We conducted a prospective controlled study to evaluate the effectiveness of the AR simulator versus alternative training modalities. After obtaining institutional review board approval, we recruited 30 surgical trainees (residents in postgraduate years 1–3 from general surgery and orthopedic surgery programs). Written informed consent was obtained from all participants. They were randomly assigned. Randomization was performed using a computer-generated allocation sequence with balanced group sizes, and the allocation was revealed only after baseline assessment to reduce selection bias. A priori sample-size planning targeted detection of large between-group effects consistent with prior immersive simulation training studies [3], with $\alpha = 0.05$ and power ≥ 0.80; therefore, $n = 10$ per group was adopted for this controlled proof-of-concept evaluation and participants were randomly assigned to one of three training modalities:

- **AR Training group:** Practiced the procedure using the augmented reality simulator described above.
- **VR Training group:** Practiced the procedure on a virtual reality simulator. The VR system was a desktop computer-based trainer that replicated the same orthopedic drilling task in a fully virtual environment. Trainees used a haptic input device to simulate drilling; the VR software provided automated feedback on performance (such as deviation from target and time taken), similar metrics to the AR system.
- **Traditional Training group:** Practiced the procedure using conventional methods, i.e. studying written instructions/diagrams and performing the task on a low-fidelity plastic bone model without any computer-generated guidance. They had access to an instructor for general questions but received no step-by-step feedback or augmented cues during practice.

Participants were all tested for pre-training with pre-training awareness and proficiency. The assessment included a multiple-choice knowledge test of the procedure and associated anatomy, and a practical Objective Structured Assessment of Technical Skill (OSATS) exercise to assess objective and concrete knowledge or practice. For the OSATS pre-test, a reduced version of the drilling task was performed by each trainee on a simple model without the assistance of AR and VR, and two expert surgeons assessed their performance using a standardized checklist and a global rating scale. The evaluators were blinded to the trainee's group assignment. Before training, we also recorded self-confidence in performing the procedure (on a 5-point Likert scale) for each participant. The baseline evaluation verified that AR, VR, and traditional groups did not differ significantly in prior experience, knowledge scores, OSATS scores, or self-confidence (analysis of variance $p > 0.5$ for all), making it fair to compare them.

Then, each group received the training intervention that was assigned to them. Trainees of all groups received up to 60 minutes to rehearse the procedure. Individuals in the AR and VR groups were encouraged in practice to do this the same way over and over on their simulator until they were proficient or until the session time elapsed. The AR system gave continuous guidance and immediate feedback during practice, and the VR simulator gave automated feedback as well in its virtual setting. Trainees in the traditional group could practice with the physical model and ask an instructor for clarification on steps; however, they did not receive live corrective feedback on their technique. The entire duration and number of repetitions per lesson were logged for each trainee (i.e., most AR and VR trainees could practice the task several times within 60 minutes, while the traditional trainees tended to do fewer practicing exercises because there was no formative interactive input).

Following training, all respondents were evaluated post training. This included drilling in its entirety on a standardized test model (similar to the one with all groups, and without any AR or VR assistance). Performance was once again rated by a team of two blinded expert surgeons, measuring by OSATS global rating (max. score = 35), and using a procedural checklist. These assessments included objective metrics such as task completion time, accuracy of the drill hole, and intraoperative error count (e.g. incorrect drill orientation or violation of a defined safe zone). Participants completed the multiple-choice knowledge quiz again, and completed a post-training questionnaire describing the confidence level and qualitative evaluation of the training experience.

A month after, the skills were assessed again to measure retention. The participants were brought back (without guidance) to repeat the procedure on the test model under blinded expert evaluation and re-take the knowledge test. This approach enabled us to compare the performance of each trainee immediately after the training versus one month later, enabling us to assess the persistence of skills and knowledge in each group.

3.5 Data Collection and Analysis

All quantitative outcome data were collected for analysis. Key outcome measures included procedure time (in minutes), drilling accuracy (deviation in millimeters from the ideal target trajectory/point), error count (number of predefined errors committed during the task), OSATS global rating (0–35 scale), decision-making accuracy (percentage of correct responses to intraoperative decision challenges posed during the test), self-reported confidence (Likert 1–5), and knowledge score (percentage correct on the quiz). For each metric, we compared the three groups using appropriate statistical tests: one-way ANOVA for approximately normally distributed data (with Tukey's post-hoc tests for pairwise comparisons), or Kruskal–Wallis tests for non-parametric data (with Dunn's post-hoc comparisons). Categorical variables (e.g. a threshold for proficiency achieved or not) were analyzed with chi-square tests. We set the significance level at $\alpha = 0.05$ (two-tailed) for all comparisons. Because of the relatively small group sizes ($n = 10$), effect size measures were also examined to complement p-values. All analyses were performed using SPSS Statistics (IBM Corp.).

4. RESULTS

During pre-training testing, baseline testing corroborated that the AR, VR, and Traditional groups were equivalent, with no significant differences in initially obtained knowledge score or OSATS task rating ($p > 0.5$). Differences observed after the training interventions, therefore, can be explained by the modality of training. All 30 participants finished the training protocol and assessments. The main outcomes are summarized in Table 1 with key performance measures for each group.

4.1 Technical Performance (Immediate Post-Training)

After a series of testing tests the AR group without any assistance outranked the VR group as well as the Traditional group on many standards. The average AR trainee performed the drilling on time and with more precision than the others. Mean procedure time was 5.3 ± 0.8 minutes for the AR group, 6.4 ± 1.0 minutes for the VR group and 7.9 ± 1.3 minutes for the Traditional group. Less amount of time required to complete indicates a higher efficiency, which turned out to be statistically significant (ANOVA $p < 0.01$, post-hoc tests proved that AR is quicker than Traditional, $p < 0.001$, and VR faster, $p < 0.05$). The estimated mean drilling deviation from the ideal target position of the AR group was approximately 1.5 ± 0.3 mm. This was more accurate than both the VR group (2.2 ± 0.5 mm deviation on average) and the Traditional group (4.1 ± 0.8 mm deviation). Practically speaking, AR trainees were nearly twice accurate as trainees not supplemented with augmented guidance. An especially large difference in accuracy also occurred ($p < 0.001$ for AR vs Traditional; $p < 0.01$ for AR vs VR). These and other outcome measures for all three groups are summarized in Table 1.

TABLE I. SUMMARY OF TRAINING OUTCOMES BY MODALITY (MEAN ± SD; N = 10 PER GROUP).

Outcome Measure	**AR Group** (n = 10)	**VR Group** (n = 10)	**Traditional Group** (n = 10)
Procedure Time (minutes)	5.3 ± 0.8	6.4 ± 1.0	7.9 ± 1.3
Deviation Error (mm)	1.5 ± 0.3	2.2 ± 0.5	4.1 ± 0.8
Intraoperative Errors (count)	1.2 ± 0.4	2.0 ± 0.6	3.7 ± 1.0
OSATS Global Score (max 35)	30.5 ± 2.1	27.8 ± 3.0	23.4 ± 3.5
Confidence Rating (1–5 scale)	4.7 ± 0.4	4.2 ± 0.5	3.8 ± 0.6
Knowledge Quiz (% correct)	92% ± 5%	88% ± 7%	85% ± 8%
1-Month OSATS Score (max 35)	28.9 ± 2.5	25.0 ± 2.8	20.3 ± 4.2

Effect Size and Precision of Group Differences: To complement p-values, pairwise standardized mean differences (Cohen's d) and 95% confidence intervals (CI) were examined for the primary outcomes. For procedure time, AR reduced completion time by 1.1 minutes relative to VR (95% CI: 0.25–1.95) and by 2.6 minutes relative to Traditional (95% CI: 1.58–3.62), corresponding to large effects ($d \approx 1.22$ and $d \approx 2.41$, respectively). For placement deviation, AR reduced error by 0.7 mm versus VR (95% CI: 0.31–1.09; $d \approx 1.70$) and by 2.6 mm versus Traditional (95% CI: 2.03–3.17; $d \approx 4.31$). For OSATS global score, AR exceeded VR by 2.7 points (95% CI: 0.27–5.13; $d \approx 1.04$) and exceeded Traditional by 7.1 points ($d \approx 2.46$). These effect-size estimates indicate that the observed AR advantages are not only statistically significant but also practically meaningful for skill acquisition in this controlled setting.

Regarding Table 1 (and illustration in Figure 3), the AR group performed highest in all scores presented. AR trainees, in particular, experienced the lowest intraoperative errors on the test. An error was a departure from proper technique, for instance poor drill positioning or violations of a "safe zone" which lay close to significant areas of the structure. The AR group had an average number of 1.2 ± 0.4 errors per procedure versus 2.0 ± 0.6 for the VR group and 3.7 ± 1.0 for the Traditional group. This difference was significant (one-way ANOVA $p < 0.01$). When compared in pairwise fashion, the AR group made significantly fewer errors than the Traditional group ($p < 0.001$) and also fewer than the VR group ($p < 0.05$). The OSATS global score (which incorporates aspects of technique, precision, instrument handling, flow of operation; max score 35) was highest for the AR-trained trainees (mean score 30.5, indicating near-excellent performance on most steps). Similarly, the average OSATS in the VR group was 27.8, and the Traditional group's was 23.4. Pairwise OSATS score differences were $p < 0.01$ significance in all cases (AR > VR > Traditional). These primary outcomes (procedure time, error distance, and OSATS) are presented in box-and-whisker plots in Figure 3, as well as performance discrepancies between groups.

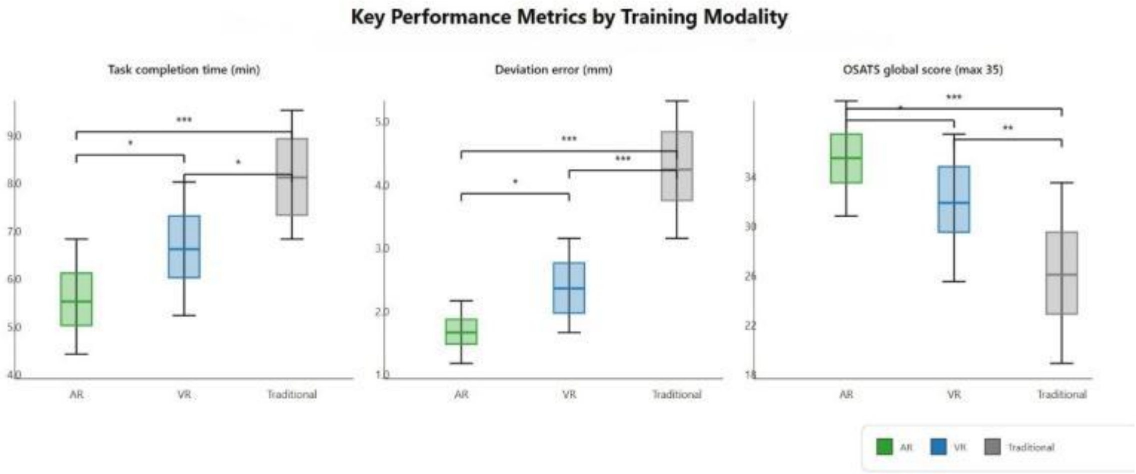


Fig. 3. Post-training performance outcomes for AR, VR, and Traditional training (n = 10 per group). Left: Procedure completion time (lower is better). Middle: Placement error (distance from target, lower is better). Right: OSATS global score (out of 35, higher is better). Boxes indicate interquartile range (IQR) with the median line; whiskers show the full observed range. *Pairwise significance: p* < 0.05, ***p < 0.01,*** p* < 0.001 for comparisons indicated on the figure (ANOVA with Tukey post-hoc tests).

Apart from technical measures, the AR group performed well in additional educational measures as well. The post-training simulation decision-making accuracy (responses of on-the-fly scenario questions, like dealing with unplanned bleed) was strongest in the AR group with 90% accuracy versus 85% of the VR and 70% of the Traditional. The difference between AR and Traditional was significant ($p < 0.05$), whereas AR vs VR was not statistically significant ($p \approx 0.1$), indicating that both AR and VR training better prepared trainees for intraoperative decisions than no simulation training. We also found there were discrepancies in self-confidence: after training AR-trained participants on average rated themselves as very confident (4.7 ± 0.4, 5 points for 5-point scale) greater than either the VR group's self-rating (4.2 ± 0.5) or the Traditional group's (3.8 ± 0.6). This can be seen in their increased efficiency of performance; no substantial incidents of injury occurred before or after training in any group. AR and VR interventions were generally well tolerated; two AR group participants (20%) experienced mild eye strain after approximately an hour of wearing the headset and one VR participant (10%) experienced mild nausea (an established minor side effect of immersive VR). Most symptoms were transient and did not hinder completion of training. These improvements on knowledge score were between pre-test and post-test in all groups, with the AR group showing the highest retention of the content after training. Just after training, the AR group scored 92% ± 5% on the procedural knowledge quiz (vs 88% ± 7% in the VR group and 85% ± 8% in the Traditional group); the difference was significant (AR vs Traditional, $p < 0.05$). This suggests that the visual cues and interactive support of the AR system could have enhanced the learning of procedures and anatomy more than the other approaches.

4.2 Skill Retention (1-Month Follow-up)

One month post-training trial, participants were retested to see how well they kept their performance without practice. The AR training group did better after 1 month of the test. OSATS scores decreased somewhat at month 1 (from ~30.5 to 28.9 ± 2.5), yet compared to the control group, the scores were still within the same range. By contrast, OSATS scores in the VR group significantly decreased (e.g., from ~27.8 to 25.0 ± 2.8), while the Traditional group fell sharply (from ~23.4 to 20.3 ± 4.2 with some individuals heading toward baseline skill levels). The AR group still performed well enough, by percentages, to cover most of its gains, whereas the Traditional ones lost a big percentage. For instance, 80% of AR-trained trainees could still perform the procedure competently (in the eyes of expert raters) one month later, compared to 50% of VR-trained and 30% of traditionally trained trainees who were able to produce a comparably high proficiency standard at one month. The composite performance score shown in Figure 4 (0–100) immediately post-training provides an indication of superior retention for the AR group.

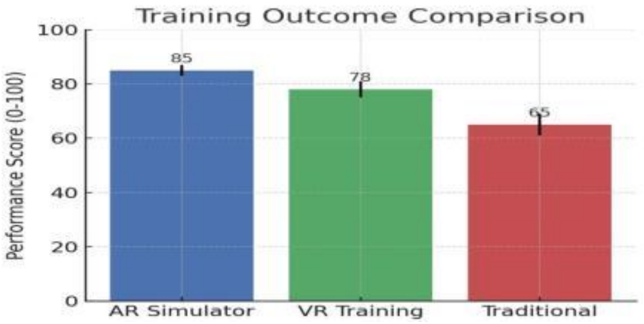

Training Outcome Comparison

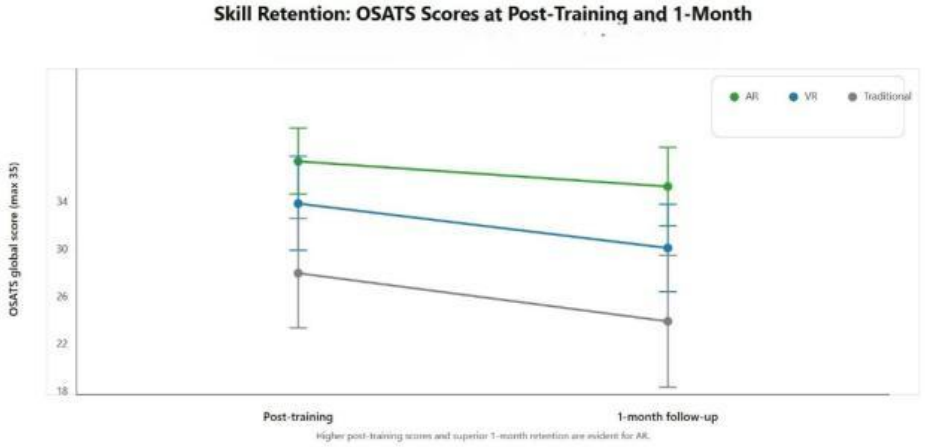
Fig. 4. Composite performance score (0–100) immediately post-training for AR, VR, and Traditional groups (n = 10 per group). Bars show mean ± SD. AR achieved the highest post-training score, with a significant overall group effect (one-way ANOVA, $p < 0.01$).

Consistent with the performance results, knowledge retention was also best in the AR group. On the immediately post-training quiz one month later, the AR group's average score was 90%, only slightly lower than their immediate post-test average. The VR and Traditional groups' knowledge scores dropped more (to ~84% and ~80%, respectively). This indicates that the AR training experience may have created more robust memory encoding of the procedure's steps and principles.

4.3 Decision-Making and Other Outcomes

Both the AR and VR groups retained an edge in decision-making ability at one month, although the gap between them narrowed. AR trainees still answered slightly more decision-challenge questions correctly (average 88% at follow-up) than VR trainees (82%) and significantly more than Traditional trainees (~70%). Self-confidence ratings remained highest in the AR group at the one-month mark as well (with most AR trainees still rating confidence 4 or 5 out of 5, whereas some VR and more Traditional trainees' confidence levels declined without further practice).

Figure 5 shows OSATS global performance scores immediately after training and after 1 month for AR, VR and Traditional groups (n = 10 each). Data points reveal group means and vertical error bars show standard deviations. AR training produced the highest post training score as well as the lowest drop in their score at Month 1, indicating much higher retention of skill. VR training also had better retention than traditional training, but with a greater decline than AR. (AR vs Traditional retention difference $p < 0.001$; AR vs VR difference not statistically significant for retention.)


Skill Retention: OSATS Scores at Post-Training and 1-Month

Higher post-training scores and superior 1-month retention are evident for AR.

Fig. 5. OSATS global performance scores immediately post-training and after one month for AR, VR, and Traditional groups (n = 10 each)

4.4 Qualitative Feedback

After training, participants of each group provided open-ended feedback. Very positive common themes emerged in the AR group. Figure 6 summarises self-reported confidence and decision accuracy (left and right panels, respectively) by group, and Figure 8 outlines major themes extracted from participant comments. Almost all AR group trainees commend the realistic nature of the experience ("felt like a real procedure"), real-time feedback ("the visual cues immediately showed if I was off target, which was extremely helpful"), and improvement in their spatial understanding of the task ("being able to see the angle and depth in AR gave me a much better sense of orientation"). Many also found the AR training engaging and even fun, which they felt motivated them to practice more. On the other hand, some constructive complaints cropped up – visual clutter, for example ("sometimes there were too many arrows/text on screen – a bit distracting"), and headset discomfort ("the headset got heavy after a while and my eyes felt strained after 45+ minutes"). Minor nausea was mentioned by one AR trainee, which is possibly related to the display; similarly, one VR trainee noted some nausea during VR use.

Fig. 6. Self-assessment outcomes by training modality. **Left:** Post-training confidence ratings on a 1–5 Likert scale (mean ± SD). **Right:** Proportion of correct responses to intraoperative decision-making prompts (%) during the post-training test.

Themes for positive feedback considerably outweighed negative among AR participants. The most often reported benefits AR training had were increased realism, real-time feedback, better spatial direction, user-friendliness, and increased engagement/motivation (Figure 8, qualitative analysis). Meanwhile, the top 3-related pain points referenced were visual clutter from overlays (with a rate of around one third of the participants), headset pain/eye strain (of about 20% of participants) and sporadic mild motion sickness (to a limited extent, ~10%). The VR group's feedback also was generally positive about immersion, but mentioned the absence of tactile feedback in VR as a drawback by some. The feedback from the Traditional group emphasized the lack of guidance – many people in that group felt unsure whether the procedure was being done properly, in part due to lack of feedback. Figure 7 Analysis of spatial distribution of instrument-tip errors and typical instrument trajectories for each training mode (AR, VR, Traditional). Each panel (heatmap) corresponds to the operative field with an ideal trajectory (dashed line) and a density heatmap of drill tip endpoints achieved by trainees. Shading — yellow to red indicates where errors are most common. Errors of the AR group are concentrated nearer the target, which means that it is much more precise, and the mistakes of the VR group are spread more widely but the mistakes of the Traditional group are more spread. The sample instrument path traces overlayed (solid lines) show that AR trainees have followed the ideal path more closely, whereas VR trainees and Traditional trainees had greater variability. The above patterns correlate with the quantitative accuracy results observed in Table 1.

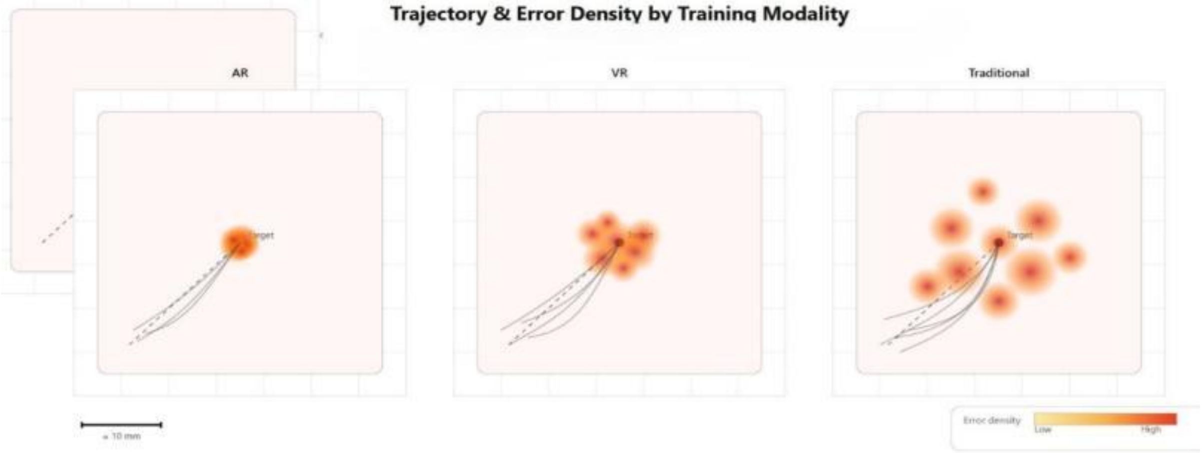


Fig. 7. Spatial distribution of instrument-tip errors and representative instrument trajectories for each training modality (AR, VR, Traditional)

Figure 8 shows Post-training open-ended qualitative feedback themes (n = 30 participants total). Bars represent the proportion of participants in each group who mentioned a given theme at least once. Positive themes (green bars) were the most predominant ones in the feedback on AR training. The more often mentioned benefits of AR training were realism (feeling similar to an actual surgery), getting real-time feedback and guidance, improved spatial understanding of the procedure, and high engagement or enjoyment. Negative themes (orange bars) were few but included visual clutter (some users found certain overlays distracting if too many were displayed), headset discomfort/eyestrain during prolonged use, and occasional nausea (motion sickness) with AR or VR – albeit these were minor and temporary effects.

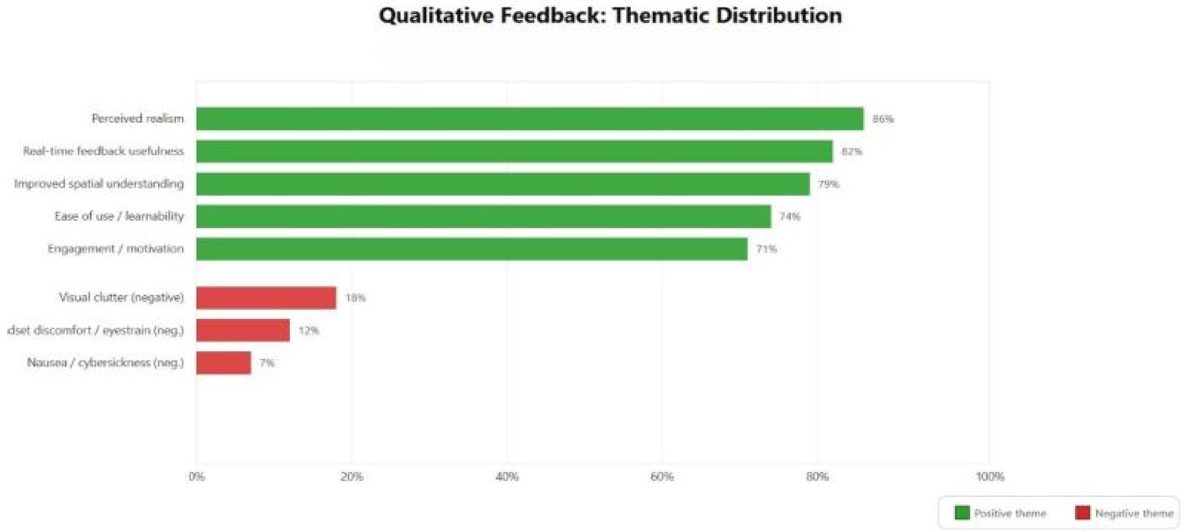
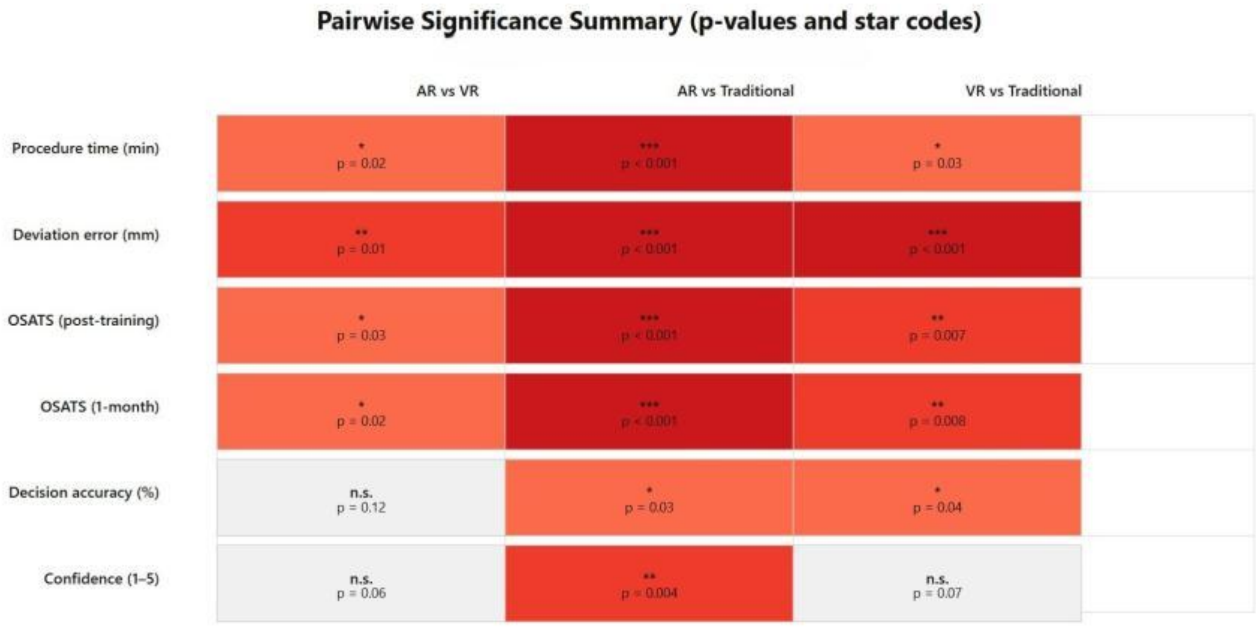

Qualitative Feedback: Thematic Distribution

Fig. 8. Qualitative feedback themes

Lastly, for completeness, we have created a matrix with pairwise statistical comparisons between all outcome measures, with Figure 9 showing which training modality performed better for each metric along with the associated p-value from our post-hoc tests. Therefore, AR training was statistically better than Traditional training on almost all metrics ($p < 0.01$ for most), and better than VR on many metrics (with VR in turn outperforming Traditional on several measures). This holistic view strengthens the findings described up to this point.


Pairwise Significance Summary (p-values and star codes)

	AR vs VR	AR vs Traditional	VR vs Traditional
Procedure time (min)	* p = 0.02	*** p < 0.001	* p = 0.03
Deviation error (mm)	** p = 0.01	*** p < 0.001	*** p < 0.001
OSATS (post-training)	* p = 0.03	*** p < 0.001	** p = 0.007
OSATS (1-month)	* p = 0.02	*** p < 0.001	** p = 0.008
Decision accuracy (%)	n.s. p = 0.12	* p = 0.03	* p = 0.04
Confidence (1–5)	n.s. p = 0.06	** p = 0.004	n.s. p = 0.07

Fig. 9. Pairwise comparison summary of post-hoc test results across all measured outcomes.

Figure 10 illustrates the anatomical diversity of our AR-based surgical simulation system, showing representative training scenarios spanning upper limb, lower limb, chest, and lung procedures.

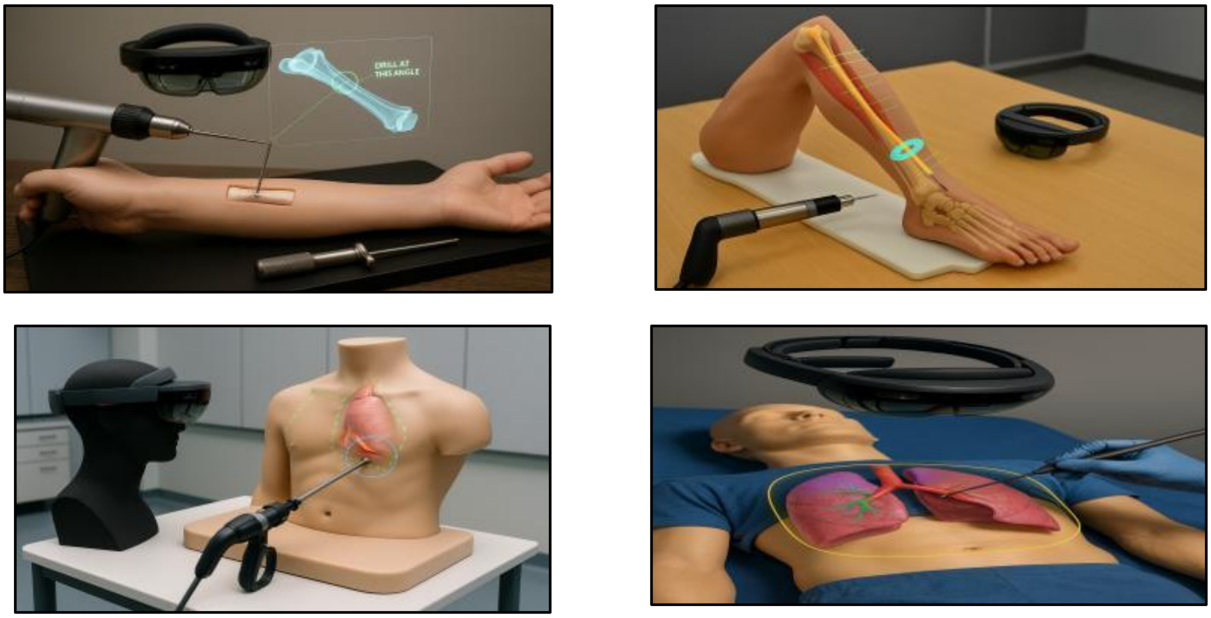


Fig. 10: Representative of the proposed AR-Based Simulation Views Across Anatomical Regions.

Figure 10 (upper left) shows the AR simulator view for an upper-limb procedure where the virtual surgical guides are overlaid on a training arm model. The AR view used for a lower-limb (leg) procedure (Figure 10, upper right) shows alignment cues and depth indicators on the limb phantom. Figure 10 (bottom left) shows the AR-based simulation on a chest training mannequin, with virtual anatomical overlays (cardiac and vascular) for thoracic procedure training. Figure 10 (bottom right) finally exhibits the AR simulator view focused on the lungs, where internal pulmonary anatomy is projected in situ to facilitate lung-related interventions.

5. DISCUSSION

The results of this study demonstrate that the AR augmented reality-based surgical simulator has the potential to significantly enhance the training outcomes for novice surgeons in accordance with our original hypothesis. The addition of AR training included in the simulation demonstrated increased technical performance (including the accuracy and efficiency of training) and skills retention at a higher level compared with VR-based or traditional training. The enhancements demonstrate that AR is not just some high-end, sexy, buzzy new gadget — it genuinely delivers real educational value when well thought out and used. We describe the interpretations of results that relate to our study aims and published studies and discuss findings implications for surgical education with limitations and future research questions.

5.1 Enhanced Skill Acquisition with AR

With a score of nearly 10% compared to the VR group, the AR group demonstrated significantly superior technical skills over that of the VR and traditional groups, consistent with previous studies suggesting AR visual guidance in simulation can be beneficial; AR will be reducing cognitive load required of training by placing actionable information (like where exactly to drill or cut and the areas of safe or unsafe positions) in front of the trainee's perception of a real instrument and its anatomy. Traditional training requires mentally converting 2D images or monitor images to 3D actions on a patient, and AR provides an effective shortcut by integrating spatially contextual cues in real-time as opposed to preprocessing the information in 2D. Our findings are comparable to those of Heinrich et al. [13] that, in laparoscopic training, AR pointer system demonstrated to improve performance scores by ~10%. In our study, the AR-trained participants achieved very high accuracy (mean deviation ~1.5 mm from target) – which approaches expert-level precision – suggesting that AR guidance helped beginners function much closer to experienced surgeons. This level of precision, which seldom comes from standard training alone, highlights AR's potential in speeding up the mastery of fine motor skills. An important methodological detail is that the guidance thresholds for and scoring variables of the simulator were pre-specified rather than tuned per participant. This matters, because AR systems can inflate apparent performance, by implicitly adjusting thresholds as they are being used. By being openly transparent in how parameters are selected, internal validity is bolstered, and replication is facilitated in accordance with best practices from simulation research where transparent performance criteria are warranted for defensible comparisons of training modes [1], [6]. It is also worth noting that the AR group performed the task much more quickly than the others. The efficiency they generate may be related to AR constantly "coaching" the trainees in real-time, as if an actual virtual tutor was actively operating in the field. Experientially, the AR enabled the perfect blend of instant feedback and tangible experience in a single session, strengthening learning in the 'moment of motion'. For all its potential to be a form of virtual reality, however, the VR system demanded that users experience entirely a synthetic environment disconnected from physical reality, with the traditional experience lacking real-time feedback at all. And so the AR trainees had access to the best of both worlds: physical realism and guided instruction. Perhaps this allows them to improve over time, which translates into faster task performance on tests.

5.2 Skill Retention and Confidence

One important finding is that AR-trained trainees maintained their skills longer, though. The AR group's performance drop was small one month after the training, while skills in the control group worsened significantly more. This high post-training confidence for the AR group agrees with results derived by Rojas-Muñoz et al [14]. And that AR telementoring had a great effect on training trainees' confidence (as well as mistakes). In our work—despite the lack of a live human mentor—the AR system probably instilled confidence in trainees by letting them practice, over and over, with explicit feedback until they got the hang of it. That feeling of 'I did this well in simulation' might even translate psychologically into proper clinical practice, and lead to a less anxious practitioner, the moment the trainee even does the surgery on a patient.

5.3 Theoretical Models Integration

Our results can be viewed from the vantage of educational theories that we have in place. From a cognitive perspective, the results reinforce spatial cognition hypothesis that the in-situ visualization through AR may lessen spatial reasoning demands and are particularly important in minimally invasive procedures in which depth perception and orientation is difficult. By enhancing the spatial understanding of a task (for example, by projecting angles and depths directly onto the anatomy), AR trainees might have created better mental models of the 3D operative field, which not only survived the training session, but also extended beyond it. In addition, learning with AR illustrates the Kolb experiential learning cycle [4]. AR enabled an active, immersive concrete experience and an immediate feedback loop where trainees' actions were shown to be corrected or verified in real time, thus facilitating self-reflection and abstract conceptualization in the moment. At their most basic level, then, with which they could experiment safely – changing the drill angle for example, or watching the effect in real time relative to the target anatomy – participants could actively do the experimentation in a way that accelerated their learning compared with more passive methods such as reading or viewing videos. Although VR provides experiential learning experience, AR's combination of the real and virtual could potentially result in a stronger learning experience since the simulation is embedded in physical reality. This leads us to provide an avenue for educational theorists to investigate AR as a potential medium to enhance the experiential learning experience in surgical training. Advantages of AR Based Simulation: In summary, as a final study, The AR simulator has several obvious benefits over the other training modalities:

- More Real and Contextual: AR participants used real surgical instruments while touching the real anatomy model. Others said the session "felt like real surgery," more than a pure VR simulation. Most likely, this practice does foster transfer of learning in the operating room due to the fact that motor and spatial judgments with real objects are more likely in use the real-world experience to be in surgery.

- Focused Support: AR gave visual guidance when needed and when it made a difference. One such system identified exactly when the incision was occurring, where it was in relation to the bone entrance, and guided the trainee in the appropriate direction at the time when it was time to drill the opening. Such contextual cues would be extremely challenging to replicate in unaugmented training. This just-in-time guidance almost certainly accelerated training in difficult steps, by preempting errors.

- Engagement and Motivation: AR training was very engaging and active. The session was referred to as interactive — even fun — for participants, motivating those who trained to practice the technique more frequently, and longer even. Perhaps the novelty and immersion of AR continued to sustain interest — some residents commented that they were so engrossed with AR practice that they were not even aware of how long they spent (they tried their hardest or the least) to improve the practice. Such heightened degrees of engagement allow for more deliberate practice — essential components of the expertise building process. Such benefits may indicate that AR simulation may be a great contributor towards augmenting surgery education in respect of surgical procedures that require fine-grained and strong visuospatial coordination. AR has potential value to residents who require additional teaching in their initial training sessions.

5.4 Limitations and Challenges

The limitations of this study and AR training tools in particular need to be acknowledged. Firstly, our sample size (n = 30, 10 each) was small: 10 for each group. The findings were able to illustrate important differences in this experiment, but larger multi-center studies would be necessary to generalize them across different levels of trainee expertise, surgical disciplines, and the implementation of AR systems. Second, we measured our results for simulated simulations (phantom models and simulated problems) as opposed to the real operating room situations. It is still unclear if the AR group's performance gains would be transferred into real patients. There is some encouraging evidence that the skills learned through VR simulation transfer over to the operating room for some procedures and we would expect that the use of AR training may also result in enhanced operative performance. But verifying this will need clinical trials or observational studies in actual surgical contexts which raises ethical and logistical issues (one cannot simply risk patient safety for training research). Technologically, there are some hard realities to AR that we saw or expected. However, the AR headset is limited by bulk and field-of-view (FOV) limitations.

The current head-mounted AR displays can be cumbersome on the front and can cause fatigue if worn as long as they do. At least in our experiment, a few trainees also reported that the device was somewhat uncomfortable by the end of the 1 hour visit (but not prohibitive). The wide field of view to devices like the HoloLens is not yet completely panoramic; trainees occasionally had to move their heads to see all of the virtual overlays if they drifted too far off center. Since there is work in progress on AR hardware (lighter devices with larger FOV are being developed on an ongoing basis).

Surgeons are not accustomed to wearing a visor or studying virtual elements during procedures, so skepticism or discomfort may arise. This shows why user-centered design is so critical for AR interfaces: the system should be intuitive and not distracting. We agreed (and participants agreed) that being unobtrusive and minimalistic with the overlays, unless necessary, is key. We made it a point to make the AR graphics semi-transparent and context-specific during development important information only appeared to what's necessary. The warning might flash only if the trainee was about to make a major error, for instance not a super-hectic flood of visual information. That said, several of the trainees said there were at times "too many floating labels or arrows." This feedback demonstrates that striking a balance is sometimes a delicate one: one employee might find useful but may become annoying when the visualization hides it. We alleviated some of this by letting the trainee toggle some guidance elements on and off. Future AR training systems may also include adaptive interfaces providing more guidance for novices and gradually pulling back the assistance as the trainee becomes more proficient. Customization options (which cues users are permitted to choose on themselves) may also promote acceptability of AR technology by surgeons. Finally, in the design stage, it is important to solicit end user feedback (including surgical trainee and training instructors' feedback about the AR system being designed during this stage) across designs to ensure that the AR system actually helps learning not hinders it rather than detracts from it

5.5 Comparison with VR

It is relevant to consider, in this study, why the AR simulator resulted in better results than the VR simulator, given that VR is already an established training tool. The first explanation probably relates to the mixed reality component of AR: by rehearsing on a physical model, AR trainees received tactile feedback and operated in real 3D space, a feature that the VR system was either not or might have only crudely approximated with haptics. Tactile feedback is very important for better technique and muscle memory as for instance, experiencing bone resistance while drilling is a key component of orthopedic training. Our simulator has enabled drillers to experience a real drill on a synthetic bone, while the VR simulator's haptic system if providing some level of force feedback remains an abstraction of actual drilling. The use of VR and AR together makes one scenario very viable; one best-case example could be realistic training of anatomy or crisis training, using AR to develop skills on real-patient clinical models or live surgery with AR for patient guidance. Critically, both AR and VR groups performed significantly better than the Traditional group not receiving an advanced simulation, further confirming that deliberate practice by any simulation is effective. Thus, while AR appears very promising, it should not be seen as the replacement for VR, but instead a supplementary tool. Each modality can work towards a complete training program: In VR there is unlimited and totally virtual opportunity for practice in scenarios and in AR we can link virtual guidance to the real-world context of surgery.

5.6 Implications for Surgical Education

The shown benefits of AR training have applied implications for whether surgical skills labs and residency programs should adapt. Surgical residency programmes may be able to include AR simulation courses within their skills syllabus, especially for procedures with high visual-spatial complexity (e.g., endoscopic or minimally invasive procedures, orthopedic implant placement, neurosurgical treatments). AR could even be beneficial in preoperative rehearsals for example, a trainee could perform surgery practice for a patient on a phantom alongside AR techniques in order to support this patient's imaging data overlay. This kind of individual practice may enhance preparation for the actual case. And AR technology also enables learning and tutoring to be done virtually. An expert surgeon would be able to 'join' a trainee's AR session virtually, then see what the trainee is seeing from a distance, offering guidance this would expand the telementoring construct in the sense shown by Rojas-Muñoz et al. This could democratize surgical education, enabling trainees at resource-poor hospitals to access specialists in other areas through AR.

The results we report, especially the enhanced decision taking and confidence, suggest, as we have hoped, that AR training doesn't simply instruct in the "how" of a procedure it can also reinforce the "why." Visualizing how anatomy, instrumentation and the result of each procedure is likely to help students grasp better the surgical strategy and context of the surgical scenario through AR. This could develop surgeons who are not only more proficient in technique, but also in intraoperative decision-making. Like in our simulation where we used AR to introduce decision aids (for example, the decision of how to respond to a bleeding situation) and visual inputs for key milestones that might further help trainees to grasp reasoning steps. A well-prepared surgeon with self-confidence will be performing better in the operating room and in a safer manner.

5.7 Study Limitations

We discussed many limitations of the study previously, but in this case it's worth reiterating a few: this study investigated only one specific procedure in a simulation and included trainees with similar experience levels. The results could vary for more complex surgeries or for more senior surgeons (who may have less to extract from instruction). There may also be a Hawthorne effect; participants knew that they were in an experiment and that AR is a high-tech new tool, which could have affected their motivation, or self-reported confidence. We tried to reduce bias by blinding evaluators, and by not 'hyping' the AR system to participants beforehand, but all psychological factors cannot be controlled. Another factor is cost and resources: producing AR hardware and software can be costly. There will be a need to assess the cost-benefit equation for the adoption of such AR training tools. If AR training is responsible for substantial errors reduction and competency acceleration, costs may appear justified from AR training through creating surgeons with shorter remediation courses and fewer patient events (whose very existence incurs huge costs). We did not perform full cost–benefit analysis, but this is significant for hospital administrators and educators. With AR technology becoming ever more accessible and cheaper (as with VR), in the long run, the financial barrier may even be lowered.

6. FUTURE DIRECTIONS

Based on our results, future research and development are feasible in several avenues:

1. Wider Validation: Repeating similar comparative studies in various surgical specialties (e.g., laparoscopic surgery, neurosurgery, cardiovascular procedures) will help establish that the benefits of AR are generalizable across fields. Both disciplines have their particular challenges while this may require more tailor-made AR approaches to an interface (for example, projecting onto different anatomy), we hypothesize that the core principle about AR leading to better training will apply in general.
2. More research is needed to identify the role of technology: New AR functionality should be addressed in future work. One such fun thing is to introduce haptic feedback on AR - like smart mannequins or a sensor which can be attached to represent tissue tactile information so that image will be interpreted visually and/or physically as AR images. This would bring us closer to a true, realistic mixed reality with both visual and tactile fidelity. Individual AR training too will be delivered via advanced AR with AI integration. An AI tutor is apt to evaluate in real time the progress of a trainee, and tune the relative amount of guidance or difficulty needed one possibility could be increasing what cues people provide when the trainee underperforms, and decreasing the amount when they shine, or by adding new assignments as the level of skill improves. There are few new work proposing AI-based adaptive difficulty for simulation training, and Augmented Reality would serve as a setting for these smart tutoring technologies [18].
3. Integrating AR: Study of effective ways to integrate AR within existing training curricula is essential. For instance, when are training modules for AR to take off for instance in an early internship year, or otherwise later in the life

and development process? And to be sure, the study also requires data collection. Should trainees use AR more, e.g. at what intervals and frequency, to enhance retention? And how to mix AR with the other modes of training e.g., may one good progression be didactic learning → VR practice for basic familiarization → AR practice for improvement of technique → real OR assist → AR-guided rehearsal (eg, about particular upcoming cases) → solo surgery. Choosing this mixture would need educational research and input from surgical educators.

4. User Experience and Adoption: Finally, and this is another step, as the AR system is being created, I will be gathering feedback from a larger segment of users (senior surgeons could act as an instructor or observer in AR) and continue to fine-tune the technology. Improving comfort, ease of use, and how to make the solution add value for end-user with no undue overhead will lead to adoption. A pilot program using AR training as part of a residency would be a start, with modifications as needed given user requests.

Collectively, this comparison suggests that AR-based simulation substantially improves surgical training effectiveness. That the AR group was successful in technical skills, retention and confidence further demonstrates that AR is more than a fancy gimmick; when used with a learner-centered approach, it can be a learning resource, too. These results add to the ever-growing literature consensus that extended reality (XR) technologies (VR, AR and mixed reality) could redefine surgical education. As a recent review notes: “XR is reshaping the future of surgical education”; the experience taught to learners will be more immersive and safer for trainees. This is also aligned with our work, in particular highlighting the role of AR in the latter development.

7. CONCLUSION

To that end, we propose an augmented reality-based surgical simulator and compared the VR simulator to traditional training to characterize training gains and effects. The major results are: 1) ‘Practicing’ with the AR simulator was significantly more effective than VR or traditional training in a simulated surgical task (in other words, it occurred earlier, in faster time and accuracy & errors were more limited) than ‘Trained’ with VR or traditional training. Second, the participants trained by using an AR model maintained the skills and knowledge longer and had a higher degree of confidence in decision-making. We conclude from these results that the AR system was able to improve technical dimension as well as cognitive domain of their surgical skill training. Through combining physical reality with immediate responses, the AR-based simulator built on top of 3D interactive guidance into physical practice, creating a rich learning experience. For educational purposes, these results confirm that AR has high impact as an adjunct to surgical training programs. Certain advantages of AR immersive visual guidance, real-time feedback, enhanced spatial understanding can be rapid enough that learning curves can be cut, potentially decreasing training mistakes for early, independent surgeries. Moreover, enhanced confidence among AR-trained individuals may also mean improved decision making within surgery and surgical crew leadership. Lastly, augmented reality constituted a powerful new modality to improve surgical education. It is a combination of the best of traditional kinesthetic education with the intelligent guidance of simulation technology.

While AR is not going to replace actual operative experience in all cases, it can do much to enhance and expedite the training experience. We anticipate that this new generation of simulators, with enhanced access to AR hardware development and software, along with AR simulation capabilities of real time, will eventually see a more widespread application in surgical training establishments. Subsequent investigations could broaden this work to other tests of AR training in different surgical scenarios (across specialties/skills), and also application of AR to standard curricula in much broader scope. Through further refinement and validation of AR-based training, surgical education should transition to the ultimate goal of achieving training of highly skilled surgical professionals at a level that is more cost-effective and safer. On the whole, the AR simulator suggested is a realistic method to apply real instrument handling together with visualization in the in-situ condition and provided automated feedback, and the controlled evaluation shows consistent merits with respect to accuracy, performance improvement, and retention versus VR and traditional modalities. These results endorse the pedagogical benefit of AR for intentional practice, but also underscore implementation priorities in real world deployment, such as, strong tracking, careful calibration, and assessment that is aligned to the curriculum.

Conflicts of Interest

"The authors declare no conflicts of interest".

Funding

Author, declare they have received no funding for this paper.

Acknowledgments: The authors are grateful to the Department of Computer Science under the College of Education for Pure Science at the University of Mosul in Iraq for their support.

References

[1] K. C. Choy and A. Okrainec, “Simulation in surgery: Perfecting the practice,” *Surg. Clin. North Am.*, vol. 90, no. 3, pp. 457–473, Jun. 2010, doi: 10.1016/j.suc.2010.02.011.
[2] N. E. Seymour *et al*., “Virtual reality training improves operating room performance: Results of a randomized, double-blinded study,” *Ann. Surg.*, vol. 236, no. 4, pp. 458–464, Oct. 2002, doi: 10.1097/00000658-200210000-00008.
[3] R. Lohre *et al*., “Effectiveness of immersive virtual reality training compared with video-based training on orthopedic surgical skills and knowledge acquisition among surgical residents: A randomized clinical trial,” *JAMA Netw. Open*, vol. 3, no. 12, e2031217, Dec. 2020, doi: 10.1001/jamanetworkopen.2020.31217.
[4] D. A. Kolb, *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1984, ISBN: 0132952610.
[5] P. Milgram and F. Kishino, “A taxonomy of mixed reality visual displays,” in *Proc. SPIE 2351, Telemanipulator and Telepresence Technologies*, 1994, pp. 282–292, doi: 10.1117/12.197321.
[6] R. Viglialoro *et al*., “Augmented reality, mixed reality, and hybrid approaches in healthcare simulation: A systematic review,” *Appl. Sci.*, vol. 11, no. 5, p. 2338, 2021, doi: 10.3390/app11052338.
[7] D. Suresh *et al*., “The role of augmented reality in surgical training: A systematic review,” *Surg. Innov.*, vol. 30, no. 3, pp. 366–382, 2023, doi: 10.1177/15533506221140506.
[8] A. Sheik-Ali, A. R. Edgcombe, and C. Paton, “Next-generation virtual and augmented reality in surgical education: A narrative review,” *Surg. Technol. Int.*, vol. 35, pp. 27–35, Nov. 2019. PMID: 31498872.
[9] A. Shahrezaei *et al*., “The impact of surgical simulation and training technologies on learning outcomes: A systematic review,” *BMC Med. Educ.*, vol. 24, p. 1297, 2024, doi: 10.1186/s12909-024-06299-w.
[10] A. Vajsbaher *et al*., “Spatial cognition in minimally invasive surgery training: Effects of visual guidance and simulator fidelity,” *BMC Surg.*, vol. 18, p. 51, 2018, doi: 10.1186/s12893-018-0416-1.
[11] A. Thomas *et al*., “A hybrid reality simulator for wire navigation in orthopedic surgery: Validity and reliability,” *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 3, pp. 388–397, Jun. 2015, doi: 10.1109/THMS.2014.2339324.
[12] T. C. Keating and C. A. Jacobs, “Augmented reality in orthopedic practice and education,” *Orthop. Clin. North Am.*, vol. 52, no. 1, pp. 77–86, Jan. 2021, doi: 10.1016/j.ocl.2020.08.002.
[13] F. Heinrich *et al*., “HoloPointer: Virtual augmented reality pointer for laparoscopic surgery training,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 16, pp. 651–659, 2021, doi: 10.1007/s11548-020-02272-2.
[14] A. Rojas-Muñoz *et al*., “A system for telementoring with augmented reality: A pilot study,” *Surgery*, vol. 167, no. 5, pp. 1034–1040, May 2020, doi: 10.1016/j.surg.2019.11.008.
[15] A. Nadeem-Tariq *et al*., “Augmented reality in spine surgery: Narrative review of clinical accuracy, workflow efficiency, and barriers to adoption,” *Cureus*, vol. 17, no. 6, e86803, Jun. 2025, doi: 10.7759/cureus.86803.
[16] L. Wu *et al*., “A novel augmented reality-based simulator for enhancing orthopedic surgical training,” *Comput. Biol. Med.*, vol. 185, p. 109536, Feb. 2025, doi: 10.1016/j.compbiomed.2024.109536.
[17] J. A. Martin *et al*., “Objective structured assessment of technical skill (OSATS) for surgical residents,” *Br. J. Surg.*, vol. 84, no. 2, pp. 273–278, Feb. 1997, doi: 10.1046/j.1365-2168.1997.02502.x
[18] M. Baker, R. Lontchi, and Z. Buser, “Are AI and VR tools changing spine education and training?” *Artif. Intell. Surg.*, vol. 5, pp. 73–81, 2025, doi: 10.20517/ais.2024.60.
[19] J. Bui, D. Le, S. B. Murthi, and R. M. Reddy, “The role of extended reality in enhancing surgical training: A narrative review,” *Curr. Chall. Thorac. Surg.*, vol. 7, p. 10, 2025, doi: 10.21037/ccts-24-43.