Advancing Medical Education and Planning Through Extended Reality: A Mini Review of XR Applications in Medicine

Irena Galić[1][0000−0002−0211−4568], Marija Habijan[1][0000−0002−0211−4568], Marin Benčević[1][0000−0003−4294−0781], Juraj Perić[1], Hrvoje Leventić[1][0000−0001−9568−1408], Krešimir Romić[1][0000−0002−8656−6712], Ivana Hartmann Tolić[1][0000−0002−6805−1328], Robert Šojo[1], Aleksandra Pižurica[3][0000−0002−9322−4999], Danilo Babin[4][0000−0002−2881−6760], Dario Muževič[2], Vjekoslav Kopačin[2][0000−0002−3854−6602], Maja Košuta Petrović[2], and Mirta Vujnovac[5]

[1] J. J. Strossmayer University of Osijek, Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Croatia, `marija.habijan@ferit.hr`

[2] Clinical Medical Center Osijek, Croatia

[3] Ghent University, Department of Telecommunications and Information Processing, TELIN-GAIM, Belgium

[4] Ghent University, Department of Telecommunications and Information Processing, imec-TELIN-IPI, Belgium

[5] III Gymnasium Osijek, Croatia

Abstract. The integration of extended reality (XR) technologies, virtual, augmented, and mixed reality with artificial intelligence (AI) is transforming medical education, diagnostics, and surgical planning. This mini review explores how established AI methods such as convolutional neural networks (CNNs), recurrent networks (RNNs), generative adversarial networks (GANs), and reinforcement learning (RL) are being used to enhance XR systems for anatomical segmentation, realistic simulation, and autonomous interaction. It also examines emerging approaches, including diffusion models (DMs), vision transformers (ViTs), and multimodal learning (MML), which enable high-fidelity synthetic data generation, contextual scene understanding, and integration of heterogeneous inputs such as imaging, text, and sensor data. Through use cases in placenta accreta diagnosis and neurovascular intervention planning, we demonstrate how AI-enhanced XR systems can deliver immersive, intelligent, and personalized experiences for clinicians and trainees. We further outline technical challenges, including real-time performance, data variability, and interpretability, and discuss strategies to ensure safe, equitable, and effective adoption of AI-driven XR in healthcare.

Keywords: Artifical Intelligence · Augmented Reality · Extended Reality · Virtual Reality · Medical Education · Personalized Education

1 Introduction

Extended reality (XR) is a term that includes virtual reality (VR), augmented reality (AR), and mixed reality (MR). These technologies merge digital information with the physical environment to varying degrees, providing immersive three-dimensional (3D) experiences [7, 20]. In XR, VR typically refers to fully immersive simulations blocking out the real world, AR overlays digital content onto the real-world view, and MR blends interactive virtual objects into the real environment with spatial anchoring [1]. Recent advances in computing and display hardware have accelerated the adoption of XR in medicine [1, 7]. Early applications demonstrated the benefits of XR's 3D visualization and intuitive interaction for medical tasks that are challenging with traditional 2D screens. For example, XR enables hands-free interaction with 3D anatomical models in sterile operating rooms, overcoming the limitations of desktops and tablets in these environments. Over the past decade, XR has evolved from experimental prototypes to practical tools in medical education and clinical practice [41, 26].

This mini review explores state-of-the-art XR applications in two key areas: (1) medical education (such as anatomy learning, diagnostic training, and clinical simulations) and (2) surgical and clinical planning (including procedural rehearsal and image-guided interventions). We discuss how XR is being used to enhance learning and preoperative planning, examine the evidence of its benefits (and risks), and highlight current limitations. Section 2 reviews educational and clinical use cases, while Section 3 highlights how AI methods enhance XR systems. Sections 4 present examples in placenta accreta diagnosis and neurovascular intervention, followed by AI pipeline integration in Section 5 and a discussion of challenges in Section 6. Section 7 concludes with future directions. The review aims to guide the development of accessible and scalable XR systems for medical education and practice.

2 XR in Medical Education and Clinical Planning

XR is widely used in teaching of anatomy, surgical skills, and clinical scenarios, often matching or surpassing traditional methods in effective learning engagement, and cost efficiency [7]. Tools like virtual anatomy platforms and AR overlays enhance spatial understanding and contextual learning, while increasing motivation and retention through multi-sensory interaction [43]. XR also improves access to high-quality simulation, offering portable, scalable alternatives to costly, centralized training labs. VR headsets allow for remote asynchronous practice, support standardized experiences, and reduce dependence on physical wear [40]. A range of XR tools support diverse educational needs: from VR surgical simulators (e.g. for endoscopy) to MR headsets such as HoloLens, which project interactive 3D anatomy for collaborative learning [15]. Virtual dissection tables bridge solo and group learning, further expanding XR's applicability. Evidence shows XR boosts procedural confidence and skill acquisition, with outcomes often comparable or superior to conventional teaching [19]. However, integration challenges remain, including the need for more research on mobile and multi-user platforms, curricular alignment, and soft skills training. As XR matures, it is poised to become a standard complement to medical education, particularly in anatomy and procedural instruction [25].

XR technologies are also becoming essential tools in surgical planning and image-guided procedures by enabling detailed, interactive visualization of patient-specific anatomy. VR allows surgeons to explore 3D reconstructions of CT/MRI scans, enhancing spatial understanding and supporting preoperative rehearsal [6]. AR and MR systems can project this data onto the patient during surgery, assisting with real-time navigation and improving precision. Studies across specialties, such as thoracic and urologic surgery, show that VR-based planning improves anatomical insight, surgical confidence, and outcomes, including reduced operative times and improved preservation of healthy tissue [10]. Intraoperatively, AR reduces the need to shift attention between monitors and patients by embedding navigation data directly into the surgeon's field of view [36]. Applications in neurosurgery and vascular surgery show promise, with benefits such as better alignment of surgical targets and more intuitive incision planning [34]. XR also supports interdisciplinary collaboration and patient education, enabling shared understanding through 3D models in *virtual consultations*. Despite its promise, XR faces challenges such as ensuring accurate hologram alignment, managing device ergonomics, and integrating systems into existing workflows. Nonetheless, early trials report positive outcomes and no adverse events, suggesting XR may soon become a standard aid in complex surgical planning and execution [9]. A structured comparison of VR, AR, and MR modalities in terms of their immersion level, educational and surgical applications, hardware, strengths, and limitations is presented in Table 1.

Table 1: Systematic comparison of XR methodologies in medicine

Feature	**VR**	**AR**	**MR**
Immersion level	Full immersion	Low to moderate	Moderate to high
Use in medical education	High-fidelity simulations, anatomy learning	Real-time overlays for anatomy and procedural steps	Interactive group learning with spatially anchored virtual content
Use in surgery	Preoperative planning, rehearsal	Intraoperative navigation and guidance	Real-time guidance with interactive overlays and tool tracking
Hardware examples	Oculus Rift, HTC Vive, Meta Quest	Smartphone-based AR, Microsoft HoloLens	Microsoft HoloLens 2, Magic Leap with spatial tracking
Strengths	Deep immersion, effective skill transfer, standardized training	Real-world integration, supports real-time feedback, cost-effective	High realism, spatial accuracy, user interaction with real-world context
Limitations	Lack of real-world context, motion sickness, hardware cost	Limited field of view, alignment challenges, requires stable tracking	High complexity, hardware cost, computational requirements

3 AI Methods for Medical XR

In this section we present eastablished and emerging AI methods. Established techniques, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), generative adversarial networks (GANs), and reinforcement learning (RL), have been foundational in tasks like image segmentation, temporal modeling, image synthesis, and interactive control. Meanwhile, newer paradigms like vision transformers (ViTs), diffusion models (DMs), and multi-modal fusion architectures are pushing boundaries by offering improved data representation, generative realism, and integrated understanding across diverse medical modalities. Together, these AI models enable XR systems to deliver more realistic, adaptive, and clinically relevant experiences for diagnosis, training, and intraoperative guidance.

3.1 Established AI Methods for Medical XR

Several well-established AI methods serve as the foundation for enhancing medical image interpretation, simulation, and interactivity in XR environments. Fig. 1 present a conceptual overview of established AI methods that drive innovation in medical XR.

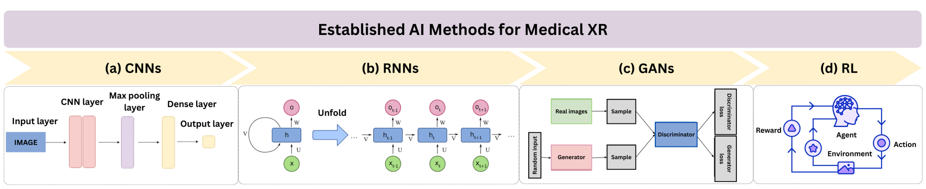

Fig. 1: Overview of established AI methods for medical XR applications: (a) CNNs for image analysis and segmentation, (b) RNNs for temporal modeling, (c) GANs for realistic image synthesis, and (d) RL for interactive control and simulation.

Convolutional Neural Networks (CNNs): CNNs form the backbone of many XR medical applications, especially for medical image segmentation and reconstruction. For example, U-Net architectures [32] are widely used to segment patient-specific anatomy (e.g. organs, vasculature) from volumetric scans, producing 3D models that can be visualized in AR/VR. In placenta accreta diagnosis, an nnU-Net (a self-configuring 3D U-Net variant) was applied to MRI data for automatic placenta segmentation, which improved diagnostic AUC beyond expert radiologists [37]. This segmentation allows the overlay of critical structures in the view of the clinician in XR, or the import of patient-specific models into simulators. CNN-based image analysis also supports XR by detecting and tracking instruments or anatomical landmarks in real time, enabling interactive feedback during simulation and image-guided interventions [30].

Recurrent Neural Networks (RNNs): While less common than CNNs in medical imaging, RNNs are employed in XR systems to model temporal dynamics and sequences. For instance, physics-informed deep networks combining CNN and LSTM layers have been used to emulate soft tissue behavior over time in surgical VR simulations [35]. By training on biomechanical simulation data, these models can predict tissue deformations and tool interactions with high fidelity, but at a fraction of the computational cost of finite element methods. This enables real-time haptic feedback and realistic tissue response in VR surgical trainers, which is crucial for interventions like neurovascular embolization that involve complex, time-dependent vessel and catheter interactions [14]. In general, RNNs can learn system dynamics, allowing XR environments to respond realistically to user actions (e.g. simulating physiologic changes or instrument motion over time).

Generative Adversarial Networks (GANs): GAN-based models are powerful for image generation and domain translation in XR medical systems [3]. They can synthesize realistic medical images or augment simulation visuals, addressing the limited realism of conventional simulators. A notable example is using GANs to create a real-time ultrasound simulation: starting from simulated ultrasound physics output, a GAN was trained to produce realistic B-mode ultrasound images that mimic speckle and tissue texture [23]. Moreover, GAN-based domain adaptation can translate between imaging modalities, such as generating ultrasound-like images from MRI scans, helping to improve multi-modal visualization in XR systems. These capabilities help create more immersive and varied XR simulations, from virtual patients with different anatomies to on-the-fly realistic imaging feedback during an AR-guided procedure [31, 29].

Reinforcement Learning (RL): RL techniques are emerging as tools to automate and optimize interactions within XR simulations [11]. In medical training simulators, RL can be used to train virtual assistants or autonomous agents that demonstrate surgical maneuvers, adjust scenario difficulty, or control virtual patients [39]. For example, researchers have applied RL (using methods like Proximal Policy Optimization) to learn automated surgical subtasks in a virtual reality setting [4]. Such RL-driven automation within XR can provide trainees with an intelligent training partner or dynamically adjust simulations based on performance [27]. Beyond training, RL has been used for planning in image-guided therapy – for instance, to optimize the path of a catheter or endoscope in a complex 3D vessel network. Overall, combining RL with high-fidelity XR environments enables skill acquisition through trial-and-error in simulation and can adaptively enhance realism and educational value in medical XR systems.

3.2 Emerging AI Techniques and Multi-Modal Systems

As medical XR continues to evolve, new AI paradigms are emerging that push the boundaries of what is possible in simulation, visualization, and interaction.

These cutting-edge approaches aim to enhance the fidelity, adaptability, and intelligence of XR systems by leveraging recent advances in generative modeling, vision architectures, and multi-modal integration. Fig. 2 present a conceptual overview of emerging AI methods that drive innovation in medical XR.

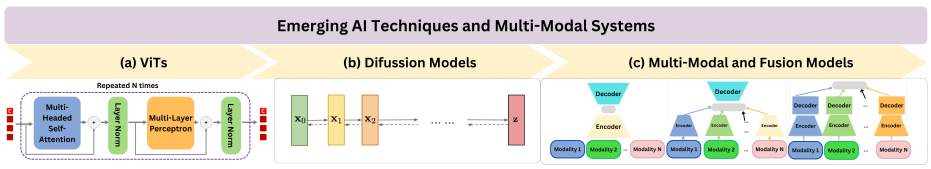

Fig. 2: Overview of emerging AI methods for medical XR applications: (a) Vision Transformers (ViTs) for capturing long-range dependencies in medical images, (b) Diffusion Models for high-fidelity generative tasks such as 3D volume synthesis, and (c Multi-Modal and Fusion Models capable of integrating diverse data types (e.g., text, imaging, and sensor data) to support intelligent XR systems.

Diffusion Models Diffusion-based generative models represent a cutting-edge approach for creating high-quality synthetic medical data and enhancing realism in XR. Unlike GANs, diffusion models iteratively refine random noise into coherent images, often yielding higher fidelity and diversity. These models have shown state-of-the-art ability to generate realistic anatomy and even assist with tasks like sparse-view reconstruction and data augmentation [21, 38, 22]. In an XR context, diffusion models could generate diverse patient cases (e.g. various placenta shapes or vascular anatomies) for simulation training, or fill in missing visual details on-the-fly to improve graphical realism. They also enable multi-modal image synthesis: for example, creating plausible MRI from ultrasound or vice versa, which could be used to enrich the visual feedback in AR fusion systems. While still an emerging research area, diffusion models hold promise for overcoming data scarcity and increasing the variability and authenticity of XR medical content [17].

Vision Transformers (ViTs) Transformer-based vision models have recently gained importance in medical imaging due to their ability to capture long-range dependencies and contextual information. ViT models and hybrid CNN-Transformer architectures (e.g. Swin UNETR) are now achieving or exceeding state-of-the-art performance on tasks like 3D segmentation and classification [16]. Transformers can be pre-trained on massive image datasets and fine-tuned with relatively small medical datasets, addressing data limitations. In XR systems, ViTs can enhance perception and understanding of the scene, for example, accurately segmenting organs or vessels in real-time to overlay onto a surgeon's AR display, or identifying clinical landmarks in a VR training scenario. By integrating ViT models, XR platforms benefit from robust and efficient analysis of complex imaging data, ultimately providing more reliable visualizations (e.g. highlighting a suspected accreta region in an AR view) and interactions. As transformer models continue to evolve (becoming more computationally efficient and hybridized with CNNs), they are expected to play a growing role in XR for tasks requiring high precision and understanding of global context in images [33].

Multi-Modal AI and Fusion XR medical applications inherently involve multiple data modalities – combining medical images (CT, MRI, ultrasound), sensor data, text (medical records or instructions), and even haptic signals. Emerging AI methods aim to fuse these modalities into a coherent understanding or output [8]. In practice, a multi-modal XR system might take as input a 3D MRI, an ultrasound video stream, and perhaps physician voice commands, and produce an output such as an AR overlay aligned to the patient with diagnostic annotations. Initial strides in this direction include models that align different imaging modalities: e.g. a learning-based system to register and fuse MRI and ultrasound images for prenatal diagnostics [2]. These XR systems could not only display merged imagery but also understand and respond to complex combinations of inputs (visual, textual, auditory, haptic), leading to more intuitive and intelligent medical training and guidance platforms.

4 Clinical Applications of AI-Enhanced XR in Medicine

The following section presents concrete examples that illustrate how the integration of AI and XR transforms medical imaging and surgical planning, focusing on placenta accreta diagnosis (ultrasound/MRI fusion) and neurovascular simulation and embolization planning, through enhanced visualization, real-time guidance, and patient-specific modeling.

4.1 Placenta Accreta Diagnosis (Ultrasound/MRI fusion)

Placenta accreta spectrum (PAS) is a high-risk obstetric condition where the placenta invades the uterine wall, often requiring complex surgery. Accurate diagnosis and pre-surgical planning are critical [12]. Ultrasound (US) is the first-line imaging modality for suspected accreta, but it can be limited in deep invasion cases, whereas MRI provides complementary soft-tissue detail. Extended reality offers a novel way to combine these imaging modalities and assist clinicians [28]. In an XR diagnostic setup, a physician wearing an AR headset could perform a live ultrasound while seeing overlaid 3D information from a prior MRI on the patient [5]. This US/MRI fusion in AR provides a rich, spatially intuitive view – for example, highlighting the placenta boundaries and areas of uterine invasion in 3D within the patient's body, rather than interpreting 2D images side-by-side. Early developments suggest that augmented reality overlays can indeed enhance the understanding of placental anatomy: in one report, real-time AR visualization of a 3D placenta model during ultrasound improved diagnostic clarity for accreta [18].

To enable such XR fusion, AI methods are pivotal. A CNN (like U-Net) can automatically segment the placenta and abnormal invasion zones on the MRI, generating a 3D volume of the placenta accreta [42]. This model is then imported into the AR system and spatially registered to the patient's anatomy (using landmarks or sensor tracking). During the ultrasound exam, the AR headset displays the segmented placenta (from MRI) correctly aligned within the patient, while the operator simultaneously sees live ultrasound slices in their field of view. The result is a composite "X-ray vision" effect – the physician looks through the abdomen and observes both the real-time ultrasound and the contextualized MRI-derived structures [13]. For placenta accreta, this could translate to better intraoperative planning (e.g. visualizing how deeply placenta penetrates and its relation to bladder or vessels), potentially reducing surprises during surgery. AI contributes further by analyzing the multimodal data. Beyond segmentation, machine learning classifiers can assess invasion depth or likelihood of placenta accreta from the images (using radiomic features or deep networks). The XR interface might also allow clinicians to interact with the data – for instance, voice-querying the system for measurements like placental thickness or getting a projected blood loss risk based on AI analysis of the case. This tight integration of AI-driven analysis with immersive visualization exemplifies how XR can augment clinical decision-making.

4.2 Neurovascular Simulation and Embolization Planning

Neurovascular procedures, such as cerebral aneurysm coiling or AVM embolization, demand a high level of skill due to intricate cerebrovascular anatomy and the risks of navigation within blood vessels. XR technology is being applied to neurosurgery training and planning to enhance understanding of complex 3D anatomy and to allow rehearsal of interventions in a safe environment. A recent systematic review found that extended reality greatly aids endovascular neurosurgery education: VR simulations were used in the majority of studies (for procedural training on aneurysm coiling, thrombectomy, etc.), while AR/MR were used for visualizing patient-specific anatomy and intraoperative guidance [30].

AI is advancing neurovascular simulators by enabling patient-specific virtual environments. Using CT/MR angiography, CNNs can segment cerebral vessels and generate 3D models for VR, allowing trainees to practice on digital twins and test catheterization strategies. AI-driven simulations model real-time blood flow and contrast propagation, while generative models like diffusion networks can create diverse aneurysm scenarios to enhance preparedness [24].

In XR-based planning, surgeons can explore 3D holograms of patient vasculature, simulate device placement, and receive AI-driven guidance on risk zones or optimal fluoroscopy angles. When paired with haptics, AI models replicate tactile feedback, helping train fine motor skills needed for delicate tasks. Reinforcement learning and RNNs improve realism by simulating catheter behavior and vessel response.

Intraoperatively, AR can overlay key anatomy onto the patient, offering "X-ray vision" during embolizations. Tracked catheter positions and real-time imaging can be integrated, though precise tool tracking remains a technical challenge.

AI helps by improving registration accuracy and predicting instrument motion, supporting safer and more effective navigation.

5 Integration of AI Models into XR Pipelines

Incorporating AI into XR pipelines requires a delicate balance between data processing and interactive, real-time visualization. A general pipeline might look like: data acquisition → AI model inference → XR rendering, all potentially happening in real or near-real time as illustrated in Fig. 3.

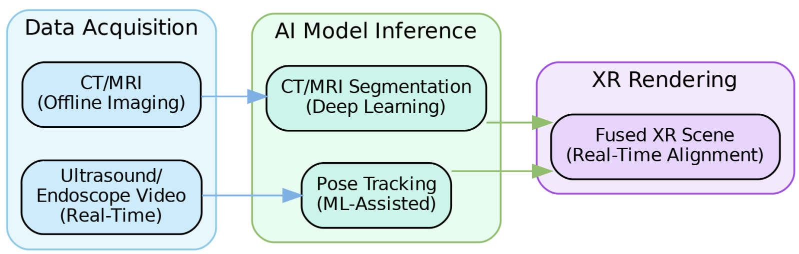

Fig. 3: AI-enhanced XR pipeline for medical visualization. The system integrates two data sources: offline imaging (CT/MRI) and real-time imaging (ultrasound or endoscopic video) which are processed by AI models for CT/MRI segmentation (via deep learning) and pose tracking (via ML-assisted methods). These outputs are combined in the XR rendering stage to generate a fused XR scene aligned in real time with the patient's anatomy, enabling accurate and immersive visualization for diagnostic or surgical guidance.

For example, consider an AR surgical guidance system: it acquires patient imaging (CT/MRI) offline and live data (ultrasound or endoscope video) in real time. Offline, a deep learning segmentation model processes the CT/MRI to produce 3D meshes of anatomy. In real time, a pose-tracking algorithm (often augmented by ML for robustness) continuously aligns the virtual content with the patient. A rendering engine then displays the fused scene to the user. Each of these stages can be improved or accelerated by AI.

Regarding real-time image analysis, XR systems often rely on understanding live sensor input (camera views, ultrasound streams). Latency is a major concern; many XR applications require < 50 ms pipeline delay to feel instantaneous. Here, efficient neural networks or even specialized hardware (GPUs, TPUs) are critical. Techniques like model quantization, distillation, and use of compact architectures (e.g. MobileNet, or neuron pruning in a U-Net) are employed to ensure AI models don't bottleneck the XR experience.

Moreover, regarding interactive graphics and image synthesis, it is important to emphesise that AI-generated content can be fed directly into the XR rendering stage. We saw how GANs can produce ultrasound images on the fly; similarly, diffusion models or neural radiance fields (NeRFs) might be used to render anatomy with realistic textures and lighting in VR. A challenge is to do this on demand as the user's viewpoint changes. One approach is to pre-compute a lot of variability (e.g. many angles of a surgical scene) and use lightweight interpolation during interaction. Another is to use neural rendering – where part of the graphics pipeline is a neural network that fills in detail. This is cutting-edge in XR: research prototypes exist where a neural network helps render unseen viewpoints of organs by learning from CT data. Ensuring stability and avoiding flicker or artifacts is non-trivial; thus, often a hybrid pipeline is used (traditional graphics for baseline visualization + AI for refinement). The integration of AI-based denoising is already more common. For example, noisy ultrasound images projected in AR can be cleaned up by a deep denoising model in real time, making the overlay clearer. Ultimately, the XR pipeline becomes a mix of deterministic algorithms and AI modules, each feeding into the next.

Nevertheless, XR is not just visual. Many medical simulators involve haptic feedback (force, tactile simulation) to mimic procedures. Integrating AI here means using learned models to simulate physics faster or more realistically. Classical haptic engines might use physics equations that are simplified for speed. By training on high-fidelity simulations or real force measurements, a neural network can learn the mapping from user actions (e.g. how far a needle is inserted) to resulting feedback (resistance, tissue deformation). Once learned, this can be extremely fast to compute. For example, a deep model was trained to predict soft tissue deformation under various forces, achieving accurate results at interactive rates, whereas a full finite element simulation would be too slow for real-time use.

6 Current Challenges and Future Directions

Integrating AI with XR in medicine offers transformative potential but also presents several challenges. A primary concern is the limitation and bias in training data, which can hinder AI models' ability to generalize across diverse patient populations and devices. Variations in imaging equipment and clinical practices may further degrade performance. To mitigate these issues, researchers are exploring large-scale, multicenter datasets and domain adaptation techniques. Synthetic data generation, such as creating labeled 3D CT scans, shows promise but requires caution to prevent overfitting to artificial characteristics. The "black-box" nature of many AI algorithms raises concerns about interpretability and trust, especially when understanding decision-making processes is crucial in clinical settings. Efforts in Explainable AI (XAI) aim to provide insights into AI decisions by highlighting influential image regions and displaying confidence levels, thereby fostering clinician trust. Achieving real-time, high-resolution AI processing in XR environments remains challenging due to the computational demands of complex models, particularly with volumetric data.

Solutions include optimizing neural networks, leveraging edge or cloud computing, and developing specialized hardware to minimize latency and ensure seamless XR experiences. Additionally, AI-integrated medical XR systems must undergo rigorous validation to meet safety and efficacy standards set by regulatory bodies like the FDA. The opacity of some AI models complicates this process, necessitating transparency, robustness, and comprehensive testing across diverse scenarios for regulatory approval and clinical adoption. Ethical and training considerations are also paramount; balancing AI assistance with the preservation of fundamental clinical skills is critical. While AI and XR can democratize medical training, there's a risk of over-reliance, potentially diminishing essential competencies. Ensuring equitable access to these technologies and addressing data privacy concerns are essential. Techniques like federated learning, which allow AI models to learn from decentralized data without compromising patient privacy, are being explored to responsibly enhance XR systems. In summary, while AI and XR hold immense promise for transforming medical education and practice, addressing these challenges through ongoing research and ethical considerations is essential for their successful integration into healthcare.

7 Conclusion and Discussion

AI-driven extended reality is poised to revolutionize medical training and intervention, fusing the strengths of human intuition with the precision and breadth of machine intelligence. We now have XR simulators where convolutional networks provide patient-specific 3D anatomy, transformers and multimodal models fuse data into rich interactive experiences, and generative models populate virtual worlds with realistic scenarios that were once scarce or unimaginable. These advances enable a trainee to palpate a virtual organ with haptic feedback informed by real physics, or a surgeon to visualize hidden pathology floating before their eyes with AR confidence markers. The result is a more immersive learning environment and potentially safer, more effective patient care – early studies already show improved procedural performance after AI-enhanced XR training. At the same time, our exploration of current limitations reminds us that this is a journey in progress. Ensuring that an AI model's superb accuracy in the lab translates to the chaotic variability of the real world is an ongoing endeavor. The future directions highlighted – from creating larger and more diverse datasets, to developing explainable and trustworthy AI, to integrating these tools seamlessly into clinical workflows – are being actively pursued by the research community. We can expect the next decade to bring XR systems that are more intelligent (with AI that can reason across modalities and scenarios), more adaptive (tailoring themselves to each user and patient), and more accepted in everyday clinical practice. Key to this will be maintaining a focus on the end-users: the medical professionals and patients. XR and AI should ultimately amplify the expertise of clinicians, providing support where humans have limitations (e.g. seeing the unseen, recalling vast case libraries), while preserving the empathy, judgment, and responsibility that only humans can provide.

In conclusion, the convergence of advanced AI methods with extended reality platforms is opening new frontiers in medical simulation, planning, and guidance. Current methods like CNNs, GANs, and RL have already shown how realistic virtual training and augmented guidance can be achieved, and emerging approaches like diffusion models, vision transformers, and multimodal learning are set to push these boundaries even further. The use cases of placenta accreta management and neurovascular interventions exemplify the tangible benefits – from improved diagnostic confidence to hands-on rehearsal of high-stakes procedures – that these technologies can deliver. By diligently addressing the challenges of generalization, interpretability, data curation, and validation, researchers and clinicians together can ensure that future XR systems are not only technologically impressive but also clinically impactful and safe. The ultimate vision is a healthcare ecosystem where learning and practice are enriched by immersive, intelligent tools: surgeons entering the operating theater having essentially "already performed" the surgery in VR with an AI tutor, and when the real procedure happens, having AI-augmented AR guidance to navigate complexity. Achieving this will mark a significant leap in how we train doctors and treat patients, fulfilling the promise of extended reality amplified by the power of machine learning.

Acknowledgements. This work was supported by the Croatian Science Foundation under the project number IP-2024-05-9492.

References

1. Andrews, C.E., Southworth, M.K., Silva, J.N.A., Silva, J.R.: Extended reality in medical practice. Current Treatment Options in Cardiovascular Medicine **21**, 1–12 (2019), https://api.semanticscholar.org/CorpusID:89618916
2. B, J.K.D.K., G, D.C.L.U.N.: Real-time virtual sonography using mri and ultrasound fusion imaging in the evaluation of cns anomalies (2015), https://api.semanticscholar.org/CorpusID:26995384
3. Bengesi, S., El-Sayed, H., Sarker, M.K., Houkpati, Y., Irungu, J., Oladunni, T.: Advancements in generative ai: A comprehensive review of gans, gpt, autoencoders, diffusion model, and transformers. IEEE Access **12**, 69812–69837 (2023), https://api.semanticscholar.org/CorpusID:265281267
4. Bourdillon, A.T., Garg, A., Wang, H., Woo, Y.J., Pavone, M., Boyd, J.H.: Integration of reinforcement learning in a virtual robotic surgical simulation. Surgical Innovation **30**, 94 – 102 (2022), https://api.semanticscholar.org/CorpusID:248503124
5. Castro, P.T., Júnior, E.A., Lopes, J., Ribeiro, G., Werner, H.: Placenta accreta: Virtual reality from 3d images of magnetic resonance imaging. Journal of Clinical Ultrasound **50**, 119 – 120 (2021), https://api.semanticscholar.org/CorpusID:244402114
6. Chang, Y.Z., Wu, C.T.: Application of extended reality in pediatric neurosurgery: A comprehensive review. Biomedical journal p. 100822 (2024), https://api.semanticscholar.org/CorpusID:274605449

7. Curran, V.R., Xu, X., Aydin, M.Y., Meruvia-Pastor, O.: Use of extended reality in medical education: An integrative review. Medical Science Educator **33**, 275 – 286 (2022), https://api.semanticscholar.org/CorpusID:254876440
8. Dai, Y., Gao, Y.: Transmed: Transformers advance multi-modal medical image classification. Diagnostics **11** (2021), https://api.semanticscholar.org/CorpusID:232170589
9. D'Amico, R.S., Barrington, N.M., Langer, D.J.: Opinion: Navigating the integration and impact of extended reality in neurosurgery. Surgical Neurology International **15** (2024), https://api.semanticscholar.org/CorpusID:270684228
10. Elsakka, A., Park, B.J., Marinelli, B., Swinburne, N.C., Schefflein, J.: Virtual and augmented reality in interventional radiology: Current applications, challenges, and future directions. Techniques in vascular and interventional radiology **26 3**, 100919 (2023), https://api.semanticscholar.org/CorpusID:264385482
11. Feng, B.: A deep reinforcement learning-based resource allocation mechanism for xr applications*. 2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB) pp. 1–6 (2023), https://api.semanticscholar.org/CorpusID:260933507
12. Fratelli, N., Fichera, A., Prefumo, F.: An update of diagnostic efficacy of ultrasound and magnetic resonance imaging in the diagnosis of clinically significant placenta accreta spectrum disorders. Current Opinion in Obstetrics & Gynecology **34**, 287 – 291 (2022), https://api.semanticscholar.org/CorpusID:251131704
13. Gadodia, G., Evans, M., Weunski, C., Ho, A., Cargill, A., Martin, C.: Evaluation of an augmented reality navigational guidance platform for percutaneous procedures in a cadaver model. Journal of Medical Imaging **11** (2024), https://api.semanticscholar.org/CorpusID:267715794
14. Gani, A., Pickering, O.J., Ellis, C., Sabri, O., Pucher, P.H.: Impact of haptic feedback on surgical training outcomes: A randomised controlled trial of haptic versus non-haptic immersive virtual reality training. Annals of Medicine and Surgery **83** (2022), https://api.semanticscholar.org/CorpusID:252500398
15. García-Robles, P., Cortés-Pérez, I., Nieto-Escámez, F.A., García-López, H., Obrero-Gaitán, E., Osuna-Pérez, M.C.: Immersive virtual reality and augmented reality in anatomy education: A systematic review and meta-analysis. Anatomical Sciences Education **17**, 514 – 528 (2024), https://api.semanticscholar.org/CorpusID:267635560
16. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. ArXiv **abs/2201.01266** (2022), https://api.semanticscholar.org/CorpusID:245668780
17. Herur-Raman, A., Almeida, N.D., Greenleaf, W.J., Williams, D., Karshenas, A., Sherman, J.H.: Next-generation simulation—integrating extended reality technology into medical education. In: Frontiers in Virtual Reality (2021), https://api.semanticscholar.org/CorpusID:237425857
18. Huang, J., Do, Q.N., Shahedi, M., Xi, Y., Lewis, M.A., Herrera, C.L., Owen, D., Spong, C.Y., Madhuranthakam, A.J., Twickler, D.M., Fei, B.: Deep-learning-based automatic segmentation of the placenta and uterine cavity on prenatal mr images. In: Medical Imaging (2023), https://api.semanticscholar.org/CorpusID:258016071
19. Jiang, H., Vimalesvaran, S., Wang, J.K., Lim, K.B., Mogali, S.R., Car, L.T.: Virtual reality in medical students' education: Scoping review. JMIR Medical Education **8** (2021), https://api.semanticscholar.org/CorpusID:245610841
20. Kazemzadeh, K., Akhlaghdoust, M., Zali, A.: Advances in artificial intelligence, robotics, augmented and virtual reality in neurosurgery. Frontiers in Surgery **10** (2023), https://api.semanticscholar.org/CorpusID:261194809

21. Kazerouni, A., Aghdam, E.K., Heidari, M., Azad, R., Fayyaz, M., Hacihaliloglu, I., Merhof, D.: Diffusion models in medical imaging: A comprehensive survey. Medical image analysis **88**, 102846 (2022), https://api.semanticscholar.org/CorpusID:253523207
22. Khader, F., Mueller-Franzes, G., Arasteh, S.T., Han, T., Haarburger, C., Schulze-Hagen, M.F., Schad, P., Engelhardt, S., Baessler, B., Foersch, S., Stegmaier, J., Kuhl, C., Nebelung, S., Kather, J.N., Truhn, D.: Medical diffusion: Denoising diffusion probabilistic models for 3d medical image generation (2022), https://api.semanticscholar.org/CorpusID:260444528
23. Lainé, N., Zahnd, G., Bernard, O., Orkisz, M., Liebgott, H.: Generation of realistic simulated b-mode image texture with a gan. 2022 IEEE International Ultrasonics Symposium (IUS) pp. 1–4 (2022), https://api.semanticscholar.org/CorpusID:254155038
24. Lee, K.H., Muniyandi, P., Li, M., Saccenti, L., Christou, A.S., Xu, S., Wood, B.J.: Smartphone technology for applications in image-guided minimally invasive interventional procedures. Cardiovascular and Interventional Radiology **48**, 142 – 156 (2024), https://api.semanticscholar.org/CorpusID:274787363
25. Moro, C., Stromberga, Z., Stirling, A.: Virtualisation devices for student learning: Comparison between desktop-based (oculus rift) and mobile-based (gear vr) virtual reality in medical and health science education. Australasian Journal of Educational Technology **33**, 1–10 (2017), https://api.semanticscholar.org/CorpusID:54012958
26. Nakaguchi, T.: The current state of the use of xr technology in medicine. The Japanese Journal of Rehabilitation Medicine (2024), https://api.semanticscholar.org/CorpusID:275365258
27. Nguyen, T.T., Nguyen, N.D., Bello, F., Nahavandi, S.: A new tensioning method using deep reinforcement learning for surgical pattern cutting. 2019 IEEE International Conference on Industrial Technology (ICIT) pp. 1339–1344 (2019), https://api.semanticscholar.org/CorpusID:57825757
28. Palaiologou, E., Chiu, W., de Souza, R.R., Goggin, P., Lofthouse, E.M., Cleal, J.K., Page, A., Lewis, R.M.: High resolution three dimensional imaging of placental villi using serial block face scanning electron microscopy. Placenta **45**, 84 (2016), https://api.semanticscholar.org/CorpusID:88963351
29. Parmar, T.: Generative adversarial networks for historical data generation in semiconductor manufacturing: Applications and challenges. International Journal For Multidisciplinary Research (2024), https://api.semanticscholar.org/CorpusID:276413421
30. Patel, S.A., Covell, M.M., Patel, S., Kandregula, S., Palepu, S., Gajjar, A.A., Shekhtman, O., Sioutas, G.S., Dhanaliwala, A., Gade, T., Burkhardt, J.K., Srinivasan, V.M.: Advancing endovascular neurosurgery training with extended reality: opportunities and obstacles for the next decade. Frontiers in Surgery **11** (2024), https://api.semanticscholar.org/CorpusID:272050580
31. Peng, B., Huang, X., Wang, S.Y., Jiang, J.: A real-time medical ultrasound simulator based on a generative adversarial network model. 2019 IEEE International Conference on Image Processing (ICIP) pp. 4629–4633 (2019), https://api.semanticscholar.org/CorpusID:202780747
32. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. ArXiv **abs/1505.04597** (2015), https://api.semanticscholar.org/CorpusID:3719281

33. S, M., S, A., S, Y., T, A.S.K.: Transforming digital interaction: An in-depth exploration of extended reality(xr) technologies. In: 2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI). pp. 937–941 (2024). https://doi.org/10.1109/ICDICI62993.2024.10810916
34. Sueyoshi, T., Sugimoto, M.: [clinical application of metaverse and extended reality in neurosurgery: Spatial image analysis, remote surgical assistance, and re-live training]. Brain and nerve = Shinkei kenkyu no shinpo **75 10**, 1129–1134 (2023), https://api.semanticscholar.org/CorpusID:264287126
35. Sundermeyer, M., Schlüter, R., Ney, H.: Lstm neural networks for language modeling. In: Interspeech (2012), https://api.semanticscholar.org/CorpusID:18939716
36. Tortora, M., Luppi, A., Pacchiano, F., Marisei, M., Grassi, F., Werner, H., Kitamura, F.C., Tortora, F., Caranci, F., Ferraciolli, S.F.: Current applications and future perspectives of extended reality in radiology. La Radiologia medica (2025), https://api.semanticscholar.org/CorpusID:277409936
37. Wang, H., Wang, Y., Zhang, H., Yin, X., Wang, C., Lu, Y., Song, Y., Zhu, H., Yang, G.: A deep learning pipeline using prior knowledge for automatic evaluation of placenta accreta spectrum disorders with mri. Journal of Magnetic Resonance Imaging **59** (2023), https://api.semanticscholar.org/CorpusID:258676898
38. Wang, H., Liu, Z., Sun, K., Wang, X., Shen, D., Cui, Z.: 3d med-diffusion: A 3d medical diffusion model for controllable and high-quality medical image generation. ArXiv **abs/2412.13059** (2024), https://api.semanticscholar.org/CorpusID:274789446
39. Wang, Z., Wang, R., Wu, J., Zhang, W., Li, C.: Dynamic resource allocation for real-time cloud xr video transmission: A reinforcement learning approach. IEEE Transactions on Cognitive Communications and Networking **10**, 996–1010 (2024), https://api.semanticscholar.org/CorpusID:267124648
40. Wong, J.Y., Azam, A.B., Cao, Q., Huang, L., Xie, Y., Winkler, I., Cai, Y.: Evaluations of virtual and augmented reality technology-enhanced learning for higher education. Electronics (2024), https://api.semanticscholar.org/CorpusID:269265828
41. ming Yang, J., Chen, B.J., Li, R.Y., Huang, B.Q., Zhao, M.H., Liu, P.R., Zhang, J.Y., Ye, Z.W.: Artificial intelligence in medical metaverse: Applications, challenges, and future prospects. Current medical science (2024), https://api.semanticscholar.org/CorpusID:274711422
42. Zheng, C., Zhong, J., Wang, Y., Cao, K., Zhang, C., Yue, P., Xu, X., Yang, Y., Liu, Q., Zou, Y., Huang, B.: Deep learning radiomic analysis of mri combined with clinical characteristics diagnoses placenta accreta spectrum and its subtypes. Journal of magnetic resonance imaging : JMRI (2024), https://api.semanticscholar.org/CorpusID:267806822
43. Ziker, C., Truman, B., Dodds, H.: Cross reality (xr): Challenges and opportunities across the spectrum. Innovative Learning Environments in STEM Higher Education pp. 55 – 77 (2021), https://api.semanticscholar.org/CorpusID:232206002