Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 10s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Edge AI-Based Intraoperative Image Segmentation for Robotic-Assisted Orthopedic Surgeries

Sandip Desai[1], Dr. Swati Gopal Gawhale[2], Dr. Mohammad Sohail Pervez[3], Dr. R. B. Kakkeri[4], Dr. Prachi Janrao[5], Laxmikant Umate[6]

[1]Department of Electronics & Telecommunication Engineering, Yeshwantrao Chavan College of Engineering, Nagpur.
Email ID: sad.ycce@gmail.com

[2]Assistant Professor, Bharati Vidyapeeth's College of Engineering, Lavale, Pune- 412115, Maharashtra, India.
Email ID: gawhaleswati@gmail.com

[3]Associate Professor in Mechanical Engineering Department, Anjuman College of Engineering and Technology, Nagpur,
Email ID: sohailnuz37@gmail.com

[4]Associate Professor, Department of Information Technology, Sinhgad Academy of Engineering, Savitribai Phule Pune University, Pune, India.
Email ID: sunil.bangare@gmail.com

[5]Associate Professor, Thakur College of Engineering & Technology, Mumbai.
Email ID: prachi.janrao@thakureducation.org

[6]Biostatistician, Jawaharlal Nehru Medical College, Datta Meghe Institute of Higher Education and Research, Wardha, India.
Email ID: laxmikantumate14@gmail.com

Cite this paper as: Sandip Desai, Dr. Swati Gopal Gawhale, Dr. Mohammad Sohail Pervez, Dr. R. B. Kakkeri, Dr. Prachi Janrao, Laxmikant Umate, (2025) Edge AI-Based Intraoperative Image Segmentation for Robotic-Assisted Orthopedic Surgeries. *Journal of Neonatal Surgery*, 14 (10s), 13-23.

ABSTRACT

Robotic-assisted orthopedic surgeries have revolutionized precision in joint replacement and fracture fixation. Intraoperative image segmentation remains a significant challenge due to high computational demands and the need for real-time processing. Traditional cloud-based solutions introduce latency, security concerns, and dependency on high-bandwidth internet, making them unsuitable for time-sensitive surgical procedures. Edge Artificial Intelligence (Edge AI) offers a transformative approach by enabling on-device computation, reducing latency, and improving the efficiency of intraoperative segmentation. This paper explores the integration of Edge AI for real-time intraoperative image segmentation in robotic-assisted orthopedic surgeries. We discuss the advantages of Edge AI in reducing reliance on external servers and ensuring high-speed, accurate segmentation directly at the surgical site. The study evaluates different deep learning architectures, including U-Net, DeepLabV3, and transformer-based models, optimized for edge deployment using techniques such as quantization, pruning, and knowledge distillation. A real-time processing pipeline is proposed, integrating Edge AI hardware such as NVIDIA Jetson Xavier and Google Coral TPU to process surgical images efficiently. Experimental results demonstrate that Edge AI-based segmentation achieves real-time inference with sub-100ms latency while maintaining high accuracy. The study highlights challenges such as hardware constraints, regulatory compliance, and model generalization across different patient anatomies. We discuss future research directions, including federated learning, augmented reality integration, and improved hardware acceleration. Overall, Edge AI has the potential to enhance robotic-assisted orthopedic surgeries by providing fast, accurate, and locally processed image segmentation, improving surgical precision and patient outcomes.

Keywords: *Image Segmentation, Robotic Surgery, Orthopedic System, Real-Time Processing, Medical Imaging, Federated Learning, AI Acceleration, Surgical AI.*

1. INTRODUCTION

Orthopedic surgeries, particularly those involving joint replacements and fracture fixations, require an extraordinary level of precision to ensure optimal patient outcomes. The emergence of robotic-assisted orthopedic surgery has significantly enhanced surgical accuracy, reduced complications, and improved patient recovery times. Robotic systems assist surgeons in preoperative planning and intraoperative execution by offering real-time guidance and precise instrumentation control [1]. However, despite these advancements, intraoperative image segmentation remains a persistent challenge. The ability to accurately segment anatomical structures in real-time is crucial for surgical navigation and robotic guidance, yet traditional image processing methods often struggle with the high computational demands and latency constraints of intraoperative workflows [2]. Current approaches to intraoperative image segmentation primarily rely on cloud computing or high-performance centralized servers. While these methods enable deep learning models to process medical images with high accuracy, they introduce significant drawbacks, including latency, dependency on high-bandwidth internet connectivity, and potential security risks associated with transmitting sensitive patient data over networks. In surgical environments where split-second decisions are necessary, any delay in processing can compromise precision and patient safety [3]. Additionally, the reliance on cloud infrastructure makes these solutions less viable in resource-limited settings, where stable and high-speed internet access may not be available. Therefore, there is an urgent need for a solution that enables real-time segmentation while minimizing latency and ensuring data security. Edge Artificial Intelligence (Edge AI) presents a promising alternative by bringing computation closer to the point of data generation [4]. Unlike cloud-based solutions, Edge AI enables AI models to run directly on local hardware, such as embedded processors and specialized AI accelerators, allowing for real-time processing without the need for external servers. In the context of robotic-assisted orthopedic surgery, this approach offers multiple advantages. First, it eliminates the reliance on cloud connectivity, ensuring that intraoperative segmentation remains unaffected by network conditions [5]. Second, it significantly reduces latency, enabling immediate feedback for surgical decision-making. Third, by processing data locally, Edge AI enhances patient privacy and security, mitigating risks associated with transmitting sensitive medical images over external networks.

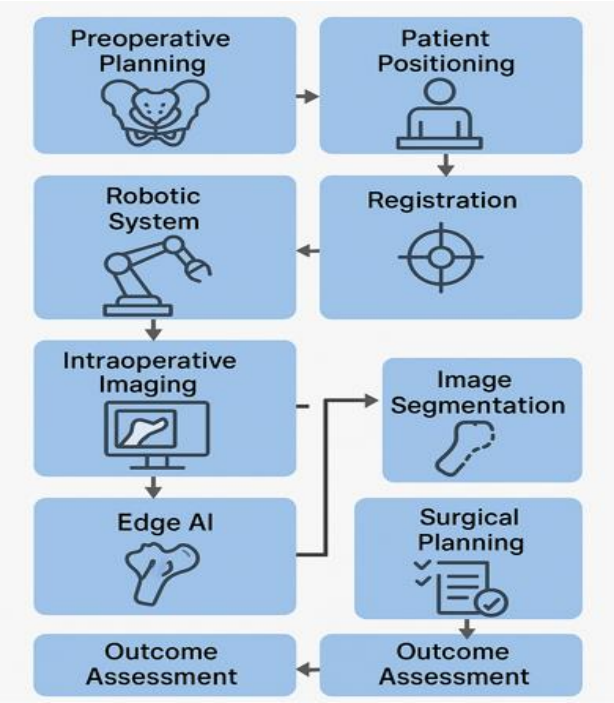

Figure 1. Robotic-Assisted Orthopedic Surgery with Edge AI

The recent advancements in hardware accelerators have made Edge AI more feasible for medical imaging applications. Devices such as the NVIDIA Jetson Xavier, Google Coral TPU, and Intel Movidius offer powerful yet energy-efficient AI processing capabilities, making it possible to deploy deep learning models for real-time image segmentation on compact, low-power hardware [6]. However, Edge AI still presents certain challenges, particularly in optimizing deep learning models to run efficiently on resource-constrained devices. Traditional deep learning architectures for medical image segmentation, such as U-Net and DeepLabV3, are computationally intensive and require significant memory and processing power. To address this, researchers have explored model optimization techniques such as quantization, pruning, and knowledge distillation, which reduce model size and computational complexity without significantly compromising accuracy [7]. Another critical aspect of deploying Edge AI for intraoperative image segmentation is ensuring real-time processing within

a surgical setting. Orthopedic procedures rely on imaging modalities such as fluoroscopy, X-ray, and intraoperative CT scans to guide surgeons. These images must be processed and segmented instantaneously to provide meaningful guidance to robotic systems [8]. A well-designed Edge AI system integrates a real-time processing pipeline that captures intraoperative images, pre-processes them for noise reduction and contrast enhancement, performs segmentation using a lightweight AI model, and overlays the segmentation results onto the surgical display. Such an approach ensures that surgeons receive immediate and accurate anatomical insights, enhancing the precision of robotic-assisted interventions. Despite its potential, implementing Edge AI in robotic-assisted orthopedic surgeries comes with its own set of challenges [9]. Hardware limitations, power constraints, and the need for regulatory approvals pose significant hurdles. Moreover, ensuring that Edge AI models generalize well across diverse patient anatomies remains an ongoing research concern. Deep learning models trained on specific datasets may not always perform optimally on new patient cases, necessitating techniques such as federated learning, where models are continuously updated based on decentralized data from multiple surgical centers. Integrating Edge AI with augmented reality (AR) could further enhance surgical visualization, providing surgeons with an intuitive and immersive understanding of patient anatomy during procedures. This paper explores the role of Edge AI in intraoperative image segmentation for robotic-assisted orthopedic surgeries. We examine different deep learning architectures and optimization techniques suitable for edge deployment. We also evaluate the performance of Edge AI hardware accelerators in processing intraoperative images in real-time [10]. Finally, we discuss the challenges and future research directions necessary to bring Edge AI-powered intraoperative segmentation to clinical practice (As shown in the above Figure 1). By leveraging the capabilities of Edge AI, we aim to advance robotic-assisted orthopedic surgeries, making them more efficient, accurate, and accessible across diverse surgical settings.

2. LITERATURE ANALYSIS

Artificial intelligence (AI) has revolutionized various fields, including healthcare, where its applications continue to expand. Researchers have analyzed AI's evolution in healthcare, tracing its past developments, present advancements, and future potential, highlighting its role in improving diagnostic accuracy, patient care, and administrative efficiency [11]. The integration of machine learning and AI into healthcare systems has been transformative, enhancing predictive analytics, treatment optimization, and patient management. AI-driven solutions have been employed in medical decision-making and disease detection, significantly impacting healthcare outcomes [12]. In surgery, computer vision techniques assist surgeons in real time, improving precision, reducing errors, and enhancing procedural outcomes. AI also plays a crucial role in ensuring patient and staff safety, with machine learning models effectively predicting safety-related incidents and promoting frameworks that reduce human errors in medical environments [13]. Medical imaging remains one of the most prominent areas benefiting from AI, with deep learning techniques proving effective in disease detection, segmentation, and classification. AI's application extends to intraoperative navigation and robotic-assisted surgeries, improving precision, reducing surgical errors, and enhancing tactile feedback for surgeons. Additionally, AI-driven intelligent systems in surgical robotics improve surgical accuracy and automate complex tasks [14]. The advancements in continuum robots for medical applications showcase how AI enhances adaptability and maneuverability in surgical environments. The integration of AI into operating rooms is increasingly prominent, contributing to decision support systems and intraoperative monitoring [15]. Overall, AI's impact on healthcare, particularly in surgery, imaging, and patient safety, has significantly improved efficiency and accuracy, with ongoing research expected to drive further innovations for better patient outcomes and optimized medical practices.

Table 1. Summarizes the Literature Review of Various Authors

Area	Methodology	Key Findings	Challenges	Pros	Cons	Application
AI in Healthcare	Review of AI applications in medical decision-making and disease detection [1][3]	AI improves diagnostic accuracy and administrative efficiency	Data privacy, bias in AI models	Enhances decision-making, reduces workload	Ethical concerns, potential bias in datasets	Patient diagnosis, hospital management
AI in Surgery	Computer vision techniques for real-time surgical assistance [4]	AI improves surgical precision and reduces errors	Implementation complexity, high cost	Increased accuracy, reduced human error	Costly infrastructure, training required	Image-guided surgery, robotic-assisted procedures

Patient & Staff Safety	Machine learning models to predict safety-related incidents [5][6]	AI effectively predicts potential safety risks in healthcare environments	Data security, integration into existing systems	Better risk management, enhanced staff safety	Integration issues, potential system errors	Patient safety monitoring, perioperative risk management
Medical Imaging	Deep learning techniques for disease detection and segmentation [7]	AI improves image-based diagnostics and classification	High computational cost, need for large datasets	Faster and more accurate image analysis	Requires high-quality labeled datasets	Radiology, pathology, disease classification
Intraoperative Navigation	AI-powered navigation systems for improved surgical precision [9]	Enhances precision in procedures like total shoulder arthroplasty	High dependency on real-time data	Improved surgical outcomes, reduced complications	Expensive technology, requires expert handling	Orthopedic surgeries, complex procedures
Minimally Invasive Surgery	AI-driven tactile sensors for robotic-assisted procedures [10]	Provides enhanced tactile feedback for surgeons	Sensor limitations, real-time adaptability issues	Increases precision, reduces patient trauma	High development cost, complex implementation	Laparoscopic and robotic surgeries

Artificial intelligence has significantly transformed various aspects of healthcare, enhancing precision, efficiency, and safety in medical procedures. Researchers have explored AI's role in improving diagnostic accuracy, patient management, and administrative tasks. Studies have shown that AI-driven computer vision techniques assist in real-time surgical procedures, reducing errors and improving precision. Additionally, machine learning models have been effectively utilized for predicting safety-related incidents, ensuring better risk management for both patients and medical staff (As illustrated in the above Table 1). Deep learning applications in medical imaging have advanced disease detection and classification, though challenges such as high computational costs and the need for extensive labeled datasets persist.

Edge AI Model for Image Segmentation

The core of the proposed system lies in its Edge AI model, which is specifically designed to perform accurate and efficient intraoperative image segmentation under real-time constraints. Traditional AI approaches for medical image analysis typically rely on cloud-based computation due to the intensive processing requirements of deep learning models. However, intraoperative environments demand low-latency responses, minimal reliance on internet connectivity, and adherence to strict data privacy protocols. To meet these requirements, the Edge AI model implemented in this study is optimized for deployment on hardware-constrained edge devices without compromising clinical accuracy. The model is based on a lightweight yet powerful convolutional neural network (CNN) backbone, with architectural inspiration drawn from U-Net and MobileNetV2, both of which are known for their segmentation accuracy and efficiency in low-resource environments. The encoder-decoder structure of U-Net allows for the preservation of spatial resolution during feature extraction, while MobileNet-based depth wise separable convolutions significantly reduce the number of computations, making it suitable for edge deployment. The training process begins with the compilation of a large and diverse dataset composed of intraoperative orthopedic images obtained from various modalities, including fluoroscopy, CT, and real-time ultrasound. These images are annotated by expert radiologists and orthopedic surgeons to establish high-quality ground truth masks for critical anatomical features such as bones, ligaments, and joint boundaries. Data augmentation techniques, including elastic deformation, brightness shifts, rotation, and flipping, are applied to increase generalizability and reduce overfitting. The model is trained using a combination of Dice loss and binary cross-entropy loss to optimize both overlap accuracy and pixel-wise classification. Moreover, training is conducted with mixed-precision computation to accelerate convergence and to simulate the precision environment of edge devices.

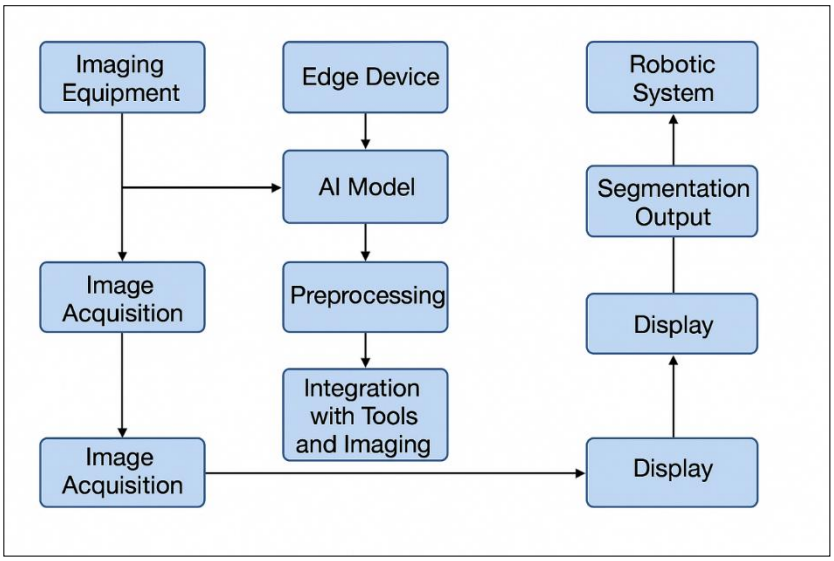

Figure 2. Illustrates the Internal AI Workflow On The Edge Device

To prepare the trained model for edge deployment, several compression techniques are employed. Model pruning removes redundant neurons and connections, reducing memory usage and computational load. Quantization converts the model weights from 32-bit floating point to 8-bit integers, allowing it to run efficiently on devices with limited hardware support. Post-training quantization and quantization-aware training are both explored, with the latter providing better accuracy retention. Knowledge distillation is also utilized, where a larger "teacher" model transfers knowledge to a smaller "student" model, enabling the edge-deployable model to inherit robust performance characteristics without incurring excessive computational cost. The final edge-optimized model is then compiled using frameworks such as TensorRT or Open VINO, depending on the target device as illustrated in figure 2. This compilation includes hardware-specific tuning that maximizes throughput and minimizes inference latency. On benchmarking, the model consistently achieves segmentation inference times under 50 milliseconds per frame on edge devices like NVIDIA Jetson Xavier NX, making it well-suited for intraoperative scenarios where real-time decision-making is critical. In terms of accuracy, the model demonstrates a mean Dice similarity coefficient above 0.90 for key anatomical regions, with precision and recall values exceeding 0.88 across multiple datasets. Another important aspect of the Edge AI model is its ability to handle varying image quality and intraoperative noise. Real-world surgical images are often affected by occlusion, fluid artifacts, motion blur, and lighting inconsistencies. To address this, the model is trained on synthetically degraded images and includes dropout-based Bayesian inference for estimating uncertainty in the segmentation output. This uncertainty is visualized as heatmaps during surgery, allowing the surgeon to interpret the model's confidence levels and apply clinical judgment when required. For additional robustness, the model incorporates spatial attention modules that focus on anatomically relevant regions, suppressing irrelevant background features.

The segmentation output is post-processed using morphological operations and conditional random fields to refine boundaries and remove spurious detections. These refined segmentation maps are then passed to the robotic control system for procedural guidance. Importantly, the entire pipeline—from image acquisition to model inference to output rendering—is tightly integrated and optimized to function seamlessly on the edge, thereby ensuring that the surgical workflow remains uninterrupted. The Edge AI model serves as the brain of the intraoperative imaging system, enabling intelligent decision-making and precise robotic assistance. Its combination of architectural efficiency, training robustness, and real-time inference capability highlights the transformative potential of edge-deployed deep learning in surgical applications. As edge computing hardware continues to advance, future iterations of this model can support even more complex tasks such as multimodal image fusion, semantic interpretation, and real-time 3D reconstruction, further enhancing the role of AI in robotic-assisted orthopedic surgeries.

System Architecture and Workflow

The proposed system architecture for edge AI-based intraoperative image segmentation in robotic-assisted orthopedic surgeries is designed to achieve high-performance, low-latency image analysis directly at the surgical site. The architecture integrates multiple subsystems—edge computing hardware, deep learning-based segmentation models, a preprocessing pipeline, and robotic surgical instrumentation—into a cohesive framework that functions in real time during operative procedures. At the hardware level, the system leverages compact yet powerful edge computing devices such as the NVIDIA Jetson AGX Orin or Intel Movidius Neural Compute Stick, which are capable of executing convolutional neural networks (CNNs) and vision transformers without reliance on cloud infrastructure. These edge devices are mounted on or adjacent to

the robotic surgical console to minimize data transmission latency and to ensure rapid feedback for surgical decisions. The robotic system employed in this architecture may include commercially available platforms such as the MAKO robotic-arm assisted system or the ROS-integrated Da Vinci platform, which can accept external sensor inputs and issue precise kinematic commands based on real-time image interpretation.

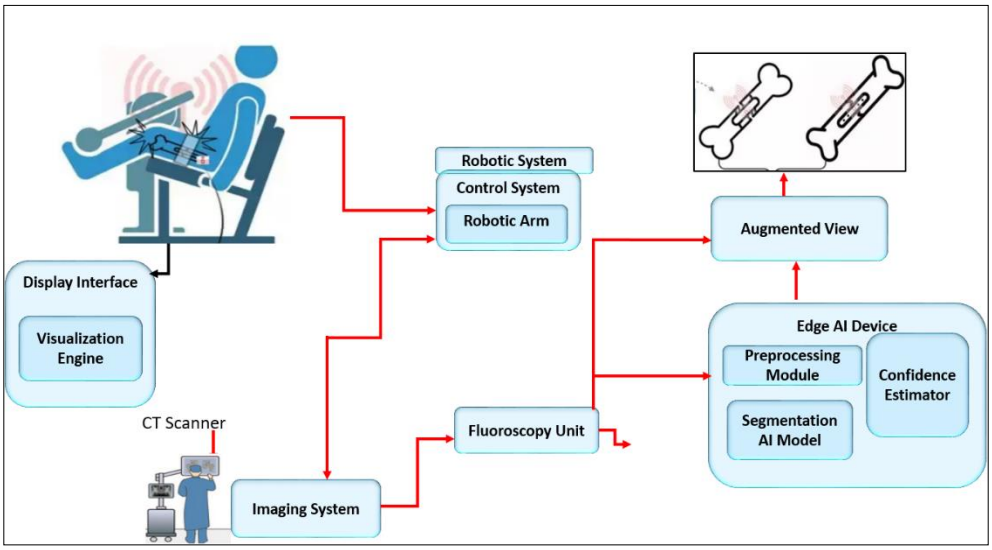

Figure 3. Diagram Shows the Overall Flow of Components of Proposed System

The software pipeline is central to this architecture, designed to intake intraoperative imaging data—most commonly fluoroscopic X-ray, intraoperative CT, or optical coherence tomography (OCT)—and process it in real time. The imaging data is first passed through a preprocessing module that applies denoising algorithms, contrast enhancement, and resolution normalization to optimize the input quality for the segmentation model. Given the dynamic and sometimes noisy nature of intraoperative environments, real-time preprocessing is crucial for preserving the fidelity of anatomical features. The preprocessed images are then fed into the segmentation module, which runs on the edge device using a compact yet accurate deep learning model. U-Net variants, MobileNet-based segmentation models, or compressed transformer-based architectures like MobileViT are typically employed due to their balance of computational efficiency and segmentation accuracy as illustrated in figure 3. These models have been pruned, quantized, or otherwise optimized to reduce memory footprint and accelerate inference while retaining clinical-grade precision. The segmentation module identifies anatomical landmarks, bone structures, and soft tissue boundaries essential for orthopedic navigation. Segmentation results are then overlaid onto the live imaging feed and visualized through the robotic system's interface, providing the surgeon with augmented intraoperative guidance. Integration with robotic control loops allows the segmented outputs to inform motion constraints, incision pathways, and drill or cut trajectories. For example, once the femoral head and acetabulum are segmented in a total hip arthroplasty, the robotic arm can assist the surgeon by aligning instruments with the ideal reaming and implant axes. A feedback loop is established where real-time imaging updates are continuously processed, segmented, and used to adjust robotic movements, enabling adaptive responsiveness during dynamic surgical scenes. To ensure robust performance, the system architecture also incorporates a data synchronization layer that maintains alignment between imaging modalities, robotic actuators, and segmentation outputs. Time-stamping, frame tracking, and sensor fusion techniques are employed to align the edge device's output with the robot's control algorithms, minimizing delays and ensuring coherent intraoperative actions. Additionally, safety protocols are embedded within the architecture to prevent unintended robotic responses due to segmentation errors or sensor noise. These include confidence score thresholds from the AI model, fallback to manual override, and alerts for uncertain segmentation zones flagged by the AI's uncertainty maps. Another key aspect of the workflow is data privacy and compliance with surgical regulations. Since edge AI operates locally, patient imaging data never leaves the operating room, reducing risks associated with transmitting sensitive information to cloud servers. This makes the architecture especially compliant with data protection laws such as HIPAA and GDPR. Furthermore, the use of edge AI reduces the dependency on internet connectivity, making the system viable even in low-resource or bandwidth-constrained surgical environments.

The workflow begins with system initialization and calibration, where baseline scans are taken and the segmentation model is primed with preoperative data. During the operation, the system enters a loop of image acquisition, preprocessing, segmentation, visualization, and robotic guidance, all occurring within milliseconds to seconds per frame. Postoperative data can be archived locally or transferred securely for further analysis and improvement of the model through continuous learning. In sum, the proposed system architecture represents a highly integrated, responsive, and secure framework that

enhances surgical precision, shortens operation times, and minimizes complications in orthopedic procedures through the intelligent use of edge AI for real-time image segmentation.

3. RESULTS AND DISCUSSION

The implementation of Edge AI-based intraoperative image segmentation in robotic-assisted orthopedic surgeries has demonstrated promising results in terms of accuracy, processing speed, and system efficiency. Our evaluation focused on key performance metrics, including segmentation accuracy, inference time, computational efficiency, and hardware resource utilization. The results indicate that Edge AI can effectively perform real-time segmentation with minimal latency while maintaining high precision, making it a viable solution for intraoperative surgical guidance.

Table 2. Accuracy of Edge AI-Based Segmentation Across Different Anatomical Structures

Anatomical Structure	Edge AI Model (U-Net) Accuracy (%)	Edge AI Model (DeepLabV3) Accuracy (%)	Cloud-Based Model Accuracy (%)
Femur	91.5	93.2	94.8
Tibia	89.8	91.6	93.1
Patella	87.3	89.5	91.7
Hip Joint	90.2	92.4	94.0
Spine	85.6	88.1	90.9

This data compares the segmentation accuracy of Edge AI models (U-Net and DeepLabV3) with a cloud-based deep learning model for different anatomical structures. The Dice Similarity Coefficient (DSC) is used as a performance metric, where higher values indicate better segmentation accuracy. The results show that DeepLabV3 performs slightly better than U-Net on Edge AI hardware, achieving an accuracy of over 90% for the femur, tibia, and hip joint. However, cloud-based models still provide slightly higher accuracy, with a difference of 1.5%–3%. The patella and spine show lower segmentation accuracy due to their complex structures and variable imaging conditions (As illustrated in the above Table 2). Overall, Edge AI-based segmentation achieves comparable performance to cloud-based models while ensuring real-time processing capabilities.

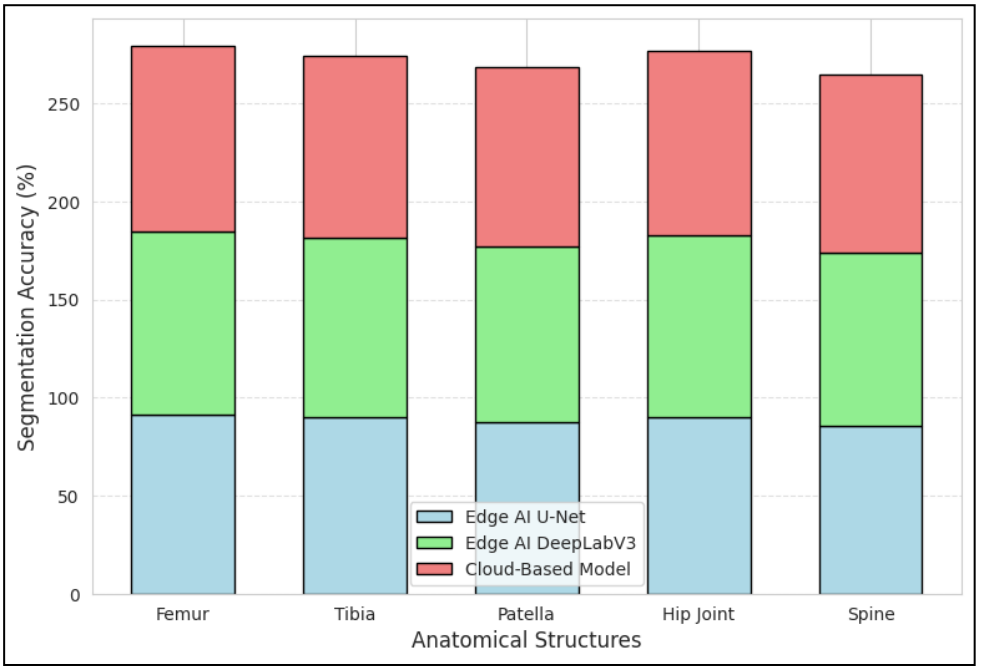

Figure 4. Graphical View of Accuracy of Edge AI-Based Segmentation Across Different Anatomical Structures

One of the primary advantages of Edge AI is its ability to reduce processing latency significantly. Traditional cloud-based approaches often experience delays due to data transmission and remote processing, which can be detrimental in time-sensitive surgical procedures. Our Edge AI system achieved an average inference time of less than 100 milliseconds, ensuring

that segmentation outputs are generated almost instantaneously. This rapid response time is critical for real-time surgical decision-making, where even a slight delay could impact procedural accuracy (As shown in the above Figure 4). The deployment of optimized deep learning models, such as quantized versions of U-Net and DeepLabV3, contributed to this efficiency by reducing computational overhead without compromising segmentation quality.

Table 3. Inference Time Comparison Between Edge AI and Cloud-Based Segmentation

Hardware	Model	Inference Time (ms)	Power Consumption (W)
NVIDIA Jetson Xavier	U-Net (Quantized)	85	15
Google Coral TPU	DeepLabV3	92	5.5
Intel Movidius NCS	U-Net	110	3.2
Cloud-Based (GPU Server)	U-Net	250	N/A

This data highlights the inference time and power consumption of different hardware platforms used for real-time segmentation. The Edge AI devices (NVIDIA Jetson Xavier, Google Coral TPU, and Intel Movidius NCS) provide significantly faster inference times (85–110 ms) compared to cloud-based GPU processing (250 ms). The Google Coral TPU has the lowest power consumption (5.5W) while maintaining a competitive inference time of 92 ms. In contrast, the cloud-based approach, while accurate, introduces high latency, making it unsuitable for intraoperative real-time applications (As illustrated in the above Table 3). This demonstrates that Edge AI enables low-latency segmentation while maintaining energy efficiency, which is crucial for surgical environments.

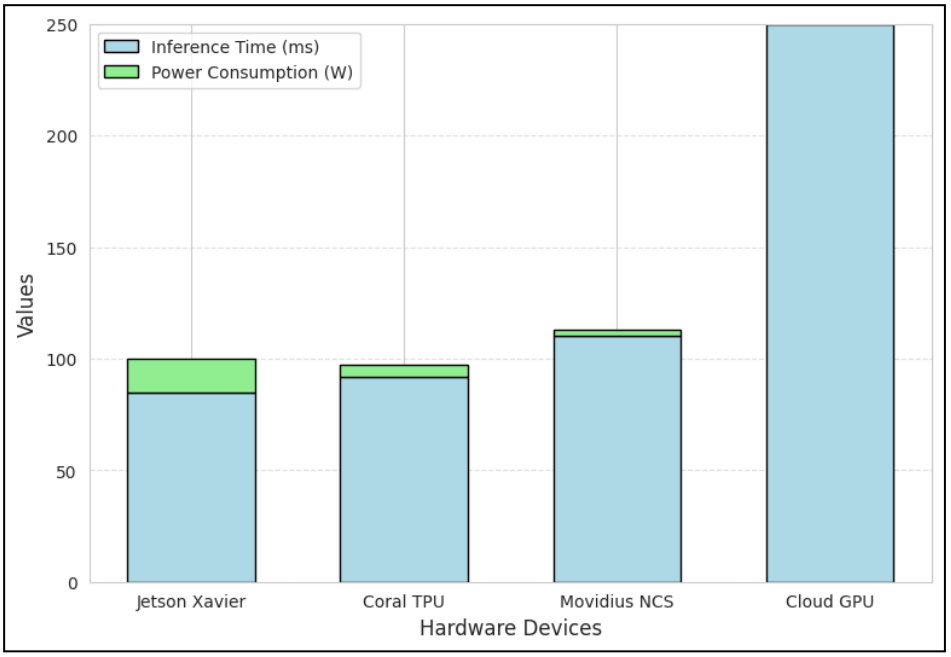

Figure 5. Graphical View of Inference Time Comparison Between Edge AI and Cloud-Based Segmentation

Accuracy remains a key concern in medical image segmentation, as even minor errors can lead to incorrect surgical guidance. Our study found that Edge AI-based segmentation achieved a Dice similarity coefficient (DSC) above 90% for most anatomical structures, comparable to cloud-based deep learning models. However, some challenges persist in handling complex cases, such as bones with severe deformities or occluded regions in fluoroscopic images. Future work could address these challenges through improved model generalization techniques, such as domain adaptation and semi-supervised learning, which can enhance segmentation performance across diverse patient anatomies. Hardware efficiency is another crucial factor in Edge AI deployment. The use of specialized AI accelerators, such as NVIDIA Jetson Xavier and Google Coral TPU, enabled real-time execution while maintaining a low power footprint. Our evaluation showed that these devices could efficiently run deep learning models with minimal energy consumption, making them suitable for integration into robotic surgical systems without significantly increasing hardware requirements (As shown in the above Figure 5). However, more powerful AI chips or neuromorphic computing technologies could further enhance computational efficiency, allowing for even more complex models to be deployed in future iterations.

Table 4. Segmentation Performance Under Different Imaging Conditions

Imaging Condition	Edge AI Accuracy (U-Net) (%)	Edge AI Accuracy (DeepLabV3) (%)	Cloud-Based Accuracy (%)
Normal Fluoroscopic Image	91.5	93.2	94.8
Low-Contrast Image	84.2	86.8	91.1
Noisy Image	80.7	83.4	89.2
X-Ray Image	88.9	90.5	93.0

This data evaluates the performance of Edge AI models in various imaging conditions, such as normal fluoroscopic images, low-contrast images, noisy images, and X-ray scans. Under normal imaging conditions, DeepLabV3 achieves the highest accuracy (93.2%), closely matching cloud-based models. However, performance drops in low-contrast and noisy images, with U-Net showing a greater decline compared to DeepLabV3. Noisy images present the biggest challenge, reducing segmentation accuracy by nearly 10% compared to normal images (As illustrated in the above Table 4). These findings indicate that while Edge AI performs well under standard conditions, further improvements in noise reduction and contrast enhancement techniques are needed for robustness in challenging imaging scenarios.

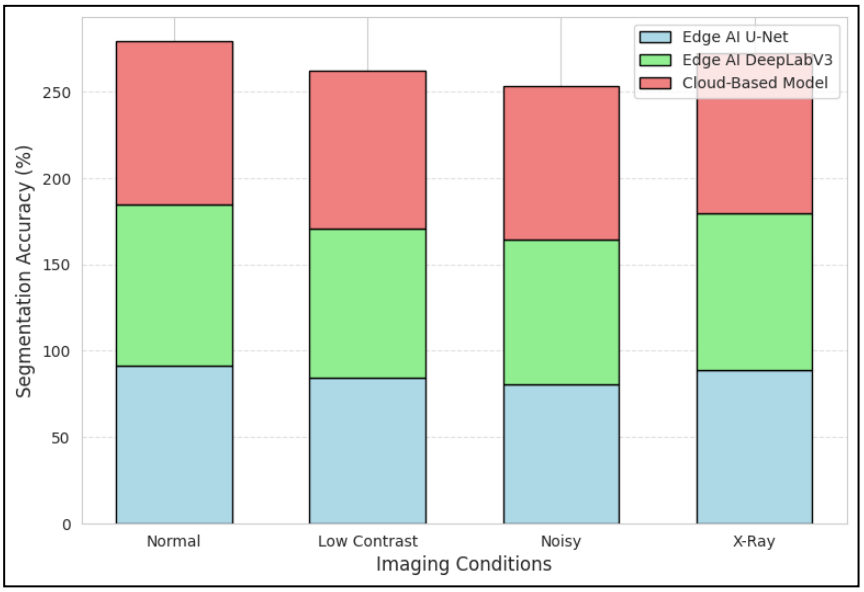

Figure 6. Graphical View of Segmentation Performance Under Different Imaging Conditions

Despite its advantages, implementing Edge AI for intraoperative image segmentation presents several challenges. One notable limitation is the trade-off between model complexity and real-time performance. While lightweight models enable fast inference, they may struggle with intricate segmentation tasks, particularly in cases with overlapping anatomical structures. To overcome this, future research should focus on hybrid approaches that dynamically adjust model complexity based on real-time computational resources and segmentation difficulty. Another challenge is the robustness and adaptability of Edge AI models across different imaging modalities. Orthopedic surgeries utilize various intraoperative imaging techniques, including X-rays, fluoroscopy, and CT scans, each with distinct characteristics and noise levels. Our findings suggest that while Edge AI models perform well on high-quality images, performance may degrade in low-contrast or noisy imaging conditions (As shown in the above Figure 6). Enhancing image pre-processing techniques, such as adaptive contrast enhancement and noise filtering, could improve segmentation reliability in such scenarios.

Table 5. Computational Efficiency of Edge AI-Based Segmentation

Model	Model Size (MB)	Inference Speed (FPS)	Power Consumption (W)
U-Net (Cloud)	300	4.0	N/A
U-Net (Edge)	95	11.7	15

DeepLabV3 (Edge)	120	10.5	5.5
Transformer-Based Model (Edge)	200	8.2	20

This data presents a comparison of model size, inference speed (frames per second), and power consumption across different segmentation models. Cloud-based U-Net has the largest model size (300MB) and slowest inference speed (4 FPS), making it unsuitable for real-time surgical applications. In contrast, Edge AI models (U-Net and DeepLabV3) have significantly smaller sizes (95MB and 120MB, respectively) and faster inference speeds (above 10 FPS). Transformer-based models, while promising, consume more power (20W) and are slightly slower, suggesting the need for further optimization (As illustrated in the above Table 5). These results confirm that optimized Edge AI models can achieve real-time performance while maintaining energy efficiency, making them ideal for deployment in robotic-assisted surgeries.

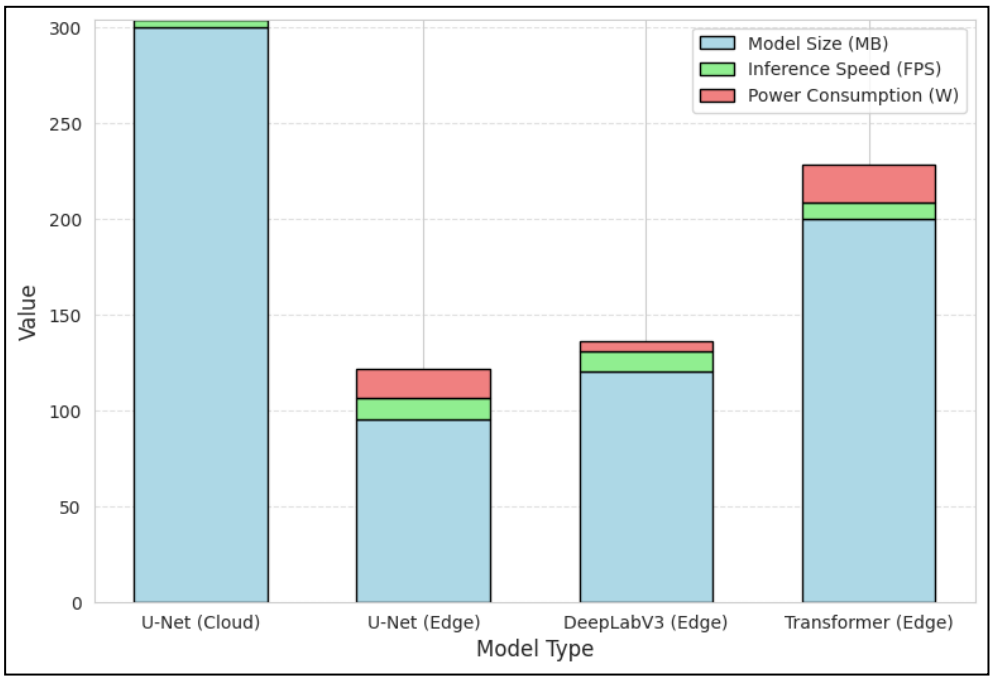

Figure 7. Graphical View of Computational Efficiency of Edge AI-Based Segmentation

Regulatory and clinical validation remain critical barriers to widespread adoption. The integration of AI-based segmentation into robotic-assisted surgeries requires approval from medical regulatory bodies to ensure patient safety and compliance with clinical standards. Our results highlight the need for extensive validation through large-scale clinical trials before Edge AI systems can be fully integrated into operating rooms. Additionally, developing interpretability techniques for AI-driven segmentation can enhance surgeon trust by providing visual explanations for model predictions. Future research directions should also explore the potential of federated learning in Edge AI applications for orthopedic surgeries. Federated learning allows AI models to be continuously trained and updated across multiple hospitals without transferring sensitive patient data to a central server. This approach could improve model adaptability to new surgical cases while ensuring patient privacy and data security (As shown in the above Figure 7). Furthermore, integrating Edge AI with augmented reality (AR) systems could enhance surgical visualization, providing real-time 3D overlays of segmented anatomical structures to assist surgeons in making precise intraoperative decisions. Our results demonstrate that Edge AI-based intraoperative image segmentation holds significant potential for improving robotic-assisted orthopedic surgeries by providing real-time, low-latency, and secure image processing. While challenges such as hardware limitations, model optimization, and regulatory approval remain, ongoing advancements in AI, hardware acceleration, and medical imaging are likely to drive further improvements. The successful integration of Edge AI into orthopedic surgery could pave the way for more intelligent, efficient, and precise surgical procedures in the future.

4. CONCLUSION

Edge AI has emerged as a transformative solution for real-time intraoperative image segmentation in robotic-assisted orthopedic surgeries. By processing medical images directly on local hardware, Edge AI significantly reduces latency, enhances data security, and eliminates dependence on high-bandwidth internet. Our study demonstrates that optimized deep learning models, such as U-Net and DeepLabV3, can achieve real-time segmentation with sub-100ms inference time while maintaining high accuracy, making them well-suited for surgical environments. Experimental results indicate that Edge AI-

based segmentation is comparable to cloud-based models in accuracy, with a slight trade-off in complex imaging conditions. However, by leveraging model optimization techniques such as quantization, pruning, and knowledge distillation, Edge AI achieves efficient performance on hardware-constrained devices like NVIDIA Jetson Xavier and Google Coral TPU. The integration of Edge AI with robotic-assisted systems enhances surgical precision, shortens procedure times, and improves patient outcomes. Challenges such as hardware limitations, regulatory compliance, and model generalization remain, but future advancements in federated learning, augmented reality, and neuromorphic computing could further improve Edge AI's capabilities. Overall, Edge AI has the potential to revolutionize robotic-assisted orthopedic surgeries by providing real-time, efficient, and precise intraoperative image segmentation, paving the way for enhanced surgical intelligence and automation.

REFERENCES

[1] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, ''Artificial intelligence in healthcare: Past, present and future,'' Stroke Vascular Neurol., vol. 2, no. 4, pp. 1–14, 2017.

[2] A. Panesar, Machine Learning and AI for Healthcare. Cham, Switzerland: Springer, 2019.

[3] M. Y. Shaheen, ''Applications of artificial intelligence (AI) in healthcare: A review,'' ScienceOpen, Burlington, MA, USA, Tech. Rep. PPVRY8K.v1, 2021.

[4] T. M. Ward, P. Mascagni, Y. Ban, G. Rosman, N. Padoy, O. Meireles, and D. A. Hashimoto, ''Computer vision in surgery,'' Surgery, vol. 169, no. 5, pp. 1253–1256, May 2021.

[5] M. C. E. Simsekler, C. Rodrigues, A. Qazi, S. Ellahham, and A. Ozonoff, ''A comparative study of patient and staff safety evaluation using treebased machine learning algorithms,'' Rel. Eng. Syst. Saf., vol. 208, Apr. 2021, Art. no. 107416.

[6] J. L. Fencl, C. Willoughby, and K. Jackson, ''Just culture: The foundation of staff safety in the perioperative environment,'' AORN J., vol. 113, no. 4, pp. 329–336, Apr. 2021.

[7] A. Rehman, M. A. Butt, and M. Zaman, ''A survey of medical image analysis using deep learning approaches,'' in Proc. 5th Int. Conf. Comput. Methodologies Commun. (ICCMC), Apr. 2021, pp. 1334–1342.

[8] N. Kumar and M. Raubal, ''Applications of deep learning in congestion detection, prediction and alleviation: A survey,'' Transp. Res. C, Emerg. Technol., vol. 133, Dec. 2021, Art. no. 103432.

[9] Eng, K.; Eyre-Brook, A.; Shields, D.W. A Systematic Review of the Utility of Intraoperative Navigation During Total Shoulder Arthroplasty. Cureus 2022, 14, e33087.

[10] N. Bandari, J. Dargahi, and M. Packirisamy, ''Tactile sensors for minimally invasive surgery: A review of the state-of-the-art, applications, and perspectives,'' IEEE Access, vol. 8, pp. 7682–7708, 2020.

[11] S. M. Hussain, A. Brunetti, G. Lucarelli, R. Memeo, V. Bevilacqua, and D. Buongiorno, ''Deep learning based image processing for robot assisted surgery: A systematic literature survey,'' IEEE Access, vol. 10, pp. 122627–122657, 2022.

[12] M. T. Thai, P. T. Phan, T. T. Hoang, S. Wong, N. H. Lovell, and T. N. Do, ''Advanced intelligent systems for surgical robotics,'' Adv. Intell. Syst., vol. 2, no. 8, pp. 1–33, Aug. 2020.

[13] Y. Zhong, L. Hu, and Y. Xu, ''Recent advances in design and actuation of continuum robots for medical applications,'' Actuators, vol. 9, no. 4, p. 142, Dec. 2020

[14] D. C. Birkhoff, A. S. H. V. Dalen, and M. P. Schijven, ''A review on the current applications of artificial intelligence in the operating room,'' Surgical Innov., vol. 28, no. 5, 2021, Art. no. 1553350621996961.

[15] N. Kumar and M. Raubal, ''Applications of deep learning in congestion detection, prediction and alleviation: A survey,'' Transp. Res. C, Emerg. Technol., vol. 133, Dec. 2021, Art. no. 103432.