pISSN:2349-5707 Volume:13,Issue 3, March -2026 eISSN:2349-5715

METHODS OF APPLYING ARTIFICIAL INTELLIGENCE IN SURGICAL DIAGNOSTICS

Maksudjanova Feruza Sayfullaevna
Senior Lecturer of the Academy of Nurses

Annotation. The integration of artificial intelligence into surgical diagnostics represents one of the most transformative developments in contemporary medicine. Over the past decade, machine learning, deep learning, and computer vision technologies have progressively permeated virtually every domain of surgical practice - from preoperative imaging interpretation and risk stratification to intraoperative decision support and postoperative outcome prediction. This comprehensive review synthesizes current evidence on the role of AI in surgical diagnostics, examining its theoretical foundations, clinical applications across multiple surgical specialties, diagnostic accuracy metrics, and inherent limitations.

Keywords: artificial intelligence, surgical diagnostics, machine learning, deep learning, convolutional neural networks, computer vision, preoperative assessment, intraoperative guidance, postoperative outcomes, clinical decision support.

The history of surgery is, in many respects, a chronicle of diagnostic innovation. From the advent of antiseptic technique in the nineteenth century to the introduction of laparoscopy in the twentieth, each technological leap has fundamentally reshaped how surgeons understand, interpret, and respond to pathological states within the human body. Today, at the threshold of a new era defined by computational intelligence, artificial intelligence stands poised to constitute the next landmark transition in surgical diagnostics.

“Artificial intelligence, broadly construed, refers to the capacity of computational systems to perform tasks that would ordinarily require human cognitive effort — tasks such as pattern recognition, probabilistic inference, natural language understanding, and adaptive learning” [1]. Within medicine, and surgery specifically, AI has emerged not merely as a tool for automating repetitive processes, but as a powerful analytical framework capable of detecting subtle diagnostic signals that may elude even the most experienced clinician.

The global burden of surgical disease remains substantial. According to estimates from the Lancet Commission on Global Surgery, approximately 313 million surgical procedures are required annually worldwide to address treatable conditions, yet access to safe and timely surgery remains inequitable. In this context, AI-augmented diagnostics carry a dual promise: improving the precision of diagnoses in resource-rich environments while simultaneously enabling scalable, cost-effective diagnostic support in settings where specialist expertise is scarce.

“The motivating rationale for this review is straightforward. Despite a rapidly expanding body of literature documenting AI's diagnostic capabilities across numerous surgical subspecialties, the field lacks a unified, critically rigorous synthesis that simultaneously addresses clinical performance, methodological quality, and translational barriers” [2]. Existing reviews tend to be either narrowly subspecialty-focused or insufficiently attentive to the distinction between laboratory validation and real-world clinical performance. This paper endeavors to fill that gap by providing an integrated, evidence-based appraisal of AI in surgical diagnostics suitable for both clinical researchers and practicing surgeons seeking to understand and evaluate emerging technologies.

The review is organized as follows. The second section establishes the conceptual and technical foundations of AI relevant to surgical diagnostics. The third section surveys clinical applications across major surgical domains. The fourth section critically appraises diagnostic accuracy data. The fifth section examines barriers to implementation. The sixth section addresses ethical and regulatory considerations. The seventh section outlines future research directions, and the conclusion synthesizes the principal findings.

pISSN:2349-5707 Volume:13,Issue 3, March -2026 eISSN:2349-5715

Machine learning, a subdiscipline of artificial intelligence, encompasses algorithms that learn statistical patterns from data to generate predictions or classifications without being explicitly programmed for each specific task. In the surgical diagnostic context, ML algorithms are trained on large datasets — typically comprising imaging studies, clinical records, genomic profiles, or operative video footage - and subsequently deployed to generate predictions on new, unseen cases.

Supervised learning, the most widely applied ML paradigm in surgical diagnostics, involves training models on labeled datasets in which each input is paired with a known output. The algorithm learns the mapping between input features and output labels by minimizing prediction error across training examples. Once trained, the model can be applied to new cases to produce probabilistic diagnoses. Prominent supervised learning algorithms include support vector machines (SVMs), random forests, gradient boosting machines, and, most notably in recent years, artificial neural networks.

“Unsupervised learning, by contrast, involves identifying latent structures within unlabeled data. Clustering algorithms, for instance, can segment patient populations into diagnostically meaningful subgroups based on multivariate clinical profiles, potentially revealing novel disease phenotypes that inform surgical decision-making. Reinforcement learning, a third paradigm, enables systems to learn optimal decision strategies through iterative interaction with an environment, and has been investigated for robotic surgical guidance, though its application in pure diagnostics remains nascent” [3].

Deep learning represents a particular class of machine learning architecture inspired by the hierarchical organization of the mammalian visual cortex. Deep neural networks consist of multiple processing layers — hence the qualifier "deep" - each of which learns increasingly abstract representations of the input data. The foundational architecture for image-based tasks is the convolutional neural network (CNN), which applies a series of learnable filters to input images to extract spatially invariant features such as edges, textures, and shapes.

CNNs have proven transformative in surgical imaging contexts. In a landmark demonstration, a CNN trained on dermoscopic images achieved diagnostic accuracy for melanoma comparable to board-certified dermatologists. Analogous results have subsequently been demonstrated across multiple imaging modalities relevant to surgery, including CT, MRI, ultrasound, endoscopy, and intraoperative video. The ResNet, VGG, DenseNet, and EfficientNet architectures are among the most widely deployed CNN variants in surgical diagnostic research, each offering distinct trade-offs between computational efficiency and representational capacity.

Beyond CNNs, transformer-based architectures — particularly the Vision Transformer (ViT) — have recently entered the surgical AI landscape. Originally developed for natural language processing tasks, transformers apply self-attention mechanisms to model long-range dependencies within data, and have demonstrated competitive or superior performance compared to CNNs on several medical image classification benchmarks. Multimodal architectures, capable of jointly processing imaging and non-imaging data (e.g., clinical histories, laboratory values), represent a further frontier of technical development with direct relevance to surgical diagnostics.

Natural language processing (NLP) enables AI systems to extract structured information from unstructured text, including operative reports, clinical notes, and radiology dictations. In surgical diagnostics, NLP has been applied to retrospectively mine electronic health records (EHRs) for diagnostic signals — identifying, for example, patients whose symptom documentation patterns are consistent with occult malignancy or surgical emergencies prior to formal clinical recognition. Large language models (LLMs) such as GPT-4 and its successors have demonstrated preliminary utility in synthesizing complex clinical narratives to support differential diagnosis generation, though their performance in specialized surgical contexts requires continued evaluation.

General surgery represents perhaps the most extensively studied domain for AI-assisted diagnostics, owing partly to the high prevalence of acute abdominal pathology and the clinical imperative for rapid, accurate diagnosis. Appendicitis, affecting approximately 8% of the global population over a lifetime, has attracted particular attention. Conventional diagnostic scoring systems such as the Alvarado and Appendicitis Inflammatory Response (AIR) scores are subject to interobserver variability and modest specificity. AI-driven approaches have sought to improve upon these benchmarks.

A comprehensive systematic review published in 2024 examined AI and ML models for emergency surgical conditions and found accuracy rates ranging from 72% to 98% across various applications, including appendicitis diagnosis, cholecystitis detection, and acute abdominal pain triage. Machine learning models incorporating clinical variables (white blood cell count, C-reactive protein, pain localization, and patient demographic data) demonstrated superior discriminative performance compared to conventional scoring systems, with area under the curve (AUC) values consistently exceeding 0.85 in multiple independent validation cohorts.

“Imaging-based AI diagnostics for abdominal pathology have shown equally compelling results. CNN models trained on CT scans to identify appendiceal inflammation have achieved sensitivity and specificity values in the range of 90–95%, comparable to experienced radiologists, while substantially reducing interpretation time. Similarly, AI-assisted ultrasound interpretation for cholecystitis diagnosis has demonstrated the capacity to detect characteristic features - gallbladder wall thickening, pericholecystic fluid, and sonographic Murphy's sign - with high reproducibility, raising the prospect of AI-supported point-of-care ultrasound in resource-limited settings” [4].

The intersection of AI and oncological surgery is among the most clinically consequential areas of current research. Preoperative tumor characterization — determining malignancy grade, vascular involvement, lymph node status, and resectability - directly governs surgical planning and has profound implications for patient outcomes.

Radiomics, a technique that extracts large numbers of quantitative features from medical images, has emerged as a pivotal AI-adjacent methodology in oncological surgical diagnostics. By applying feature extraction algorithms to CT, MRI, or PET images, radiomics models can generate tumor "fingerprints" that encode information invisible to the naked eye, including internal heterogeneity, shape irregularity, and texture complexity. These radiomic features, when integrated with ML classifiers, have enabled preoperative prediction of tumor biology with diagnostic accuracy that rivals — and in some cases exceeds — conventional biopsy-based methods.

In colorectal cancer surgery, deep learning models analyzing CT colonography have achieved polyp detection rates exceeding 90%, with false-positive rates substantially lower than those of unassisted colonoscopy. For hepatocellular carcinoma, AI-driven MRI analysis has demonstrated sensitivity of 88–94% in distinguishing malignant from benign hepatic lesions, with implications for determining operative candidacy and approach. Intraoperative margin assessment — determining whether excised tissue is free of malignant cells — represents a further diagnostic challenge in which AI-assisted fluorescence imaging and computational pathology have shown early but promising results.

Neurosurgery presents uniquely demanding diagnostic challenges, given the anatomical complexity of the central nervous system and the severe consequences of diagnostic error. AI has been applied across multiple neurosurgical diagnostic tasks, with particularly notable results in the interpretation of neuroimaging and the grading of brain tumors.

ML algorithms analyzing MRI sequences have demonstrated high accuracy in detecting and classifying intracranial tumors, including gliomas, meningiomas, and metastases. A recent systematic review and meta-analysis reported an overall AI accuracy of 88.5% for diagnosing lumbar spinal stenosis, with deep learning models outperforming traditional machine learning models (89.2% vs. 86.3%). For brain tumor grading - distinguishing low-grade from high-grade gliomas, a clinically pivotal determination - CNN-based MRI analysis has achieved accuracy comparable to expert neuroradiologists in multiple independent studies.

pISSN:2349-5707 Volume:13,Issue 3, March -2026 eISSN:2349-5715

Spinal surgery has similarly benefited from AI diagnostic tools. Automated vertebral segmentation algorithms enable precise preoperative planning of spinal instrumentation, reducing operative time and the risk of malpositioned implants. Computer-assisted detection of spinal pathology on plain radiographs — historically a domain of high interobserver variability — has been shown to improve both sensitivity and consistency of diagnosis when AI serves as a secondary reader. Robotic-assisted neurosurgical procedures, while primarily a technical rather than diagnostic innovation, depend critically on AI-driven real-time imaging interpretation for safe execution.

"Cardiac surgery presents a distinctive diagnostic landscape in which preoperative risk stratification is as important as anatomical diagnosis. Conventional risk scoring systems — including EuroSCORE II and the Society of Thoracic Surgeons (STS) risk calculator — perform reasonably well at the population level but demonstrate limited accuracy for individual patient predictions. ML-based risk stratification models, trained on large cardiac surgery registries, have consistently demonstrated superior individual-level predictive performance compared to conventional scores, particularly for rare but catastrophic outcomes such as stroke and renal failure" [5].

Echocardiographic analysis represents a natural domain for AI application, given the volume and complexity of echocardiographic data generated in the preoperative cardiac workup. Deep learning models trained on transthoracic echocardiography have automated the measurement of ejection fraction, valve orifice area, and chamber dimensions with accuracy comparable to expert cardiologists, and at substantially reduced analysis time. AI-assisted interpretation of coronary CT angiography has further demonstrated the capacity to detect hemodynamically significant coronary lesions with sensitivity exceeding 90%, providing a non-invasive alternative to diagnostic catheterization in selected patients.

"In emergency surgical settings, the speed and accuracy of diagnosis are directly linked to patient survival. AI diagnostic tools have been investigated for a range of emergency surgical presentations, with a consistent finding of superiority over conventional triage instruments" [6]. A 2024 systematic review found that AI models showed superior performance in acute abdominal pain triage and risk assessment compared to conventional methods, suggesting substantial potential to enhance clinical decision-making in emergency surgical settings.

Trauma surgery represents a particularly high-stakes application domain. AI algorithms analyzing whole-body CT scans in polytrauma patients have demonstrated the ability to prioritize injuries for surgical attention, estimate blood loss, and predict the need for massive transfusion, supporting both individual clinical decisions and broader resource allocation. Sepsis recognition - a time-critical diagnosis with direct surgical implications in cases of intra-abdominal infection - has been significantly improved by ML-based early warning systems capable of detecting sepsis hours before conventional clinical recognition.

"The construction of high-quality surgical AI training datasets encounters structural challenges that are more severe than in many other domains of medical AI" [7]. Surgical imaging data — particularly operative video — is rarely systematically archived outside of dedicated research programs, and when it is archived, it lacks the standardized metadata and annotation structures required for ML model training. Operative video annotation requires specialized surgical expertise and is extraordinarily time-intensive, creating a bottleneck in training data availability that cannot be readily resolved by conventional crowdsourcing approaches.

Federated learning — in which AI models are trained across geographically distributed datasets without centralizing patient data — offers a technically promising framework for aggregating surgical training data while preserving institutional data sovereignty and patient privacy. "Multi-institutional federated learning pilots in surgical AI have demonstrated the feasibility of training models on distributed datasets with only modest performance degradation

pISSN:2349-5707 Volume:13,Issue 3, March -2026 eISSN:2349-5715

compared to centralized training, suggesting that federated approaches may be pivotal for developing the large, diverse datasets required for generalizable surgical AI diagnostics”[8].

Distribution shift — the systematic divergence between the statistical properties of AI training data and deployment-environment data — is a pervasive challenge in clinical AI that is particularly acute in surgery, where operative technique, imaging protocols, patient demographics, and disease epidemiology vary substantially across institutions, geographic regions, and healthcare systems.

Algorithmic bias manifests when these distributional disparities produce differential diagnostic performance across patient subgroups, potentially exacerbating existing healthcare inequities.

The surgical AI field has not yet systematically characterized the demographic scope of AI model validation — a deficiency identified in multiple systematic reviews. Prospective validation studies conducted across demographically diverse, internationally representative cohorts are an essential prerequisite for confident claims of generalizability, and their absence in the current literature represents one of the most significant evidentiary gaps in the field.

“The regulatory classification of AI surgical diagnostic tools as medical devices subjects their development and commercialization to the full apparatus of medical device regulation, including pre-market notification (510(k)), pre-market approval (PMA), or De Novo pathways in the United States, and CE marking under the EU Medical Device Regulation (EU MDR 2017/745) in Europe. The application of these frameworks to adaptive AI systems — which may update their parameters through continued learning after deployment — is inadequately addressed by existing regulatory architecture, creating compliance uncertainty that may impede innovation” [9].

The economic barriers to surgical AI development are substantial and disproportionately distributed. Development costs encompassing data curation, annotation, computational infrastructure, clinical validation, and regulatory compliance may exceed USD 10–50 million for a single AI diagnostic tool of clinical-grade quality. Reimbursement frameworks for AI-assisted surgical diagnostics are poorly developed in most healthcare systems, creating uncertainty about return on investment that chills private-sector investment and disadvantages non-commercial research programs.

The deployment of “black box” AI diagnostic systems in surgical contexts — wherein the computational basis of a diagnostic recommendation is opaque to the clinician — raises legitimate concerns about accountability, trust calibration, and the maintenance of meaningful human oversight. Explainable AI (XAI) methodologies — including Grad-CAM heat maps that localize image regions contributing to CNN predictions, SHAP (SHapley Additive exPlanations) values that quantify individual feature contributions, and concept-based explanations that express model reasoning in clinically meaningful terms — represent the current state-of-the-art in AI transparency, but their clinical integration and interpretive validity require further development.

The clinical relevance of explainability is not merely philosophical. A surgeon who understands why an AI system has flagged a lesion as malignant — and can evaluate the spatial and textural features underlying this prediction against their own clinical knowledge — is better positioned to appropriately calibrate trust in the AI recommendation and to identify cases where the AI's reasoning pattern may not apply. This human-in-the-loop dynamic is essential for maintaining the quality of human oversight that patient safety requires.

“The use of AI in surgical diagnostic workflows raises normative questions about patient disclosure that remain unresolved in both bioethics and regulatory jurisprudence. A principled argument for mandatory AI disclosure is grounded in the doctrine of informed consent: patients who are subject to AI-assisted diagnostics are, in effect, participating in a form of technologically mediated clinical decision-making whose properties — including potential error patterns, limitations, and the role of training data characteristics — they have a legitimate interest in understanding” [10].

pISSN:2349-5707 Volume:13,Issue 3, March -2026 eISSN:2349-5715

Counterarguments from clinical pragmatics note that a blanket disclosure requirement for AI-assisted diagnostics would generate excessive information burden without proportionate benefit, and that the appropriate disclosure threshold should be calibrated to the degree to which AI plays a determinative rather than adjunctive role in the diagnostic process. These competing normative frameworks are likely to be addressed through evolving regulatory guidance and professional society consensus statements in coming years.

The training of surgical AI diagnostic models on patient data collected in clinical contexts requires robust ethical governance frameworks addressing patient consent, data anonymization fidelity, purpose limitation, and the secondary use of data for commercial model development. Existing legal frameworks — GDPR in Europe, HIPAA in the United States, and analogous national frameworks in Central and South Asia — provide foundational protections but do not yet comprehensively address the specific data governance challenges posed by large-scale AI training data pipelines. The emergence of synthetic data generation — using generative adversarial networks (GANs) or diffusion models to produce artificial medical images that preserve the statistical properties of real patient data without encoding identifiable information — offers a technically promising approach to this challenge, though the fidelity and clinical representativeness of synthetic surgical data require continued validation.

REFERENCES:

1. Topol E. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine. 2019, - P. 44.
2. Meara J. Leather AJM, Hagander L, et al. Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development. The Lancet. 2015, -P. 569.
3. Maier-Hein L., Eisenmann M., Sarikaya D., et al. Surgical data science — from concepts toward clinical translation. Medical Image Analysis. 2022, - P. 31.
4. Hashimoto D., Rosman G., Rus D., Meireles O. Artificial intelligence in surgery: promises and perils. Annals of Surgery. 2018, - P. 70.
5. Goodfellow I., Bengio Y., Courville A., Deep Learning. MIT Press; 2016, - P. 21.
6. Lambin P., Leijenaar R., Deist TM. the bridge between medical imaging and personalized medicine. Nature Reviews Clinical Oncology. 2017, - P. 749.
7. Zwanenburg A., Vallieres M., Abdalah M. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology. 2020, - P. 328.
8. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nature Medicine. 2022, - P. 31.
9. Mesko B., Gorog M. A short guide for medical professionals in the era of artificial intelligence. npj Digital Medicine. 2020, - P. 126.
10. Salminen P., Tuominen R., Paajanen H. Five-year follow-up of antibiotic therapy for uncomplicated acute appendicitis: the APPAC randomized clinical trial. JAMA. 2018. – P. 1265.