SPINE RESEARCH

OPEN

Artificial intelligence in spine surgery

Cheng Zhang[a,b,c], Shanshan Liu[a,b,c], Jialin Shi[d], Xingyu Zhou[e], Peter Passias[f], Nanfang Xu[a,b,c,*] Weishi Li[a,b,c,*]

Abstract

Artificial intelligence (AI) technology has rapidly advanced in recent years, particularly in fields such as computer vision and natural language processing, where significant breakthroughs have been made. The emergence of large language models has greatly enhanced AI's ability to understand and generate text, accelerating its application across various domains. The AI-generated content has maintained a trend of rapid growth, with ChatGPT (OpenAI, USA) and DeepSeek-V3 (DeepSeek, China) gaining global attention due to their outstanding performance. AI development in spinal surgery is still in its early stages. Although some hospitals have pioneered the deployment of deep learning models in imaging and surgical assistance systems, AI tools that are widely adopted and routinely integrated into the daily practice of most spinal surgeons remain scarce. Developing models and tools with high accuracy, strong interpretability, and trustworthiness remains one of the primary goals for AI development in spinal surgery. This review summarizes the recent advancements in AI within the field of spinal surgery, exploring the current challenges, transformations, and future opportunities of AI in spinal surgery. The aim of this review is to enhance the understanding of AI's role in spinal care among clinicians, clinical researchers, AI scientists, and patients. Our goal is to promote interdisciplinary collaboration and development, thereby fostering a comprehensive understanding of AI's potential in improving spinal care.

Keywords: artificial intelligence, ChatGPT, DeepSeek, large language models, spine

1. Introduction

Artificial intelligence (AI) is broadly acknowledged as a branch of computer science dedicated to the research and development of intelligent machines capable of executing tasks that require human intelligence, such as perception, recognition, and decision-making. Over the past few years, with the assistance of AI, medicine and life sciences have made continuous breakthroughs, driving remarkable progress and ushering basic and clinical medical research into a new era.[1–6] Computer vision has made significant strides in the medical field, with widespread applications in the clinical practices of dermatology,[7–9] ophthalmology,[10–12] pathology,[13–15] and medical imaging.[16] The advancement of natural language processing (NLP) has seen general AI, particularly large language models (LLMs) like ChatGPT, achieve performance levels comparable to humans across various professional and academic benchmarks. Compared with earlier NLP models, LLMs exhibit exceptional intelligence. For instance, GPT-4[17] has successfully passed the US Medical Licensing Examination. GPT-4o[18] and DeepSeek-V3,[19] are capable of simulating human-like natural language communication in spoken form, thus being highly anticipated for medical applications. The latest releases of LLMs, such as ChatGPT-o3 and Deepseek-R1, are capable of handling complex coding and reasoning tasks with remarkable efficiency. Additionally, AI technology, unlike traditional medical technology transformation models, demonstrates a shorter transformation cycle. The US Food and Drug Administration (FDA) is approving AI products at an increasing pace.[20] These advancements are rapidly reshaping the traditional paradigms of medical practice.

Spinal diseases pose a significant global health threat. According to a large-scale study conducted across 195 countries, low back pain, a prevalent spinal condition, is responsible for the highest global loss of productivity as measured in years. In 126 countries, it ranks first among the causes of years lived with disability.[21] Additionally, low back pain imposes a substantial socioeconomic burden on numerous countries, with annual expenditures in the United States alone estimated to exceed one hundred billion dollars.[22] Thus, integrating AI into spinal

CZ, SL, and JL contributed to this article equally.

[a] Department of Orthopaedics, Peking University Third Hospital, Beijing, China, [b] Engineering Research Center of Bone and Joint Precision Medicine, Ministry of Education, Beijing, China, [c] Beijing Key Laboratory of Spinal Disease Research, Beijing, China, [d] The University of Science and Technology Beijing, Beijing, China, [e] Peking University Health Science Center, Beijing, China, [f] NYU Langone Orthopedic Hospital, New York City, New York.

* Correspondence: Weishi Li, Department of Orthopaedics, Peking University Third Hospital, 49 North Garden Road, Beijing 100191, China (e-mail: puh3liweishi@163.com), and Nanfang Xu, Department of Orthopaedics, Peking University Third Hospital, 49 North Garden Road, Beijing 100191, China (e-mail: xunanfang@foxmail.com)

Spine Research (2025) 11:1(e0005)

Received: 08 April 2025 / Received in final form: 30 April 2025 / Accepted: 30 April 2025

http://dx.doi.org/10.1097/br9.0000000000000005

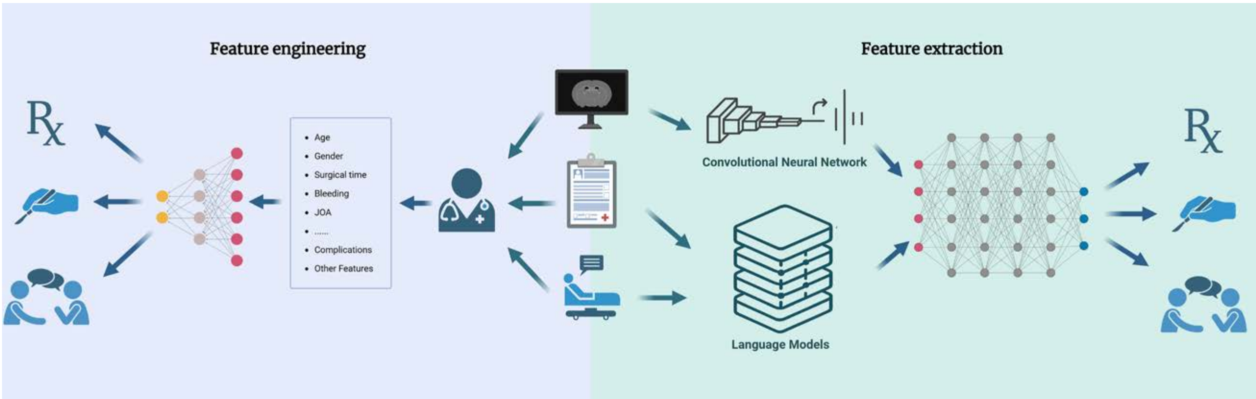

Figure 1. Schematic representation of traditional feature engineering versus deep learning-based feature extraction for processing multi-modal information. Traditional feature engineering requires physicians to manually select features for model incorporation based on prior knowledge, whereas feature extraction employs deep learning networks such as convolutional neural networks or language models to automatically identify features deemed important for the outcome by the algorithm. Created in BioRender. Zhang, C. (2025) BioRender.com/pxmc1co.

surgery has the potential to significantly reduce human labor costs, lower healthcare expenses associated with spinal diseases, and enhance medical quality and disease prognosis, ultimately improving the quality of life for the global population.

In recent years, the establishment of AI clinical research paradigms such as Standards for Reporting of Diagnostic Accuracy Study - Artificial Intelligence[23,24] has led to the standardization of AI-related application research in clinical settings. Research methodologies have also advanced significantly, evolving from single-center, single-modality, retrospective studies to multi-center, multi-modality, and prospective studies,[25] as well as large-sample validations in real-world settings and randomized controlled trials.[25,26] Despite these advancements, high-quality studies in spinal surgery remain relatively scarce. In this review, we summarize the progress of AI research in spinal surgery over the recent years. We highlight studies characterized by rigorous design, high levels of evidence, significant innovation, and real-world applications. Additionally, this review delves into the current challenges, transformations, and future opportunities for AI in spinal surgery. The goal of this review is to enhance the understanding of AI's role in spinal healthcare among clinicians, clinical researchers, AI scientists, and patients. We aim to promote interdisciplinary collaboration and development, thereby fostering a comprehensive appreciation of AI's potential in improving spinal care.

2. Recent progress of AI techniques in spine surgery

Developing models to address surgical challenges relies heavily on feature selection, a process commonly referred to as feature engineering. Before the widespread application of computer vision AI technologies, feature selection predominantly relied on manual design. This approach involved leveraging prior knowledge to identify relevant features and subsequently utilizing classical mathematical methods or machine learning for model construction (Fig. 1). These models are often characterized by their strong interpretability and are still widely used to this day.[27–41] In the past few years, the widespread adoption of deep learning has driven a paradigm shift in feature extraction and model construction, transitioning from expert-guided methodologies to data-driven approaches. Depending on the data modality, technologies such as NLP and computer vision techniques have been extensively applied. Furthermore, advancements in automatic speech recognition (ASR) technology have effectively transformed speech-related challenges into problems that can be addressed using NLP. These technologies can operate independently or be combined to address more complex tasks (Fig. 1). This section will provide an overview of advancements in AI in spinal surgery from the perspective of AI techniques.

2.1. NLP and LLMs in Spine Surgery

Over the past 2 years, the most notable advancement in the field of AI has been the advent of LLMs, such as ChatGPT, DeepSeek, Gemini, etc. LLMs refer to pretrained language models (PLMs) characterized by a vast number of parameters and extensive training on massive datasets. In recent years, they have become a central focus of research and innovation in AI.[42,43] What distinguishes LLMs from smaller-scale PLMs is their remarkable emergent abilities to tackle complex tasks. Research has shown that LLMs, such as GPT-3 with around 175 billion parameters, demonstrate a significant advancement in NLP capabilities compared with smaller PLMs like GPT-2, which has approximately 1.5 billion parameters.[43,44] In medical scenarios, any task requiring natural language understanding and generation is a potential application area for LLMs.[45–48] These potential scenarios encompass preconsultation advice, disease diagnosis and classification, interpretation and structuring of imaging reports, medical record documentation, summarization and generation of medical records, simplifying medical terminology for patient education, treatment recommendations, communication of surgical risks, generation of informed consent documents, intraoperative support, and postoperative complication prevention advice.

Numerous pioneers have explored the application of LLMs in various contexts related to spine surgery. One of the most direct implementations of LLMs is in question-and-answer formats. For common spine-related conditions or symptoms—such as low back pain,[49] lumbosacral radicular pain,[50] spinal deformities,[51] and spinal cord injuries[52]—LLMs can quickly generate easily readable responses, facilitating comprehension for individuals without formal medical training or those with lower levels of basic education. Although current LLMs do not yet achieve an accuracy of 100%, this application highlights their potential to function as accessible, on-demand personal medical consultants. Additionally, studies suggest that LLMs can reduce the reading difficulty of online educational materials on spine surgery from a high school reading level to approximately a sixth-grade level without introducing factual errors or inaccuracies. This reduction in complexity is critical for fostering a more equitable and transparent environment for science education.[53]

For spine surgeons, LLMs can serve as versatile copilots, aiding in routine tasks. For example, LLMs are capable of interpreting imaging reports and classifying the severity of scoliosis based on textual descriptions.[54] In 1 study, 4 commonly used models were tested on 56 cases, with ChatGPT 4 and Scholar AI Premium achieving 100% sensitivity and specificity in classification. Furthermore, LLMs can generate text for patient interactions, such as low back pain evaluation forms[55] and informed consent documents, which include notes on alternative treatments and precautions.[56] However, it remains necessary for physicians to review and validate the generated content to ensure accuracy and appropriateness.

Despite these advancements, the current performance of LLMs in critical medical tasks raises concerns. For instance, when ChatGPT 4 was used to diagnose cases from the "Case of the Month" series in the *American Journal of Neuroradiology*, the accuracy in spine-related subgroups was only 55%. This underscores the fact that LLMs cannot yet be fully relied upon in clinical settings. Unlike other medical devices, LLMs have not undergone extensive clinical trials to validate their utility before public release. Thus, comprehensive research is needed to assess LLM performance across a full spectrum of spine surgery diseases and scenarios. Nevertheless, LLMs, such as ChatGPT, are undeniably beginning to redefine the workflow of spine surgeons.

Additionally, the extensive use of LLMs does not signify the obsolescence of traditional NLP algorithms. In fact, NLP models trained on specific datasets can still achieve high accuracy in specialized tasks. Furthermore, NLP models with fewer parameters offer advantages in medical practice that LLMs lack, such as reduced storage requirements, lower demands of computational resources, and the capability to be deployed offline on appropriate devices, thus effectively safeguarding patient privacy.

2.2. Spinal Imaging

The application of AI in spinal imaging remains a rapidly advancing field. This growth is driven, on one hand, by the crucial role of medical imaging in the diagnosis and assessment of spinal disorders, providing a broad range of AI application scenarios. On the other hand, there is still a relative scarcity of AI-powered spinal imaging analysis products integrated into clinical workflows, indicating that substantial market demand persists. Furthermore, spinal surgery is heavily reliant on imaging studies, suggesting that AI-driven technological advancements are poised to lead the next major innovation in spinal surgery.

AI applications in imaging diagnosis represent one of the earliest areas of integration into spinal surgery. The fundamental approach involves utilizing deep learning models to extract imaging features, followed by applying high-performance classifiers to categorize these features, ultimately completing the diagnostic process. For conditions with more distinctive imaging characteristics, deep learning models have achieved high diagnostic accuracy. Examples include scoliosis,[57–59] spinal stenosis,[60–63] and osteoporosis.[64–69] Other conditions, such as spinal compression fractures[70] and spinal cord injuries,[71] have also been successfully classified using AI techniques. Additionally, predictive modeling for the prognosis of spinal diseases has emerged as another significant application of classification algorithms.[36,72,73]

At the same time, advances in algorithmic techniques have facilitated more precise detection and segmentation outcomes. Vertebral segmentation and detection are among the most commonly studied topics, with a primary application being the measurement of spinal parameters based on segmentation and detection results. However, the irregular structure and morphological similarity of vertebrae pose challenges, for instance, segmentation and stage identification, making algorithm development in this area particularly complex. Zhang et al.[74] developed a robust vertebral landmark detection system for spinal X-rays, which enables rapid and accurate measurement of the Cobb angle. Similar efforts have achieved high accuracy in vertebral segmentation using CT and MRI data.[75,76] Disc segmentation, critical in diagnosing and evaluating degenerative disc disease, has also seen progress. Semantic segmentation algorithms for MRI disc images have enabled high-accuracy quantification of degenerative disc signals.[77] Additionally, intraoperative radiation-free automatic registration based on segmentation algorithms has provided critical support for surgical navigation in spinal procedures.[78] The continuous development of segmentation technologies also offers valuable insights for downstream applications.[79–82]

Generative AI in imaging remains both a significant challenge and an area of substantial potential for innovation. One of the most prominent applications of generative AI is the synthesis of new imaging data, which contributes to the creation of larger and more diverse datasets. This, in turn, enhances the generalizability of algorithms while improving patient privacy. The application of SpineGAN for image synthesis in transfer learning has demonstrated high recall rates.[83] Furthermore, generative AI techniques have been employed to reconstruct 3D CT images from 2D X-rays, thereby enhancing the diagnostic capabilities of X-rays while simultaneously reducing the radiation exposure associated with CT scans.[83,84] These methods hold promise for generating X-rays or CT images without radiation, further minimizing radiation exposure and reducing the risk of organ damage and tumorigenesis—an especially important consideration for pediatric populations. For instance, children with adolescent idiopathic scoliosis require frequent radiographic evaluations during diagnosis and follow-up. Through the use of optical devices, in combination with light-based depth sensing and deep learning techniques, adolescent idiopathic scoliosis can be analyzed, and synthetic images can be generated,[85] significantly reducing radiation exposure for pediatric patients. Furthermore, by leveraging deep learning techniques, vertebral morphology detection and automatic measurement of spinal parameters can be performed through radiation-free ultrasound examinations.[86,87]

Developing low- or no-radiation screening and diagnostic techniques is expected to become a major focus of future research and clinical application. Collaboration between spinal surgeons and AI researchers will be essential for the development of higher-value technologies that can be seamlessly integrated into clinical practice.

3. Case studies and real-world applications in spine surgery

From the perspective of spinal orthopedic disease management, the entire treatment process can be divided into 3 key stages: preoperative, intraoperative, and postoperative, with the preoperative phase further encompassing the preadmission period and the postoperative phase extending into the postdischarge period. AI can offer distinct functionalities across these stages of care. In this section, we will examine the potential impact of AI on real-world workflows, integrating clinical case examples to demonstrate its practical applications and influence in clinical settings.

The preoperative application scenario is depicted in Figure 2. When patients present with symptoms such as back pain, online LLMs applications can serve as a valuable tool for consultation, providing relevant advice and guiding patients on whether they should seek hospital care.[88–90] Once admitted, AI can assist clinicians in preoperative diagnosis, disease classification, and the quantitative assessment of disease progression and severity.[1,2,6] It can also assess spinal parameters, predict surgical efficacy, complications, and clinical outcomes based on preoperative characteristics, and identify patients who meet surgical indications.[40] Furthermore, AI has the potential to support physicians in drafting medical records and informed consent forms, reducing the complexity of medical texts and enhancing communication between doctors and patients.[48]

The role of AI models in surgery primarily focuses on decision support, imaging and navigation, and robotics. Additionally, AI can significantly enhance the efficiency

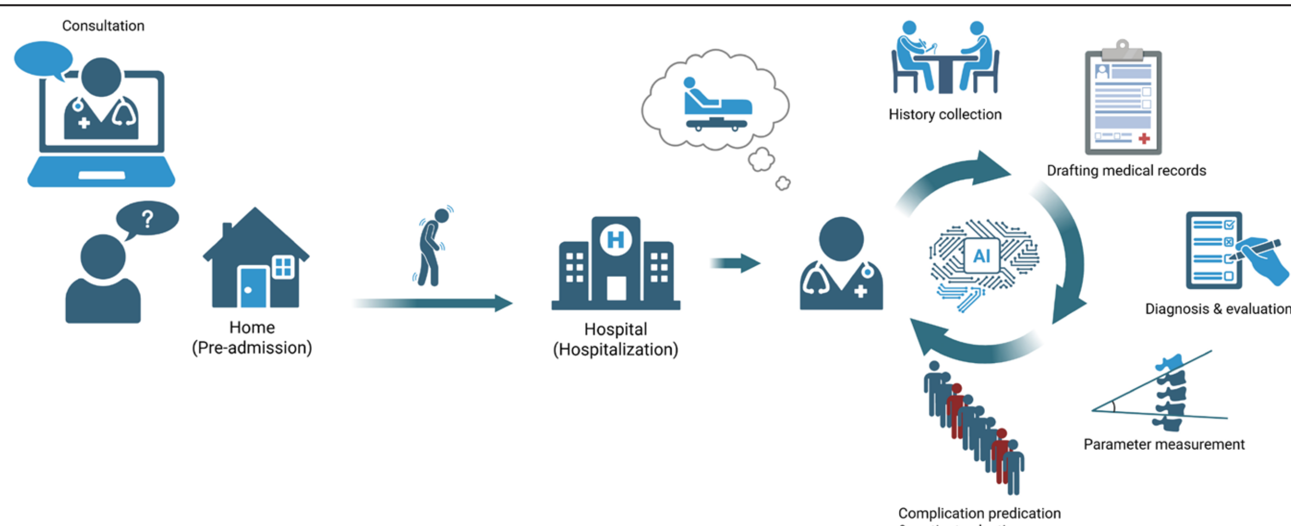

Figure 2. Preoperative application scenarios, including preadmission home-based remote online consultations and screening, as well as postadmission automated medical history collection, clinical records drafting, diagnosis, evaluation, imaging measurements, surgical candidate selection based on indications, and complication prediction. Created in BioRender. Zhang, C. (2025)

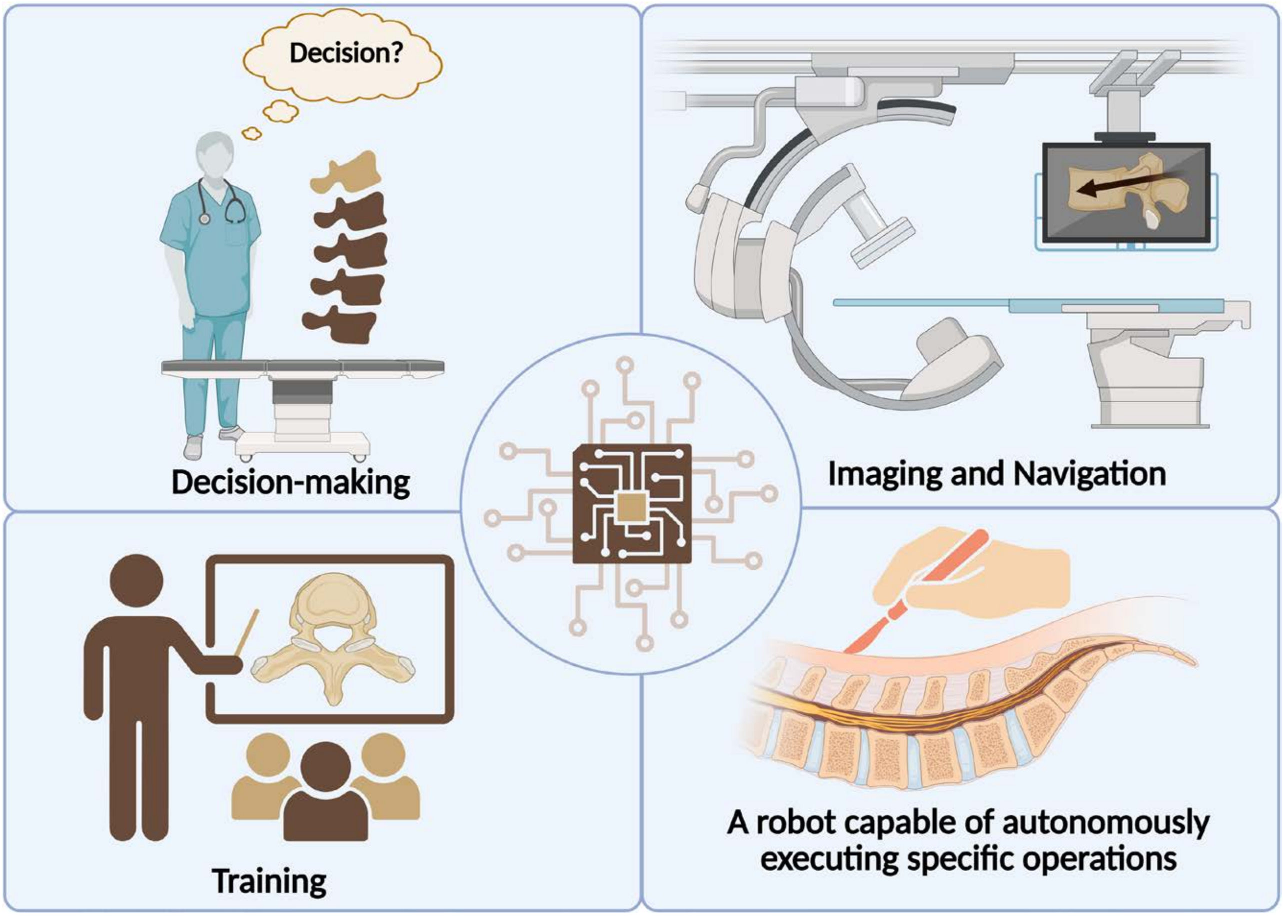

Figure 3. Intraoperative application scenarios, including decision-making, navigation and auto-planning, surgical education and training, and autonomously executing specific operations. Created in BioRender. Zhang, C. (2025)

and effectiveness of surgical education, improving the quality of surgical training for physicians (Fig. 3). Furthermore, AI models for risk prediction can also be utilized to forecast postoperative prognosis, clinical outcomes, and surgical complications.[33,35,91–96] Furthermore, for spinal patients undergoing rehabilitation, the integration of real-time motion analysis can support functional recovery and strength rehabilitation, potentially enhancing rehabilitation efficiency and facilitating.

3.1. Case 1

For instance, in the case of degenerative lumbar spondylolisthesis, diagnosis is often subjective. In certain cases, such as subtle vertebral slippage, whether to diagnose spondylolisthesis primarily depends on the physician's experience, leading to significant inconsistency between different physicians. The deep learning system built on the multi-task, proposal correlation, feature fusion-net (MPF-net), through vertebral detection segmentation and case-based learning, can autonomously diagnose lumbar spondylolisthesis and quantify the degree of slippage. Moreover, the system can accurately measure the sagittal parameters of the spinal sequence via its frontend interface and visualize the results, providing valuable insights for the formulation of precise corrective treatment plans (Fig. 4). Owing to its excellent scalability, the model is capable of handling multiple tasks across various preoperative scenarios.

3.2. Case 2

Laminectomy, a complex procedure involving the removal of the lamina to alleviate spinal cord compression, is a critical step in spinal surgery. This operation typically requires a highly experienced spinal surgeon. Due to the lamina's proximity to the spinal cord, the surgeon must exercise caution when using a Kerrison rongeur, a process that is time-consuming and relies on the surgeon's experience and tactile feedback to control the depth and force of lamina resection, which carries an inherent risk of spinal cord injury. Surgical robots, developed using control theory, AI-based image processing and analysis, and high-precision signal pattern recognition technologies, can personalize the surgical approach for laminectomy, automatically identify biological tissue features during the resection, and perform precise, autonomous lamina cutting with submillimeter accuracy. This innovation significantly reduces the risks associated with laminectomy and advances orthopedic robotic surgery to a higher level of automated decision-making and execution[97–99] (Fig. 5).

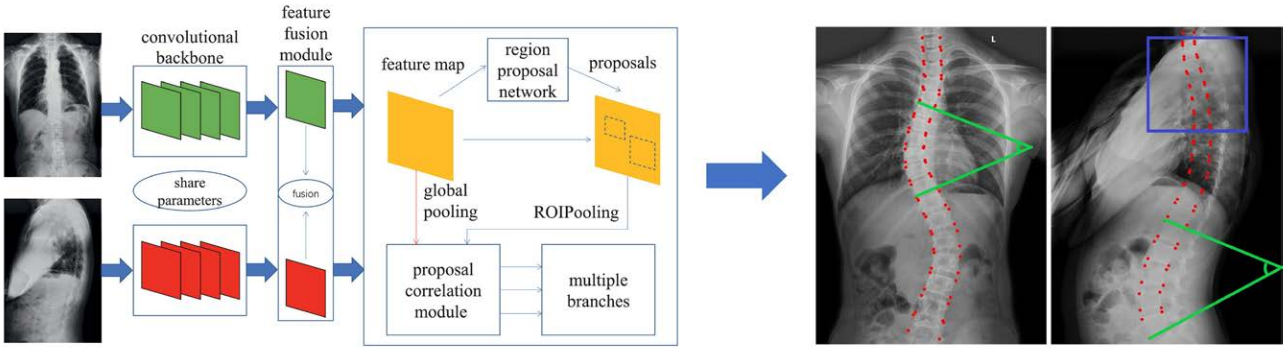

Figure 4. Schematic representation of MPF-net workflow (Reprinted with permission from Zhang et al[74]). MPF-net = multi-task, proposal correlation, feature fusion-net.

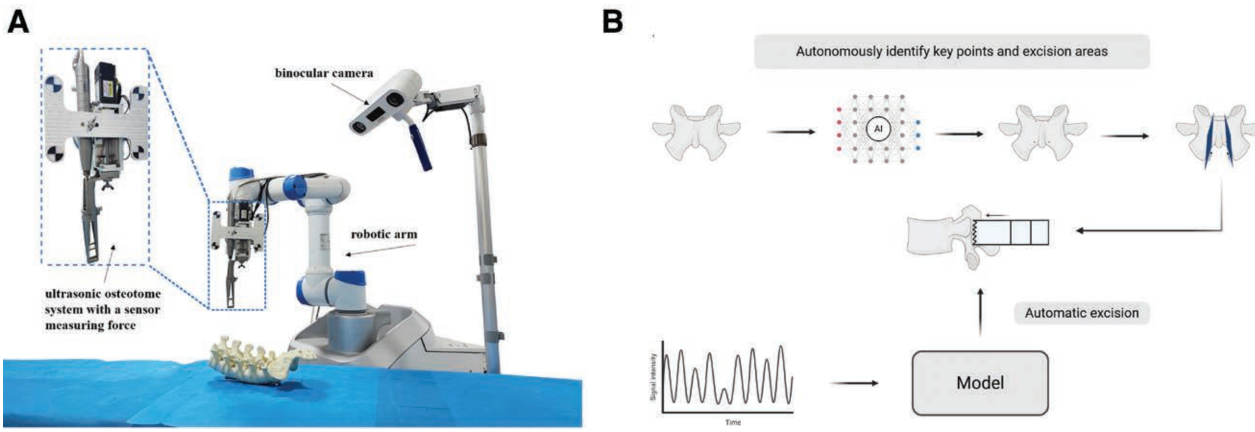

Figure 5. Surgical robot for laminectomy procedures. (A) An overview of the composition of surgical robots. (B) The concise workflow for autonomous identification and laminectomy. (A was reprinted with permission from Li X et al[99]; B was created in BioRender. Zhang, C. (2025)

4. Challenges in the current and future

Despite the significant advancements in AI research within the field of spine surgery and its promising potential for future technological breakthroughs, there remain substantial limitations and challenges in its clinical application. The complexity of clinical practice and the stringent requirement for low error tolerance imply that, for the foreseeable future, AI will primarily function as an adjunct or "copilot" to assist in clinical work rather than fully replace spine surgeons. This section will discuss the inherent limitations of AI and the specific challenges it faces in spine surgery.

Deep learning models often exhibit a "black box" nature, wherein they provide results without a clear explanation of their internal mechanisms. This opacity significantly hinders their application in the medical domain. Clinicians, when utilizing AI-generated medical decisions, typically need to understand the underlying rationale behind these decisions. The absence of explainability in AI models compromises clinicians' trust, which subsequently restricts their adoption in clinical workflows. Furthermore, these AI applications raise ethical concerns, particularly regarding the attribution of responsibility. When AI systems influence clinical decision-making, assigning responsibility becomes more complex in the event of adverse outcomes, including medical errors or accidents. The issue of responsibility allocation remains a subject of ongoing debate within both academic circles and clinical practice. Moreover, if the widespread adoption of AI alters the role of clinicians in medical decision-making, it may erode the trust between healthcare providers and patients, thereby introducing additional ethical dilemmas. At present, regardless of the technological factors contributing to medical malpractice, doctors remain largely responsible. As such, doctors must exercise caution when using AI models and strengthen the review of AI-generated content. Additionally, it is crucial to refine and establish regulations regarding the access, application, and accountability of AI technologies to ensure the safe and effective deployment of AI software in healthcare environments.

The performance of AI models is heavily dependent on the quality of the training data, with data accuracy and completeness being crucial factors. When models are trained on biased or incomplete datasets, their predictions may carry considerable uncertainty, posing serious risks when applied to clinical practice. For example, *Nature* retracted a paper concerning the application of machine learning in cancer diagnosis because the data used in the study did not meet quality standards.[100] Hence, rigorous data quality control is essential during both the training and testing phases of AI models. Additionally, medical data inherently contains sensitive personal health and biometric information. Balancing the protection of patient privacy with the utilization of medical data remains a critical challenge for AI applications in healthcare. In collaboration with institutions that manage such data, federated learning and data anonymization techniques have been employed to mitigate the risk of data breaches. Federated learning allows AI models to be trained without the need to share raw data, while data anonymization enhances data security by removing personally identifiable information.

The specialized nature of spine surgery introduces unique challenges for the application of AI. Many classification systems and scoring criteria for spinal diseases are inherently subjective, leading to inconsistent labeling across different medical centers and even among physicians within the same center. These inconsistencies pose significant obstacles to AI model development, as models trained on single-center data often exhibit limited generalizability and poor external validation. This hampers the widespread adoption of AI models in clinical practice. Currently, models trained and tested using retrospective single-center datasets are unlikely to be widely recognized. Training with data from multiple centers and evaluating models in real-world environments may become essential prerequisites for model release.

The high degree of specialization required in spine surgery also complicates the process of generating high-quality labeled data, as expert annotation is often necessary. This significantly increases the difficulty, cost, and time required for data preparation, making the development of high-quality AI models for specific spinal conditions particularly challenging. In cases where expert annotations are insufficient, AI models struggle to achieve optimal performance. This issue is particularly pronounced in the development of models for rare spinal diseases, such as craniovertebral junction anomalies. On the one hand, the limited number of cases available makes it difficult to amass large datasets. On the other hand, the lack of standardized treatment protocols exacerbates data heterogeneity, even when aggregating multicenter datasets, which vary in quality. This heterogeneity further constrains the development and implementation of AI models in spine surgery. Despite the immense capabilities of GPT-4, which integrates knowledge from nearly all public medical platforms, it remains difficult to resolve the diagnostic and therapeutic challenges posed by rare diseases (Fig. 6). To address this issue, it is crucial to establish a large-scale, multicenter network focused on standardized treatment of rare diseases. This network should emphasize the alignment of data across centers and promote collaborative training among physicians to ensure the creation of high-quality datasets. Such efforts are essential for enabling AI technologies to effectively support the diagnosis and management of rare diseases.

5. Future development trends and technological outlook

The emergence of ChatGPT has generated unprecedented interest among researchers in general-purpose AI models. In contrast to specialized models, which are limited to tasks within a specific domain, the hallmark of

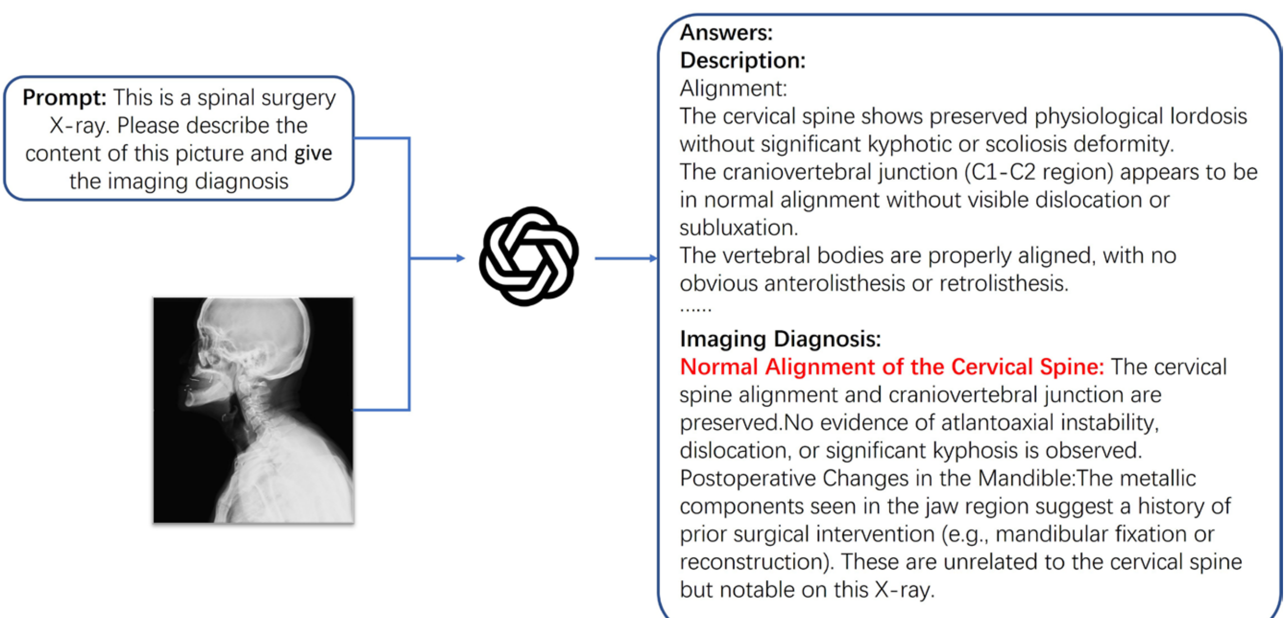

Figure 6. The GPT's incorrect response regarding rare diseases. This medical imaging depicts a patient with os odontoideum accompanied by posterior atlantoaxial dislocation with instability. When this case was submitted through the official GPT-4 interface, the model erroneously interpreted the cervical spine alignment as normal.

general-purpose models lies in their versatility, enabling them to perform a broad range of tasks across various fields. In the medical domain, for instance, a general-purpose AI model can leverage the extensive existing knowledge to address questions across different specialties. Furthermore, some research has incorporated image processing algorithms to embed medical images, allowing for multimodal input. This integration facilitates reasoning that combines both medical texts and medical images, bringing the model closer to the cognitive processes of real physicians.

Despite their potential, current medical models remain in the early stages of development. Chen et al.[101] constructed a benchmark to evaluate widely used medical general-purpose models through multiple-choice questions. In tests related to orthopedics, the highest accuracy achieved was 52.09%, which is significantly lower than GPT-4's performance of 62.80% in the same test. This highlights the fact that the accuracy of general-purpose medical models is still far from meeting the clinical standards required for practical application. Consequently, further advancements are needed to improve model accuracy. Additionally, as these large models are trained on vast amounts of publicly available data, it is essential to avoid using benchmarks and public datasets derived from the same sources during model testing to prevent errors due to data leakage. Thus, the creation of high-quality, nonpublic, closed-source datasets and benchmarks will be a crucial area of focus for future model development.

Beyond medical consultation and diagnosis, surgery constitutes the core of surgical practice. Consequently, the application of AI to optimize surgical workflows represents a critical challenge that surgical researchers around the world must collaboratively address.

Currently, surgical robotic technology is continuously evolving and undergoing iterative advancements. Although current spinal surgical robots are capable of autonomously planning the resection area of the lamina and automatically identifying the lamina to be removed, other aspects of the surgery still rely on manual operation. As such, there remains considerable potential for further development in the field of surgical robotics.

Technologies associated with surgery, including augmented reality, mixed reality, navigation systems, and robotics, have already demonstrated their effectiveness and reliability.[102–104] Moving forward, these technologies need to undergo further iteration and integration to create more intelligent, digitalized operating rooms. The goal is to improve surgical efficiency and precision while enhancing safety, minimizing invasiveness, and reducing radiation exposure for both patients and surgeons. Ultimately, this approach is expected to yield significant benefits for both surgical teams and patients.

6. Summary

In summary, the era of AI in medicine is already upon us, and AI applications in spinal surgery have attracted considerable attention. However, despite the wealth of published research, the number of AI applications actually implemented in clinical practice remains minimal. The adoption of AI software faces significant challenges in terms of trust and regulation. It is anticipated that, as algorithms progress, models become more transparent, and regulations evolve, these obstacles will be addressed. The advent of multimodal, general-purpose large models has greatly expanded AI's functionality and application potential, bringing it closer to the cognitive processes of practicing clinicians. As a fundamental component of surgical practice, the intelligentization of surgery, particularly through robotics and smart operating rooms, has long been a focal point in AI research. As AI gains increasing acceptance among both doctors and patients, researchers will continue advancing towards more refined and intelligent surgical solutions, ultimately enhancing the quality of spinal surgical practices.

Acknowledgments

The authors gratefully acknowledge the assistance of Deepseek and ChatGPT in the translation and linguistic refinement of this manuscript. The AI tool was used exclusively for language-related purposes.

Ethical statement

Not applicable.

Conflicts of interest

Weishi Li is the Editor-in-Chief of Spine Research. He was excluded from the peer-review process and all editorial decisions regarding the acceptance and publication of this article. Peer review was independently conducted by the other editors to minimize bias.

Funding source

This work was supported by the National Natural Science Foundation of China (No. 82272577). The founders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript..

Data availability statement

Not applicable.

Author contributions

All authors contributed significantly to the research.
Conceptualization and study design: Weishi Li and Nanfang Xu.
Initial manuscript drafting: Cheng Zhang, Shanshan Liu, and Jialin Shi.
Engaged in discussions and revised the manuscript: Shanshan Liu, Xingyu Zhou, Peter Passias.

References

[1] Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28:31–8.

[2] Huang T, Xu H, Wang H, et al. Artificial intelligence for medicine: progress, challenges, and perspectives. Innov Med. 2023;1:100030.
[3] Topol EJ. Learning the language of life with AI. Science. 2025;387:eadv4414.
[4] Machine learning helps to determine the diverse conformations of RNA molecules [published online ahead of print December 18, 2024]. Nature. 2024. doi: 10.1038/d41586-024-04134-2.
[5] Liu Y, Chen Y, Han L. Bioinformatics: advancing biomedical discovery and innovation in the era of big data and artificial intelligence. Innov Med. 2023;1:100012.
[6] Varghese C, Harrison EM, O'Grady G, Topol EJ. Artificial intelligence in surgery. Nat Med. 2024;30:1257–68.
[7] Webster P. How AI-powered handheld devices are boosting disease diagnostics - from cancer to dermatology. Nat Med. 2024;30:914–5.
[8] Daneshjou R, Vodrahalli K, Novoa RA, et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. Sci Adv. 2022;8:eabq6147.
[9] Kose K, Rotemberg V. The promise and drawbacks of federated learning for dermatology AI. JAMA Dermatol. 2024;160:269–70.
[10] Cen LP, Ji J, Lin JW, et al. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. Nat Commun. 2021;12:4828.
[11] Varadarajan AV, Bavishi P, Ruamviboonsuk P, et al. Predicting optical coherence tomography-derived diabetic macular edema grades from fundus photographs using deep learning. Nat Commun. 2020;11:130.
[12] Teo ZL, Ting DSW. AI telemedicine screening in ophthalmology: health economic considerations. Lancet Glob Health. 2023;11:e318–20.
[13] Lu MY, Chen TY, Williamson DFK, et al. AI-based pathology predicts origins for cancers of unknown primary. Nature. 2021;594:106–10.
[14] Abdurrachim D, Lek S, Ong CZL, et al. Utility of AI digital pathology as an aid for pathologists scoring fibrosis in MASH. J Hepatol. 2025;82:898–908.
[15] Pulaski H, Harrison SA, Mehta SS, et al. Clinical validation of an AI-based pathology tool for scoring of metabolic dysfunction-associated steatohepatitis. Nat Med. 2025;31:315–22.
[16] Rajpurkar P, Lungren MP. The current and future state of AI interpretation of medical images. N Engl J Med. 2023;388:1981–90.
[17] Achiam OJ, Adler S, Agarwal S, et al. GPT-4 Technical Report; 2023.
[18] Hurst OA, Lerer A, Goucher AP, et al. GPT-4o system card. ArXiv. 2024;abs/2410.21276.
[19] Liu A, Feng B, et al.; DeepSeek-AI. DeepSeek-V3 technical report ArXiv. 2024;abs/2412.19437.
[20] Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med. 2020;3:118.
[21] GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392:1789–858.
[22] Knezevic NN, Candido KD, Vlaeyen JWS, Van Zundert J, Cohen SP. Low back pain. Lancet. 2021;398:78–92.
[23] Sounderajah V, Ashrafian H, Aggarwal R, et al. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: the STARD-AI Steering Group. Nat Med. 2020;26:807–8.
[24] Sounderajah V, Ashrafian H, Golub RM, et al; STARD-AI Steering Committee. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open. 2021;11:e047709.
[25] Li J, Guan Z, Wang J, et al. Integrated image-based deep learning and language models for primary diabetes care. Nat Med. 2024;30:2886–96.
[26] Wan P, Huang Z, Tang W, et al. Outpatient reception via collaboration between nurses and a large language model: a randomized controlled trial. Nat Med. 2024;30:2878–85.
[27] Shi X, Cui Y, Wang S, Pan Y, Wang B, Lei M. Development and validation of a web-based artificial intelligence prediction model to assess massive intraoperative blood loss for metastatic spinal disease using machine learning techniques. Spine J. 2024;24:146–60.
[28] Phellan Aro R, Hachem B, Clin J, Mac-Thiong J-M, Duong L. Real-time prediction of postoperative spinal shape with machine learning models trained on finite element biomechanical simulations. Int J Comput Assist Radiol Surg. 2024;19:1983–90.
[29] Jiang F, Li X, Liu L, Xie Z, Wu X, Wang Y. Automated machine learning-based model for the prediction of pedicle screw loosening after degenerative lumbar fusion surgery. Biosci Trends. 2024;18:83–93.
[30] Wu X, Wang Z, Zheng L, et al. Construction and verification of a machine learning-based prediction model of deep vein thrombosis formation after spinal surgery. Int J Med Inform. 2024;192:105609.
[31] Karabacak M, Schupper AJ, Carr MT, Bhimani AD, Steinberger J, Margetis K. Development and internal validation of machine learning models for personalized survival predictions in spinal cord glioma patients. Spine J. 2024;24:1065–76.
[32] Fan G, Yang S, Qin J, et al. Machine learning predict survivals of spinal and Pelvic Ewing's Sarcoma with the SEER database. Global Spine J. 2024;14:1125–36.
[33] Cui YP, Shi XD, Qin Y, et al. Establishment and validation of an interactive artificial intelligence platform to predict postoperative ambulatory status for patients with metastatic spinal disease: a multicenter analysis. Int J Surg. 2024;110:2738–56.
[34] Fan GX, Liu HQ, Yang S, et al. Early prognostication of critical patients with spinal cord injury a machine learning study with 1485 case. Spine. 2024;49:754–62.
[35] Guo ZJ, Wang PY, Ye SH, et al. Interpretable machine learning models based on shapley additive explanations for predicting the risk of cerebrospinal fluid leakage in lumbar fusion surgery. Spine. 2024;49:1281–93.
[36] Al-Shawwa A, Ost K, Anderson D, et al. Advanced MRI metrics improve the prediction of baseline disease severity for individuals with degenerative cervical myelopathy. Spine J. 2024;24:1605–14.
[37] Mohanty S, Hassan FM, Lenke LG, et al; International Spine Study Group. Machine learning clustering of adult spinal deformity patients identifies four prognostic phenotypes: a multicenter prospective cohort analysis with single surgeon external validation. Spine J. 2024;24:1095–108.
[38] Schönnagel L, Caffard T, Vu-Han TL, et al. Predicting postoperative outcomes in lumbar spinal fusion: development of a machine learning model. Spine J. 2024;24:239–49.
[39] Schönnagel L, Tani S, Vu-Han TL, et al. Predicting conversion of ambulatory ACDF patients to inpatient: a machine learning approach. Spine J. 2024;24:563–71.
[40] Berg B, Gorosito MA, Fjeld O, et al. Machine learning models for predicting disability and pain following lumbar disc herniation surgery. JAMA Netw Open. 2024;7:e2355024.
[41] Pedro KM, Alvi MA, Hejrati N, Quddusi AI, Singh A, Fehlings MG. Machine learning-based cluster analysis identifies four unique phenotypes of patients with degenerative cervical myelopathy with distinct clinical profiles and long-term functional and neurological outcomes. Ebiomedicine. 2024;106:105226.
[42] Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023;29:1930–40.
[43] Zhao WX, Zhou K, Li J, et al. A survey of large language models. ArXiv. 2023;abs/2303.18223.
[44] Wei J, Tay Y, Bommasani R, et al. Emergent abilities of large language models. ArXiv. 2022;abs/2206.07682.
[45] Roberts K. Large language models for reducing clinicians' documentation burden. Nat Med. 2024;30:942–3.
[46] Omiye JA, Gui H, Rezaei SJ, Zou J, Daneshjou R. Large language models in medicine: the potentials and pitfalls: a narrative review. Ann Intern Med. 2024;177:210–20.
[47] Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature. 2023;620:172–80.
[48] Zhang C, Liu S, Zhou X, et al. Examining the role of large language models in orthopedics: systematic review. J Med Internet Res. 2024;26:e59607.
[49] Shrestha N, Shen Z, Zaidat B, et al. Performance of ChatGPT on NASS clinical guidelines for the diagnosis and treatment of low back pain: a comparison study. Spine (Phila Pa 1976). 2024;49:640–51.
[50] Gianola S, Bargeri S, Castellini G, et al. Performance of ChatGPT compared to clinical practice guidelines in making informed decisions for lumbosacral radicular pain: a cross-sectional study. J Orthop Sports Phys Ther. 2024;54:222–8.
[51] Kasthuri VS, Glueck J, Pham H, et al. Assessing the accuracy and reliability of AI-generated responses to patient questions regarding spine surgery. J Bone Joint Surg Am. 2024;106:1136–42.
[52] Temel MH, Erden Y, Bagcier F. Information quality and readability: ChatGPT's responses to the most common questions about spinal cord injury. World Neurosurg. 2024;181:e1138–44.
[53] Kirchner GJ, Kim RY, Weddle JB, Bible JE. Can artificial intelligence improve the readability of patient education materials? Clin Orthop Relat Res. 2023;481:2260–7.
[54] Fabijan A, Polis B, Fabijan R, Zakrzewski K, Nowosławska E, Zawadzka-Fabijan A. Artificial intelligence in scoliosis classification: an investigation of language-based models. J. Pers. Med. 2023;13:1695.
[55] Coraci D, Maccarone MC, Regazzo G, Accordi G, Papathanasiou JV, Masiero S. ChatGPT in the development of medical questionnaires. The example of the low back pain. Eur J Transl Myol. 2023;33:12114.

[56] Stroop A, Stroop T, Alsofy SZ, et al. Large language models: are artificial intelligence-based chatbots a reliable source of patient information for spinal surgery? Eur Spine J. 2024;33:4135–43.
[57] Kokabu T, Kanai S, Kawakami N, et al. An algorithm for using deep learning convolutional neural networks with three dimensional depth sensor imaging in scoliosis detection. Spine J. 2021;21:980–7.
[58] Mandel W, Oulbacha R, Roy-Beaudry M, Parent S, Kadoury S. Image-guided tethering spine surgery with outcome prediction using spatio-temporal dynamic networks. IEEE Trans Med Imaging. 2021;40:491–502.
[59] Zhang L, Pei BQ, Zhang SJ, et al. A new method for scoliosis screening incorporating deep learning with back images. Glob. Spine J. 2024;15:2062–74.
[60] Won D, Lee HJ, Lee SJ, Park SH. Spinal stenosis grading in magnetic resonance imaging using deep convolutional neural networks. Spine (Phila Pa 1976). 2020;45:804–12.
[61] Li KY, Weng JJ, Li HL, Ye H-B, Xiang J-W, Tian N-F. Development of a deep-learning model for diagnosing lumbar spinal stenosis based on CT images. Spine. 2024;49:884–91.
[62] Suzuki H, Kokabu T, Yamada K, et al. Deep learning-based detection of lumbar spinal canal stenosis using convolutional neural networks. Spine J. 2024;24:2086–101.
[63] Won D, Lee HJ, Lee SJ, Park SH. Lumbar spinal stenosis grading in multiple level magnetic resonance imaging using deep convolutional neural networks. Glob. Spine J. 2024;15:2309–17.
[64] Löffler MT, Jacob A, Scharr A, et al. Automatic opportunistic osteoporosis screening in routine CT: improved prediction of patients with prevalent vertebral fractures compared to DXA. Eur Radiol. 2021;31:6069–77.
[65] Nian SQ, Zhao YY, Li CJ, et al. Development and validation of a radiomics-based model for predicting osteoporosis in patients with lumbar compression fractures. Spine J. 2024;24:1625–34.
[66] Oh J, Kim B, Oh G, Hwangbo Y, Ye JC. End-to-end semi-supervised opportunistic osteoporosis screening using computed tomography. Endocrinol Metab (Seoul). 2024;39:500–10.
[67] Oh S, Kang WY, Park H, et al. Evaluation of deep learning-based quantitative computed tomography for opportunistic osteoporosis screening. Sci Rep. 2024;14:9.
[68] Wu Y, Yang XP, Wang MY, et al. Artificial intelligence assisted automatic screening of opportunistic osteoporosis in computed tomography images from different scanners. Eur Radiol. 2024;35:2287–95.
[69] Zhang J, Xia L, Zhang XL, et al. Development and validation of a predictive model for vertebral fracture risk in osteoporosis patients. Eur Spine J. 2024;33:3242–60.
[70] Li YC, Chen HH, Horng-Shing Lu H, Hondar Wu H-T, Chang M-C, Chou P-H. Can a deep-learning model for the automated detection of vertebral fractures approach the performance level of human subspecialists? Clin Orthop Relat Res. 2021;479:1598–612.
[71] Ma S, Huang Y, Che X, Gu R. Faster RCNN-based detection of cervical spinal cord injury and disc degeneration. J Appl Clin Med Phys. 2020;21:235–43.
[72] Kishikawa J, Kobayakawa K, Saiwai H, et al. Verification of the accuracy of cervical spinal cord injury prognosis prediction using clinical data-based artificial neural networks. J. Clin. Med. 2024;13:253.
[73] Zhang ZF, Li N, Ding Y, Cheng H. An integrative nomogram based on MRI radiomics and clinical characteristics for prognosis prediction in cervical spinal cord injury. Eur Spine J. 2024;34:1164–76.
[74] Zhang K, Xu N, Guo C, Wu J. MPF-net: an effective framework for automated cobb angle estimation. Med Image Anal. 2022;75:102277.
[75] Zhang D, Chen B, Li S. Sequential conditional reinforcement learning for simultaneous vertebral body detection and segmentation with modeling the spine anatomy. Med Image Anal. 2021;67:101861.
[76] Fang Y, Li W, Chen X, et al. Opportunistic osteoporosis screening in multi-detector CT images using deep convolutional neural networks. Eur Radiol. 2021;31:1831–42.
[77] Zheng HD, Sun YL, Kong DW, et al. Deep learning-based high-accuracy quantitation for lumbar intervertebral disc degeneration from MRI. Nat Commun. 2022;13:12.
[78] Liebmann F, von Atzigen M, Stütz D, et al. Automatic registration with continuous pose updates for marker-less surgical navigation in spine surgery. Med Image Anal. 2024;91:103027.
[79] Pang SM, Pang CL, Su ZH, et al. DGMSNet: spine segmentation for MR image by a detection-guided mixed-supervised segmentation network. Med Image Anal. 2022;75:102261.
[80] Xie L, Wisse LEM, Wang JC, et al. Deep label fusion: a generalizable hybrid multi-atlas and deep convolutional neural network for medical image segmentation. Med Image Anal. 2023;83:17.
[81] Huang MY, Zhou SL, Chen XM, Lai H, Feng Q. Semi-supervised hybrid spine network for segmentation of spine MR images. Comput Med Imaging Graph. 2023;107:102245.
[82] Serrador L, Villani FP, Moccia S, Santos CP. Knowledge distillation on individual vertebrae segmentation exploiting 3D U-Net. Comput Med Imaging Graph. 2024;113:102350.
[83] Saravi B, Guzel HE, Zink A, et al. Synthetic 3D spinal vertebrae reconstruction from biplanar X-rays utilizing generative adversarial networks. J Pers Med. 2023;13:1642.
[84] He Z, Lu N, Chen Y, et al. Conditional generative adversarial network-assisted system for radiation-free evaluation of scoliosis using a single smartphone photograph: a model development and validation study. Eclinicalmedicine. 2024;75:102779.
[85] Meng N, Wong KK, Zhao M, Cheung JPY, Zhang T. Radiograph-comparable image synthesis for spine alignment analysis using deep learning with prospective clinical validation. EClinicalMedicine. 2023;61:102050.
[86] Gafencu MA, Velikova Y, Saleh M, et al. Shape completion in the dark: completing vertebrae morphology from 3D ultrasound. Int. J. Comput. Assist. Radiol. Surg. 2024;19:1339–47.
[87] Wong JS, Reformat M, Parent E, Lou E. Validity and accuracy of automatic cobb angle measurement on 3D spinal ultrasonographs for children with adolescent idiopathic scoliosis: SOSORT 2024 award winner. Eur Spine J. 202534:1622–30.
[88] Ward M, Maity A, Brown EDL, et al. Analysis of ChatGPT in the triage of common spinal complaints. World Neurosurg. 2024;192:e273–80.
[89] Subramanian T, Araghi K, Amen TB, et al. Chat generative pretraining transformer answers patient-focused questions in cervical spine surgery. Clin. Spine Surg. 2024;37:E278–81.
[90] Nian PP, Saleet J, Magruder M, et al. ChatGPT as a source of patient information for lumbar spinal fusion and laminectomy. Clin. Spine Surg. 2024;37:E394–403.
[91] Ito S, Nakashima H, Yoshii T, et al. Deep learning-based prediction model for postoperative complications of cervical posterior longitudinal ligament ossification. Eur Spine J. 2023;32:3797–806.
[92] Lee CH, Jo DJ, Oh JK, et al. Development and validation of an online calculator to predict proximal junctional kyphosis after adult spinal deformity surgery using machine learning. Neurospine. 2023;20:1272–80.
[93] Zhang Y, Wan DH, Chen M, et al. Automated machine learning-based model for the prediction of delirium in patients after surgery for degenerative spinal disease. CNS Neurosci Ther. 2023;29:282–95.
[94] Gonzalez-Suarez AD, Rezaii PG, Herrick D, et al. Using machine learning models to identify factors associated with 30-day readmissions after posterior cervical fusions: a longitudinal cohort study. Neurospine. 2024;21:620–32.
[95] Shahrestani S, Reardon T, Brown NJ, et al. Developing mixed-effects models to compare the predictive ability of various comorbidity indices in a contemporary cohort of patients undergoing lumbar fusion. Neurosurgery. 2024;94:711–20.
[96] Wang SK, Wang P, Li ZE, Li X-Y, Kong C, Lu S-B. Development and external validation of a nomogram for predicting postoperative adverse events in elderly patients undergoing lumbar fusion surgery: comparison of three predictive models. J. Orthop. Surg. Res. 2024;19:11.
[97] Li Z, Jiang S, Song X, et al. Collaborative spinal robot system for laminectomy: a preliminary study. Neurosurg Focus. 2022;52:E11.
[98] Li Z, Wang C, Song X, et al. Accuracy evaluation of a novel spinal robotic system for autonomous laminectomy in thoracic and lumbar vertebrae: a cadaveric study. J Bone Joint Surg Am. 2023;105:943–50.
[99] Li X, Chen J, Wang B, et al. Evaluating the status and promising potential of robotic spinal surgery systems. Orthop Surg. 2024;16:2620–32.
[100] Poore GD, Kopylova E, Zhu Q, et al. RETRACTED ARTICLE: Microbiome analyses of blood and tissues suggest cancer diagnostic approach. Nature. 2020;579:567–74.
[101] Chen P, Ye J, Wang G, et al. GMAI-MMBench: a comprehensive multimodal evaluation benchmark towards general medical AI. ArXiv. 2024;abs/2408.03361.
[102] Kochanski RB, Lombardi JM, Laratta JL, Lehman RA, O'Toole JE. Image-guided navigation and robotics in spine surgery. Neurosurgery. 2019;84:1179–89.
[103] Móga K, Hölgyesi A, Zrubka Z, Péntek M, Haidegger T. Augmented or mixed reality enhanced head-mounted display navigation for in vivo spine surgery: a systematic review of clinical outcomes. J Clin Med. 2023;12:3788.
[104] Azad TD, Warman A, Tracz JA, Hughes LP, Judy BF, Witham TF. Augmented reality in spine surgery - past, present, and future. Spine J. 2024;24:1–13.