

AIIS **Artificial Intelligence Surgery**

Elevating Spine Surgery with AI: Current Landscape and Future Frontiers



Guest Editor:



Prof. Peter G. Passias

Topic: Elevating Spine Surgery with AI: Current Landscape and Future Frontiers



Prof. Peter G. Passias

Division of Spinal Surgery, Departments of Neurological and Orthopaedic Surgery, Duke University Hospital and Duke School of Medicine, Durham, United States.

Prof. Peter G. Passias is a world leader in the surgical treatment of spinal disorders. His clinical practice focuses on the treatment of both degenerative conditions of the spine, with an emphasis on complex spinal deformities including thoracolumbar revision procedures and scoliosis. He has a particular interest in the management of complex cervical spine disorders, and is widely recognized as a pioneer in this field. Prof. Passias is the first of his peers to perform several minimally invasive procedures aimed at expediting recovery and achieving the very best clinical outcomes.

Special Issue introduction:

The advancement of research geared toward a more comprehensive understanding of spinal conditions continually provides information about ever-evolving treatment options. However, tailoring optimal treatment plans for individual patients is intricate, often requiring a comprehensive consideration of their medical history and individual characteristics. Furthermore, clinicians and researchers face difficulties in accessing relevant data due to patient privacy concerns, hindering their seamless integration with current guidelines and novel indices. Artificial Intelligence (AI) is a technology that emulates human cognitive abilities through specialized programs and computers. Its applications in everyday clinical practice include surgical robots that enhance the precision of spinal surgeries and the prediction of patient outcomes based on various radiographic, clinical, and surgical parameters. Additionally, such systems can create virtual artificial patient populations mirroring real patient characteristics. The purpose of this Special Issue is to explore the contributions and merits of AI and to investigate the current state and future trends of AI in the field of spinal surgery.

EDITORIAL BOARD

Editor Emeritus

Brice Gayet (France)

Editor-in-Chief

Andrew Gumbs (Germany)

Honorary Regional Editors

Michel Gagner (Canada)

Takeaki Ishizawa (Japan)

Jacques F. Marescaux (France)

Computer Vision Editor

Adrian Park (USA)

Robotic Tele-manipulation Editor

Sharon B Ross (USA)

Tele-Medicine Editor

Dennis L. Fowler (USA)

Autonomous Actions Editor

Lee Swanström (France)

AI Endoscopic Surgery Editor

Silvana Perretta (France)

Associate Editors

Mohammed Abu Hilal (Italy)

Elie Chouillard (France)

Roland Croner (Germany)

Eyad Elyan (UK)

Isabella Frigerio (Italy)

Gaya Spolverato (Italy)

Technology Development Editor

Luca Milone (USA)

Statistics Editor

Clarissa Ferrari (Italy)

Ethics Editor

Antonella Ficorilli (Italy)

Director of Social Media and Strategy

S. Vincent Grasso (USA)

Editorial Board Members

Sagheer Abbas (Saudi Arabia)

Gudrun Aspelund (Iceland)

Fabio Ausania (Spain)

Elena Giovanna Bignami (Italy)

Madison Cuffy (USA)

Belinda De Simone (France)

Michele Diana (Switzerland)

Michael Friebe (Germany)

Arnar Geirsson (USA)

Suzanne S. Gisbertz (The Netherlands)

Jianxing He (China)

Alfredo Illanes (Germany)

Chrisina Jayne (UK)

Long R Jiao (UK)

Ulf Dietrich Kahlert (Germany)

W. Konrad Karcz (Germany)

Stephen Kavic (USA)

Zain Khalpey (USA)

Michael Kreissl (Germany)

Marialuisa Lugaresi (Italy)

Michele Masetti (Italy)

Beat P. Müller (Switzerland)

Zbigniew Nawrat (Poland)

Derek O'Reilly (China)

Mehmet Mahir Özmen (Turkey)

Peter Passias (USA)

Subroto Paul (USA)

Carla Pugh (USA)

Qahwaji Rami (UK)

Francesca Ratti (Italy)

Bettina M. Rau (Germany)

Thomas Schnelldorfer (USA)

Vivian Strong (USA)

Patricia Sylla (USA)

Amir Szold (Israel)

Martin Teraa (The Netherlands)

Yogesh K. Vashist (Saudi Arabia)

Xin Wang (China)

GENERAL INFORMATION

About the Journal

Artificial Intelligence Surgery (AIS) is a surgical journal that focuses on the interface of computers and the art of surgery. The aim of this journal is to create a space for discourse in terms of topics related to artificial intelligence (AI) in surgery such as smart surgical technology/digital surgery, computer-assisted surgical systems and surgical data science. The intention research direction on robotics includes clinical and technological work on tele-manipulation, robotic-assisted technologies, robotic Natural Orifice Transluminal Endoscopic Surgery (NOTES) and fully autonomous robots. Priority will be given to research and clinical studies of augmented reality, image-guidance, radiomics and 3D printing. As opposed to many other journals on surgical technology, this journal will not be limited to laparoscopy or other forms of minimally invasive surgery alone, but accept submissions that possess significance and scientific excellence in AI, computer-enhanced imaging and any automation as it pertains to open surgery, endoscopy, NOTES and interventional radiology.

Information for Authors

Manuscripts should be prepared in accordance with Author Instructions.

Please check www.oaepublish.com/ais/author_instructions for details.

All manuscripts should be submitted online at <https://www.oaecenter.com/login?JournalId=ais>.

Copyright

The entire contents of the *AIS* are protected under international copyrights. The journal, however, grants to all users a free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, perform and display the work publicly and to make and distribute derivative works in any digital medium for any reasonable purpose, subject to proper attribution of authorship and ownership of the rights. The journal also grants the right to make small numbers of printed copies for their personal use under the Creative Commons Attribution 4.0 License.

Copyright is reserved by ©The Author(s) 2025.

Permissions

For information on how to request permissions to reproduce articles/information from this journal, please visit www.oaepublish.com/ais.

Disclaimer

The information and opinions presented in the journal reflect the views of the authors and not of the journal or its Editorial Board or the Publisher. Publication does not constitute endorsement by the journal. Neither the *AIS* nor its publishers nor anyone else involved in creating, producing or delivering the *AIS* or the materials contained therein, assumes any liability or responsibility for the accuracy, completeness, or usefulness of any information provided in the *AIS*, nor shall they be liable for any direct, indirect, incidental, special, consequential or punitive damages arising out of the use of the *AIS*. The *AIS*, nor its publishers, nor any other party involved in the preparation of material contained in the *AIS* represents or warrants that the information contained herein is in every respect accurate or complete, and they are not responsible for any errors or omissions or for the results obtained from the use of such material. Readers are encouraged to confirm the information contained herein with other sources.

Publisher

OAE Publishing Inc.

245 E Main Street st112, Alhambra, CA 91801, USA

Website: www.oaepublish.com

Contacts

E-mail: editorialoffice@aisjournal.net

Website: www.oaepublish.com/ais

CONTENTS

Topic: Elevating Spine Surgery with AI: Current Landscape and Future Frontiers

- 1 Assessing the accuracy and utility of ChatGPT responses to patient questions regarding posterior lumbar decompression**
Alec M. Giakas, Rajkishen Narayanan, Teeto Ezeonu, Jonathan Dalton, Yunsoo Lee, Tyler Henry, John Mangan, Gregory Schroeder, Alexander Vaccaro, Christopher Kepler
Art Int Surg 2024;4:233-46 <http://dx.doi.org/10.20517/ais.2024.24>
- 2 Machine learning applications in adult spinal deformity corrective surgery: a narrative review**
Nader Toossi, Ozhan Jerry
Art Int Surg 2024;4:258-66 <http://dx.doi.org/10.20517/ais.2024.27>
- 3 Patient perspectives on AI: a pilot study comparing large language model and physician-generated responses to routine cervical spine surgery questions**
Ezra T. Yoseph, Aneysis D. Gonzalez-Suarez, Siegmund Lang, Atman Desai, Serena S. Hu, Corinna C. Zygourakis
Art Int Surg 2024;4:267-77 <http://dx.doi.org/10.20517/ais.2024.38>
- 4 Navigating artificial intelligence in spine surgery: implementation and optimization across the care continuum**
Antony A. Fuleihan, Arjun K. Menta, Tej D. Azad, Kelly Jiang, Carly Weber-Levine, A. Daniel Davidar, Andrew M. Hersh, Nicholas Theodore
Art Int Surg 2024;4:288-95 <http://dx.doi.org/10.20517/ais.2024.39>
- 5 Redefining precision: the current and future roles of artificial intelligence in spine surgery**
Ryan W. Turlip, Harmon S. Khela, Mert Marcel Dagli, Daksh Chauhan, Yohannes Ghenbot, Hasan S. Ahmad, Jang W. Yoon
Art Int Surg 2024;4:324-30 <http://dx.doi.org/10.20517/ais.2024.29>
- 6 Beyond AI and robotics: the dawn of surgical automation in spine surgery**
Nishanth S. Sadagopan, Dillan Prasad, Rishi Jain, Christopher Ahuja, Nader S. Dahdaleh, Najib E. El Teele
Art Int Surg 2024;4:387-400 <http://dx.doi.org/10.20517/ais.2024.34>
- 7 Pseudarthrosis following adult spinal deformity surgery may be predicted with preoperative MRI adipose tissue features: an artificial intelligence study on raw 3D imaging**
Graham W. Johnson, Hani Chanbour, Derek J. Doss, Ghassan S. Makhoul, Amir M. Abtahi, Byron F. Stephens, Scott L. Zuckerman
Art Int Surg 2024;4:401-10 <http://dx.doi.org/10.20517/ais.2024.40>

- 8 Deep learning for automated spinopelvic parameter measurement from radiographs: a meta-analysis**
Dylan Glaser, Ahmad K. AlMekkawi, James P. Caruso, Candace Y. Chung, Eshal Z. Khan, Hicham M. Daadaa, Salah G. Aoun, Carlos A. Bagley
Art Int Surg 2025;5:1-15 <http://dx.doi.org/10.20517/ais.2024.36>
- 9 Spino-plastic surgery, back to the future**
Casey Martinez, Camryn Payne, Jonathan L. Jeger, Jonathan L. Jeger, Sebastian Winocou, Maziyar A. Kalani, Michael Bohl, Michael Bohl, Edward M. Reece
Art Int Surg 2025;5:16-23 <http://dx.doi.org/10.20517/ais.2024.73>
- 10 Has artificial intelligence in spine surgery lived up to the hype? A narrative review of recent approaches, current challenges, and the path forward**
Vardhaan S. Ambati, Satvir Saggi, Abraham Dada, Nima Alan
Art Int Surg 2025;5:53-64 <http://dx.doi.org/10.20517/ais.2024.45>

Original Article

Open Access



Assessing the accuracy and utility of ChatGPT responses to patient questions regarding posterior lumbar decompression

Alec M. Giakas , Rajkishen Narayanan, Teeto Ezeonu, Jonathan Dalton, Yunsoo Lee, Tyler Henry, John Mangan, Gregory Schroeder, Alexander Vaccaro, Christopher Kepler

Department of Orthopaedic Surgery, Rothman Institute at Thomas Jefferson University, Philadelphia, PA 19107, USA.

Correspondence to: Dr. Rajkishen Narayanan, Department of Orthopaedic Surgery, Rothman Institute at Thomas Jefferson University, 925 Chestnut Street Floor 5, Philadelphia, PA 19107, USA. E-mail: rajkishen.narayanan@gmail.com

How to cite this article: Giakas AM, Narayanan R, Ezeonu T, Dalton J, Lee Y, Henry T, Mangan J, Schroeder G, Vaccaro A, Kepler C. Assessing the accuracy and utility of ChatGPT responses to patient questions regarding posterior lumbar decompression. *Art Int Surg* 2024;4:233-46. <https://dx.doi.org/10.20517/ais.2024.24>

Received: 25 Apr 2024 **First Decision:** 22 Jul 2024 **Revised:** 15 Aug 2024 **Accepted:** 22 Aug 2024 **Published:** 4 Sep 2024

Academic Editor: Andrew A. Gumbs **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Aim: To examine the clinical accuracy and applicability of ChatGPT answers to commonly asked questions from patients considering posterior lumbar decompression (PLD).

Methods: A literature review was conducted to identify 10 questions that encompass some of the most common questions and concerns patients may have regarding lumbar decompression surgery. The selected questions were then posed to ChatGPT. Initial responses were then recorded, and no follow-up or clarifying questions were permitted. Two attending fellowship-trained spine surgeons then graded each response from the chatbot using a modified Global Quality Scale to evaluate ChatGPT's accuracy and utility. The surgeons then analyzed each question, providing evidence-based justifications for the scores.

Results: Minimum scores across all ten questions would lead to a total score of 20, whereas a maximum score would be 100. ChatGPT's responses in this analysis earned a score of 59, just under an average score of 3, when evaluated by two attending spine surgeons. A score of 3 denoted a somewhat useful response of moderate quality, with some important information adequately discussed but some poorly discussed.

Conclusion: ChatGPT has the ability to provide broadly useful responses to common preoperative questions that patients may have when considering undergoing PLD. ChatGPT has excellent utility in providing background



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



information to patients and in helping them become more informed about their pathology in general. However, it often lacks the specific patient context necessary to provide patients with personalized, accurate insights into their prognosis and medical options.

Keywords: Artificial intelligence, ChatGPT, lumbar decompression, spine surgery

INTRODUCTION

In recent decades, internet search engines have transformed the way patients seek health information and interact with the health care system^[1,2]. Accordingly, almost 75% of adult internet users have searched for health information online. Within the orthopedic patient population, up to two-thirds of patients have used the Internet to search for information about their conditions^[3,4]. Recently, open-access artificial intelligence (AI) has become a popular tool for information-gathering in healthcare^[5-7]. ChatGPT, launched in November 2022, is an AI-powered language processing tool that has the ability to integrate billions of datapoints and fine-tune successive responses to mimic human conversation^[8,9]. Recently, the AI chatbot garnered attention within the medical community for earning a passing grade on a United States Medical Licensing Examination practice test, an exam that involves the application of knowledge and critical thinking^[10].

As patients become more familiar with AI platforms, they may increasingly look to these programs to provide instantaneous answers to questions about their health. Previous studies have examined ChatGPT's utility in clinical decision making and patient education within orthopedics^[11-16]. In a recent paper, Mika *et al.* attempted to determine if ChatGPT could answer frequently asked questions related to total hip arthroplasty^[17]. In their conclusion, the authors noted ChatGPT's utility in the clinical space by providing evidence-based responses, but acknowledged that responses often required at least some clarification. Similarly, in another paper, AI-generated answers to common patient questions regarding minimally invasive spine surgery (MISS) were explored. While authors acknowledged ChatGPT's ability to accurately describe procedures, indications, outcomes, and complications of MISS, they also noted its tendency to make unsubstantiated claims and potentially confuse patients as they make decisions regarding their treatment^[18]. Given the concerning discrepancies that these prior studies have identified, additional scrutiny into AI's ability to accurately aid patients and physicians in the shared decision-making process is necessary^[19].

Degenerative spine conditions are a prominent health concern, particularly within the aging population, that can cause patients significant physical, mental, and emotional distress. Posterior lumbar decompression (PLD), a procedure performed to treat degenerative spine conditions, is one of the most commonly performed procedures by spine surgeons^[20,21]. The utility and limitations of ChatGPT in answering common preoperative questions for patients considering PLD have yet to be explored. Furthermore, while prior studies in the adult reconstruction literature have attempted to score and analyze ChatGPT responses, no previous study has methodologically scored AI's responses to common questions asked by patients considering spine surgery. As such, the purpose of the present study is to evaluate the clinical accuracy and applicability of ChatGPT answers to commonly asked questions from patients considering PLD.

METHODS

Methods

Question selection

First, a literature review was performed for commonly asked questions by patients considering orthopedic surgery - 17 questions were identified^[4,12,15,18]. Next, we reviewed the “frequently asked questions” on the websites of 10 prominent healthcare institutions to further identify potential patient questions. 5 additional questions were pooled from this website review. Lastly, we queried ChatGPT itself for common questions patients ask when considering spine surgery. 12 questions were elicited from ChatGPT itself. Ultimately, 10 questions were selected from these three sources by two attending, fellowship-trained orthopedic spine surgeons at a high-volume academic institution. The questions selected were felt to encompass some of the most common questions and pressing concerns patients may have regarding lumbar decompression surgery.

ChatGPT queries

The selected 10 questions were then posed to ChatGPT (Version 4.0) using its free online interface (<https://chat.openai.com>). Initial responses were then recorded, and no follow-up or clarifying questions were permitted. Responses have been truncated in this manuscript for the purpose of publication. Complete responses can be found in the [Supplementary Materials](#).

Analysis

The two previously mentioned attending orthopedic spine surgeons analyzed each response from the chatbot using a modified Global Quality Scale to assign a score out of 5 possible points evaluating ChatGPT’s accuracy and utility [Table 1]^[22]. A score of 5 was reserved for evidence-based, comprehensive responses of excellent quality that were both accurate and very useful for patients. A score of 4 denoted good quality, useful responses with minimal unsubstantiated claims. A score of 3 signified a response of moderate quality, somewhat useful for patients with some poorly discussed important information. A score of 2 denoted a generally poor-quality response, with many important topics missing and some unsubstantiated claims that have very little use to patients. Finally, a score of 1 signified a poor-quality response lacking any utility with missing information and multiple unsubstantiated claims. Scores given to all responses by each attending were recorded [Table 1].

RESULTS

1. When should I get lumbar decompression surgery?

The decision to undergo lumbar decompression surgery should be made in consultation with a qualified healthcare professional, preferably a spine specialist or orthopedic surgeon. The timing of surgery depends on various factors, including the severity of your symptoms, the underlying condition causing the symptoms, your overall health, and the effectiveness of non-surgical treatments. Here are some considerations for determining when you might consider lumbar decompression surgery:

- (1) Conservative treatment failure
- (2) Persistent severe pain
- (3) Progressive neurological deficits
- (4) Loss of function
- (5) Quality of life impact
- (6) Emergent cases
- (7) Health status
- (8) Patient preference

Table 1. Modified Global Quality Scale items

#	Item
1	Poor quality, multiple unsubstantiated claims, most information missing. Not at all useful for patients
2	Generally poor quality, some unsubstantiated claims, some information listed but many important topics missing. Very limited use to patients
3	Moderate quality, some important information is adequately discussed, but others poorly discussed. Somewhat useful for patients
4	Good quality, minimal unsubstantiated claims, most of the relevant information is listed, but some topics not covered. Useful for patients
5	Excellent quality, comprehensive claims that are accurate and overall evidence-based. Very useful for patients

Attending 1 Score: 4

Attending 2 Score: 3

Analysis: ChatGPT provided a basic, but diverse general overview of why a patient might consider PLD. The answer correctly identifies several general reasons for considering PLD, notably including “emergent cases”. However, it does not mention any specific diagnoses that would be indications for surgery, such as neurogenic claudication or radiculopathy^[23-25]. Responses such as “health status” and “patient preference” are quite vague. In particular, “patient preference” may be interpreted broadly as indicating that surgery is simply optional, and many patients may use this as justification to avoid surgery. This answer should be qualified with a warning that this is highly dependent on both the clinical scenario and underlying pathology and that irreversible pain or neurologic deficit may result from delaying or avoiding surgery without consulting with a spine surgeon. Finally, this answer overall lacks evidence-based responses that might be useful to patients considering surgery.

2. What will happen if I delay lumbar decompression surgery?

Delaying lumbar decompression surgery might have various implications depending on your specific condition and symptoms. While some conditions may tolerate a certain degree of delay, others might worsen over time. It is important to consult with a healthcare professional, ideally a spine specialist, to understand the potential consequences of delaying surgery in your individual case. Here are some general considerations:

- (1) Progression of symptoms: If you are experiencing pain, weakness, numbness, or other neurological symptoms due to nerve compression, delaying surgery might lead to the progression of these symptoms. Nerve compression can cause permanent damage if left untreated.
- (2) Functional limitations: Nerve compression can affect your ability to perform daily activities, work, and participate in recreational activities. Delaying surgery might limit your quality of life and functional abilities.
- (3) Quality of life: The impact of your symptoms on your overall well-being and quality of life is an important factor to consider. If your symptoms are significantly affecting your day-to-day life, addressing them sooner through surgery might be beneficial.
- (4) Impact on surgical outcome: In some cases, delaying surgery might lead to more complex surgery later, as conditions can progress, and tissue changes might occur over time.
- (5) Psychological impact: Chronic pain and limited mobility can have psychological effects, including anxiety and depression. Addressing these issues through timely treatment might prevent their exacerbation.
- (6) Risk of complications: If your symptoms are related to structural issues such as herniated discs, delaying surgery could increase the risk of complications associated with nerve compression or other related problems.

Attending 1 Score: 3

Attending 2 Score: 3

Analysis: This ChatGPT response provides a somewhat useful overview of the potential results of delaying surgery. However, it lacks any specific information as to the consequences of “nerve damage” that might be useful to the patient considering surgery. Instead of providing specific information, such as increasing muscle weakness, or loss of bowel or bladder function, the response focuses more on how delaying surgery may impact a patient’s quality of life in general^[26,27]. This response is reasonable, but lacks the specificity and comprehensiveness needed by patients considering PLD. The complications associated with delaying lumbar decompression surgery are vague and do not convey the potential severity involved with this decision. Additionally, no data are provided to the patient in terms of the percentage of complications from non-operative treatment, which may provide some context to the patient trying to understand the scope of this choice.

3. What are the alternatives to lumbar decompression surgery?

There are several alternatives to lumbar decompression surgery that may be considered depending on your specific condition, symptoms, and overall health. It is important to discuss these alternatives with a qualified healthcare professional, preferably a spine specialist, to determine the most suitable treatment plan for your individual situation. Here are some non-surgical alternatives to lumbar decompression surgery:

- (1) Physical therapy: Physical therapy can help strengthen muscles, improve flexibility, and provide strategies to manage pain. Therapists can design exercise programs tailored to your needs.
- (2) Medications: Non-prescription pain relievers (e.g., acetaminophen, NSAIDs) or prescription medications (e.g., muscle relaxants, neuropathic pain medications) can help manage pain and inflammation.
- (3) Epidural steroid injections: These injections can help reduce inflammation and provide temporary pain relief. They are often used when pain is primarily due to nerve irritation.
- (4) Lifestyle modifications: Making changes to your daily habits, such as maintaining a healthy weight, improving posture, and avoiding activities that exacerbate your symptoms, can help manage pain and prevent worsening of your condition.
- (5) Chiropractic care: Chiropractors use manual manipulation techniques to improve spinal alignment and alleviate pain. However, caution should be exercised if chiropractic care is considered for certain conditions.
- (6) Bracing: Some conditions might benefit from wearing a supportive brace to help stabilize the spine and reduce pain during certain activities.
- (7) Watchful waiting: In some cases, especially for mild symptoms, your healthcare provider might recommend monitoring your condition over time to see if it improves or worsens before considering more invasive treatments.

Attending 1 Score: 3

Attending 2 Score: 4

Analysis: This response provides a good summation of alternative options to surgery that might be useful to patients. However, it gives minimal detail regarding when these alternatives might be sufficient treatments, and thus risks patients delaying or avoiding necessary surgery. In such instances when surgery is indicated, initiating or continuing non-operative treatment might lead to progressive and permanent neurological impairment for patients^[28,29]. As such, while the response lists, in appropriate detail, options that might be appropriate for some patients, pursuing these options might be highly detrimental for other patients.

Furthermore, these responses do not elaborate on the potential adverse effects associated with various treatment options, thus providing a weak and incomplete risk-benefit analysis for patients to consider when deciding whether to proceed with PLD. Within the descriptions of the alternative therapies, even for the appropriate patient, there is minimal detail on what these methods entail. For example, the physical therapy response does not describe what muscles are strengthened or why, nor how much time may elapse prior to symptom alleviation.

4. What are the risks of lumbar decompression surgery?

Lumbar decompression surgery is generally considered safe and effective for relieving symptoms caused by nerve compression in the lower back. However, like any surgical procedure, it carries certain risks and potential complications. It is important to discuss these risks with your surgeon before undergoing the surgery. Here are some potential risks of lumbar decompression surgery:

- (1) Infection: There is a risk of developing a surgical site infection, which can occur at the incision site or within the deeper tissues. Infections may require additional treatment, such as antibiotics or drainage.
- (2) Bleeding: While uncommon, there is a risk of excessive bleeding during or after surgery. In some cases, a blood transfusion might be necessary.
- (3) Nerve injury: Surgery involves working near nerves, and there is a small risk of damaging adjacent nerves, which could result in new or worsened neurological symptoms.
- (4) Dural tear: The dura is a thin, protective membrane surrounding the spinal cord and nerves. During surgery, the dura can accidentally be torn, resulting in cerebrospinal fluid leakage, which may require additional treatment.
- (5) Persistent or worsened symptoms: In some cases, surgery may not completely relieve symptoms, or there could be temporary worsening of symptoms immediately after surgery due to inflammation or nerve irritation.
- (6) Blood clots: Prolonged immobility during and after surgery can increase the risk of developing blood clots, which can be dangerous if they travel to the lungs (pulmonary embolism) or other parts of the body.
- (7) Anesthesia complications: There are risks associated with anesthesia, including allergic reactions, breathing difficulties, and adverse reactions to medications.
- (8) Instrumentation failure: If hardware (such as screws or rods) is used to stabilize the spine, there is a risk of instrumentation failure, including loosening, breakage, or migration.

Attending 1 Score: 3

Attending 2 Score: 3

Analysis: The response provided a somewhat useful, albeit cursory, overview of the potential complications involved in PLD. These complications are often addressed or raised in informed consent conversations with patients prior to surgery and, therefore, may serve as a useful introduction to these issues prior to an office visit. However, the mention of “instrumentation failure”, among other complications, might confuse patients regarding the use of hardware in a simple, uninstrumented PLD. Additionally, the response does not provide any clarity regarding the actual prevalence of these adverse events for patients undergoing PLD, which would provide crucial context to anyone considering surgery.

5. What is the recovery time after lumbar decompression surgery?

Here’s a general timeline for recovery after lumbar decompression surgery:

Immediate postoperative period (days to weeks):

- Hospital stay: Most patients stay in the hospital for a short period, usually a day or two, after lumbar decompression surgery.
- Pain management: Pain and discomfort are common after surgery. Pain medication and other methods are used to manage pain.
- Mobility: Patients are encouraged to start walking and moving around shortly after surgery to prevent complications like blood clots.

First two months:

- Gradual increase in activity: Patients can gradually increase their activity level under the guidance of their healthcare provider.
- Physical therapy: Physical therapy may be recommended to help improve strength, flexibility, and overall mobility.
- Return to work: Depending on the type of work, patients may be able to return to work within several weeks to a few months, but this varies.

Long-term:

- Full recovery: It can take up to a year or more to fully recover and achieve the maximum benefits of the surgery.
- Activity and lifestyle: Patients are often able to return to their normal activities and routines, though some modifications may be necessary to prevent future issues.

Attending 1 Score: 2

Attending 2 Score: 4

Analysis: Attendings differed in their opinion of the adequacy of Chat GPT's response to this question. It was acknowledged that this response was a good overview of a recovery timeline after PLD, while noting that it might mislead patients by not mentioning that many patients can be discharged on the day of surgery^[30]. Another deficiency was the response's lack of specific information regarding patient activities postoperatively, which is one of the main concerns of patients considering surgery. Additionally, the response was seen as vague regarding return to work and potential restrictions. Lastly, the full recovery timeline was perceived to be inaccurate, as many patients return to full functionality well before a year after PLD^[31].

6. How will lumbar decompression surgery affect my range of motion?

Lumbar decompression surgery can have varying effects on your range of motion depending on the specific procedure performed, the extent of the decompression, and individual factors. Here is a general overview of how lumbar decompression surgery might affect your range of motion:

- (1) Minimally invasive approaches: Many minimally invasive lumbar decompression surgeries aim to minimize disruption to surrounding tissues and muscles. As a result, these procedures can potentially lead to less alteration in your range of motion compared to traditional open surgeries.
- (2) Traditional open surgery: Traditional open lumbar decompression surgery might involve larger incisions and more tissue disruption. There might be some impact on the flexibility and range of motion of the spine in the immediate postoperative period.
- (3) Microdiscectomy: If your surgery is focused on removing a herniated disc or portion of a disc, the impact on range of motion might be minimal.

(4) Laminectomy or laminotomy: These procedures involve removing part of the lamina (a bony arch in the spine) to relieve pressure on the spinal cord or nerves. While this can increase space for nerve structures, it might have a slight impact on stability and range of motion.

(5) Fusion procedures: If fusion is performed in conjunction with decompression, it might lead to a reduction in the range of motion at the fused segment.

It is important to note that your surgeon will strive to achieve a balance between decompression and preserving spinal stability and function. They will discuss the potential impact on your range of motion during the preoperative consultation, taking into consideration your specific condition, treatment goals, and lifestyle.

Attending 1 Score: 4

Attending 2 Score: 2

Analysis: Attendings disagreed on the adequacy of ChatGPT's response to this question. A lower score from one attending was given because ChatGPT's claim regarding the superiority of MISS in preserving range of motion is not conclusively supported by current spine surgery literature^[32]. It was also felt that the mention of fusion could confuse patients considering PLD, which is typically handled without a concomitant fusion procedure. It was concluded that the response provided a satisfactory but uncomprehensive outline of motion preservation for different techniques and indications for PLD.

7. Will lumbar decompression surgery help my back pain?

Lumbar decompression surgery is primarily aimed at relieving symptoms caused by nerve compression, such as leg pain, weakness, numbness, and other neurological symptoms. While it can lead to a significant reduction or elimination of leg-related symptoms, its impact on back pain itself can vary. Here are some key points to consider:

(1) Secondary improvement: Lumbar decompression surgery is primarily designed to alleviate symptoms related to nerve compression. If your back pain is directly related to nerve compression or irritation, surgery might lead to a reduction in back pain as well.

(2) Mixed results: The impact of lumbar decompression surgery on isolated back pain (pain that is not primarily due to nerve compression) can be less predictable. Some patients experience relief, while others might have varying degrees of improvement.

(3) Degenerative disc disease: If your back pain is due to degenerative disc disease and is not primarily nerve-related, lumbar decompression surgery might not target the underlying cause of your pain. Other surgical options, like spinal fusion, might be considered in such cases.

(4) Patient expectations: It is important to have realistic expectations about the outcomes of surgery. Discuss your goals and expectations with your surgeon during your consultations.

Attending 1 Score: 4

Attending 2 Score: 2

Analysis: The reviewers generally agreed regarding the accuracy of the content provided in this response. While the response was seen as useful for patients considering PLD, it was noted that it also could be perplexing for patients without a medical background who might have numerous follow-up questions regarding the difference between "back pain" vs. "nerve pain". While many medical professionals can identify the difference between such etiologies, some patients may not be able to differentiate between these sources of pain. Therefore, patients could be confused as to how PLD might address their specific

symptoms. Another major deficiency in this response is the lack of data regarding outcomes after surgery, which contributed to a reduction in the overall score of the response.

8. What is the success rate of lumbar decompression surgery?

The success rate of lumbar decompression surgery can vary based on several factors, including the specific condition being treated, the underlying cause of symptoms, the surgical technique used, the patient's overall health, and the criteria used to define success. Here is a general overview:

- (1) Herniated disc: Microdiscectomy or other decompression surgeries for herniated discs have generally high success rates, often above 80%-90%.
- (2) Spinal stenosis: Success rates vary, but many studies report favorable outcomes in terms of symptom relief.
- (3) Foraminal stenosis: Surgery for foraminal stenosis, which involves widening the nerve exit openings, can lead to relief of leg pain and other symptoms. Success rates vary but can be quite favorable.
- (4) Degenerative disc disease: Success rates are often good, but outcomes can vary based on factors like the extent of degeneration and overall health.
- (5) Spondylolisthesis: Success rates depend on factors like the grade of slippage and the success of fusion. Overall, many patients experience symptom improvement.
- (6) Recurrent symptoms: While many patients experience relief after lumbar decompression surgery, there is a small risk of recurrent symptoms due to factors such as scar tissue formation, incomplete decompression, or new issues.

Attending 1 Score: 2

Attending 2 Score: 2

Analysis: This response was generally of poor quality and of little use to patients. The reviewers both felt that the response did not delineate the differing success rates of PLD for any of the indications listed after "herniated disk". Secondly, the response repeats the same general answer for almost every indication - some of which, including spondylolisthesis, are typically treated by spinal fusion and not a standalone PLD. This could be seen as a point of frustration on the side of the patient as they begin to ask deeper questions from ChatGPT but are met with a plateauing effect of new information without specific resources to address lingering questions.

9. What are the different approaches for lumbar decompression surgery?

There are several surgical approaches for lumbar decompression, each designed to address specific conditions and anatomical considerations. The choice of approach depends on factors such as the underlying condition, the level of the spine affected, the surgeon's expertise, and patient characteristics. Here are some common approaches for lumbar decompression surgery:

- (1) Traditional open approach:
 - Involves a larger incision and more tissue dissection.
 - Provides direct visualization of the surgical area.
 - Allows for comprehensive decompression and treatment of complex conditions.
 - Examples include laminectomy, laminotomy, and discectomy.
- (2) Minimally invasive approach:
 - Utilizes smaller incisions and specialized instruments.

- Causes less tissue disruption and reduces postoperative pain.
- Often leads to faster recovery and shorter hospital stays.
- Different techniques include tubular retractor, endoscopic, and microendoscopic approaches.

Attending 1 Score: 3

Attending 2 Score: 2

Analysis: While the complete response offered a broad list of potential approaches to PLD, the descriptions for many approaches were incomplete and misleading. For example, the reviewers felt that in the description of a corpectomy, ChatGPT neglected to mention that after the removal of a vertebral body, some patients may require cage placement to restore height. Additionally, several of the approaches, such as facetectomy and corpectomy, would not be used for a simple PLD and would necessitate concomitant fusion. The claims regarding some of the advantages of MISS were felt to not be fully representative of the literature^[33-35]. As such, these responses were viewed as not wholly accurate and possibly confusing to patients considering their options before surgery.

10. Which approach is better for lumbar decompression?

The choice of approach for lumbar decompression surgery depends on various factors, including your specific condition, the underlying cause of your symptoms, the surgeon's expertise, and your individual needs. There are different surgical approaches that can be used for lumbar decompression, and what might be considered "better" depends on your unique situation.

- Condition: The type and severity of your spinal condition play a significant role in determining the appropriate approach.
- Health status: Your overall health, medical history, and any existing medical conditions might impact which approach is safer for you.
- Surgical expertise: The surgeon's experience and expertise with a particular approach are important factors.
- Patient preference: Discuss your goals, concerns, and expectations with your surgeon to choose an approach that aligns with your needs.
- Recovery goals: Consider your desired recovery timeline and how much downtime you can afford.

Attending 1 Score: 3

Attending 2 Score: 3

Analysis: This response provides a cursory overview of different approaches. However, it fails to provide evidence-based advantages or drawbacks for the different techniques. In addition, the responses are vague and do not sufficiently address the question or the nuances thereof. As such, it lacks both the specificity and evidence-based support to be helpful to patients considering surgery. Like the prior responses, there appears to be a plateauing of new information, even with different questions surrounding the same topic.

Average Attending Score: 2.95

DISCUSSION

The advent of online search engines has incited a change in how patients interact with medical information and, therefore, their physicians. New, open-access AI chatbot technology represents a revolutionary moment in medicine regarding how patients learn about their medical conditions. Given the potential

ramifications of such programs and tools, it is crucial to evaluate ChatGPT's utility and accuracy in disseminating orthopedic information. Chatbot responses may impact patients' perceptions of treatment options and risks prior to an evaluation by a physician. Several studies have analyzed the utility of ChatGPT for patients considering orthopedic surgery^[11-17]. Assessing ChatGPT's usefulness for preoperative patient education in spine surgery is especially critical due to the relatively high risk of spine surgery and the nuances that often guide decision making regarding the indications for different operations. To our knowledge, the present study is the first to use a modified validated scoring system to appraise and evaluate ChatGPT's responses to common patient questions when considering PLD surgery.

Minimum scores across all ten questions would lead to a total score of 20, whereas a maximum score would be 100. ChatGPT's responses in this analysis earned a score of 59, just under an average score of 3, when evaluated by two attending, fellowship-trained orthopedic spine surgeons. A score of 3 denoted a somewhat useful response of moderate quality, with some important information adequately discussed but some poorly discussed [Figure 1].

In the present study, ChatGPT was generally able to provide an accurate, albeit cursory, overview of relevant surgical indications, techniques, complications, and alternate therapies. However, some of these answers, when evaluated individually, lacked the clarification necessary to provide patients with a thorough understanding to inform their medical decision making. Some of the answers have the potential to be harmful to patients, especially those answers suggesting alternative therapy without the necessary context of the patient's particular history and symptom severity. In some instances, for example, PLD might be necessary to reverse or prevent further neurologic injury, especially for urgent and emergent indications. Suggesting alternative, non-operative treatment options for these patients could worsen or adversely impact patient outcomes. Concordantly, a prior study reported that ChatGPT had a 53% mismanagement rate, which would be especially deleterious for serious underlying pathology^[36]. Furthermore, non-operative treatment option descriptions were often vague, such as physical therapy to "strengthen muscles". This could lead some patients to pursue inadequate or harmful treatment, which may exacerbate or accelerate their disease processes.

Additionally, several of the claims were not fully substantiated by current spine surgery literature and several of the listed indications (spondylolisthesis and degenerative disc disease) may be better treated with other procedures, such as spinal fusion. As noted in previous literature, ChatGPT has been trained to generate definitive responses to questions, even when the existing literature may not be conclusive enough to make a specific recommendation^[37,38]. In particular, the chatbot seemed to indicate the superiority of MISS over the traditional open approach. While there is increasing research regarding the potential benefits of minimally invasive surgery, there are still gaps in the literature, which can be most appropriately addressed by a trained and experienced surgeon^[33,34]. These discrepancies may be confusing to patients considering PLD and could potentially lead to a delay in care. Nevertheless, ChatGPT did repeatedly emphasize that its responses should be taken in conjunction with consultation with a spine surgeon. This inability to address appropriate, patient-specific context affirms the findings of previous literature supporting the spine surgeon's role in providing individualized clinical recommendations^[36].

One limitation of any study attempting to characterize the utility of online sources of medical information to patients prior to a doctor's visit is the inherent subjectivity with which the online source is evaluated. To combat this weakness, the present analysis implemented a more objective, validated numeric scoring system. Additionally, the responses were analyzed by two attending spine surgeons, both of whose scores were presented, providing additional insight from physicians with differing levels of experience and areas of

Summary of Attending Scores to ChatGPT Responses

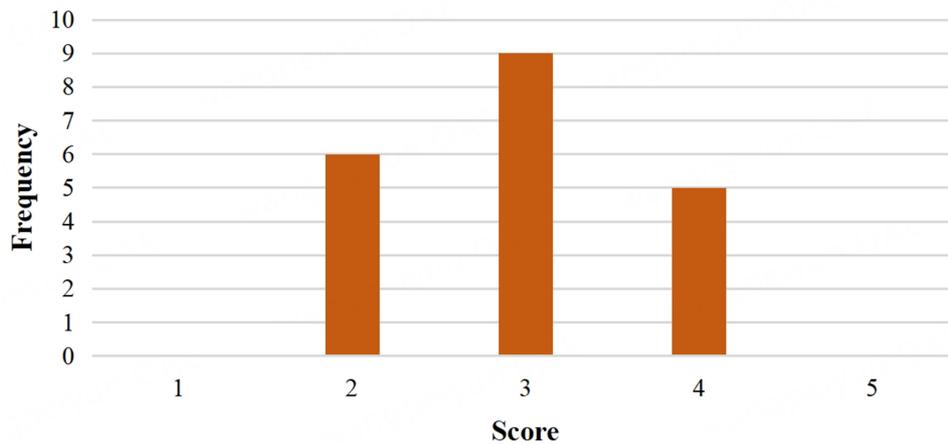


Figure 1. Summary of scores to ChatGPT responses.

focus. A current limitation of ChatGPT software is its inability to provide patient-specific recommendations based on individual clinical data. One way to improve ChatGPT's software, therefore, would be to include a method by which a patient could upload imaging in a HIPPA-compliant manner. ChatGPT could then use this diagnostic information, in conjunction with patient-reported symptoms, to make more specific recommendations. Finally, the present study attempts to analyze ChatGPT's responses to common patient questions, as it is the most widely utilized and easily accessible AI chatbot presently available for consumers. However, future studies could aim to compare the clinical accuracy and utility of responses given by different chatbots to systematically determine which might be best suited to providing patients with useful preoperative information.

In summation, ChatGPT has the ability to provide broadly useful responses to common preoperative questions that patients may have when considering undergoing PLD. ChatGPT has excellent utility in providing background information to patients and in helping them become more informed about their pathology in general. However, it often lacks the specific patient context, especially patient imaging data such as X-ray, CT, and MRI, necessary to provide patients with proper, accurate insights into their personal prognosis and medical options. Fortunately, ChatGPT does state its own limitations and consistently recommends using any information it provides in context with the consultation of a spine surgeon. In such instances, where patients use ChatGPT as background information to prepare for in-person appointments with a spine surgeon, this additional knowledge can be beneficial in enabling them to actively participate in the shared decision-making process regarding their medical care and take shared responsibility for their outcomes.

DECLARATIONS

Authors' contributions

Made substantial contributions to the conception and design of the study and performed data analysis and interpretation: Giakas AM, Narayanan R, Ezeonu T, Dalton J, Mangan J, Schroeder G, Kepler C

Performed data acquisition, as well as providing administrative, technical, and material support: Giakas AM, Narayanan R, Ezeonu T, Mangan J, Schroeder G, Kepler C

Conceptualization, methodology, project administration, writing - original draft: Lee Y
Conceptualization, methodology, project administration, writing - review and editing: Dalton J
Formal analysis, investigation, visualization, writing - original draft: Henry T
Conceptualization, project administration, supervision, writing - review and editing: Vaccaro A

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Financial support and sponsorship

None.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2024.

REFERENCES

1. Van Riel N, Auwerx K, Debbaut P, Van Hees S, Schoenmakers B. The effect of Dr Google on doctor-patient encounters in primary care: a quantitative, observational, cross-sectional study. *BJGP Open* 2017;1:bjgpopen17X100833. DOI PubMed PMC
2. Cocco AM, Zordan R, Taylor DM, et al. Dr Google in the ED: searching for online health information by adult emergency department patients. *Med J Aust* 2018;209:342-7. DOI PubMed
3. Fraval A, Ming Chong Y, Holcdorf D, Plunkett V, Tran P. Internet use by orthopaedic outpatients - current trends and practices. *Australas Med J* 2012;5:633-8. DOI PubMed PMC
4. Kasthuri V, Homer A, Alsoof D, et al. Modern internet search analytics and spine: what are patients asking and reading online? *N Am Spine Soc J* 2023;14:100214. DOI PubMed PMC
5. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell* 2023;6:1169595. DOI PubMed PMC
6. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare* 2023;11:887. DOI PubMed PMC
7. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *J Med Syst* 2023;47:33. DOI PubMed PMC
8. OpenAI. Introducing ChatGPT. 2022. Available from: <https://openai.com/blog/chatgpt>. [Last accessed on 27 Aug 2024].
9. Lock S. What is AI chatbot phenomenon ChatGPT and could it replace humans? 2022. Available from: <https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans>. [Last accessed on 27 Aug 2024].
10. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ* 2023;9:e45312. DOI PubMed PMC
11. Kaarre J, Feldt R, Keeling LE, et al. Exploring the potential of ChatGPT as a supplementary tool for providing orthopaedic information. *Knee Surg Sports Traumatol Arthrosc* 2023;31:5190-8. DOI PubMed PMC
12. Dubin JA, Bains SS, Chen Z, et al. Using a Google web search analysis to assess the utility of ChatGPT in total joint arthroplasty. *J Arthroplasty* 2023;38:1195-202. DOI PubMed
13. Duey AH, Nietsch KS, Zaidat B, et al. Thromboembolic prophylaxis in spine surgery: an analysis of ChatGPT recommendations. *Spine J* 2023;23:1684-91. DOI PubMed
14. Fayed AM, Mansur NSB, de Carvalho KA, Behrens A, D'Hooghe P, de Cesar Netto C. Artificial intelligence and ChatGPT in orthopaedics and sports medicine. *J Exp Orthop* 2023;10:74. DOI PubMed PMC
15. Hodakowski AJ, McCormick JR, Damodar D, et al. Rotator cuff repair: what questions are patients asking online and where are they getting their answers? *Clin Shoulder Elb* 2023;26:25-31. DOI PubMed PMC
16. Hurley ET, Crook BS, Lorentz SG, et al. Evaluation high-quality of information from ChatGPT (artificial intelligence-large language

- model) artificial intelligence on shoulder stabilization surgery. *Arthroscopy* 2024;40:726-31.e6. DOI PubMed
17. Mika AP, Martin JR, Engstrom SM, Polkowski GG, Wilson JM. Assessing ChatGPT responses to common patient questions regarding total hip arthroplasty. *J Bone Joint Surg Am* 2023;105:1519-26. DOI PubMed
 18. Subramanian T, Shahi P, Araghi K, et al. Using artificial intelligence to answer common patient-focused questions in minimally invasive spine surgery. *J Bone Joint Surg Am* 2023;105:1649-53. DOI PubMed
 19. Lattig F, Fekete TF, O'Riordan D, et al. A comparison of patient and surgeon preoperative expectations of spinal surgery. *Spine* 2013;38:1040-8. DOI PubMed
 20. Deyo RA, Mirza SK, Martin BI, Kreuter W, Goodman DC, Jarvik JG. Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA* 2010;303:1259-65. DOI PubMed PMC
 21. O'Lynnner TM, Zuckerman SL, Morone PJ, Dewan MC, Vasquez-Castellanos RA, Cheng JS. Trends for spine surgery for the elderly: implications for access to healthcare in North America. *Neurosurgery* 2015;77 Suppl 4:S136-41. DOI PubMed
 22. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. *Am J Gastroenterol* 2007;102:2070-7. DOI PubMed
 23. Kreiner DS, Hwang SW, Easa JE, et al; North American Spine Society. An evidence-based clinical guideline for the diagnosis and treatment of lumbar disc herniation with radiculopathy. *Spine J* 2014;14:180-91. DOI PubMed
 24. Zaina F, Tomkins-Lane C, Carragee E, Negrini S. Surgical versus non-surgical treatment for lumbar spinal stenosis. *Cochrane Database Syst Rev* 2016;2016:CD010264. DOI PubMed PMC
 25. Katz JN, Zimmerman ZE, Mass H, Makhni MC. Diagnosis and management of lumbar spinal stenosis: a review. *JAMA* 2022;327:1688-99. DOI PubMed
 26. Kuris EO, McDonald CL, Palumbo MA, Daniels AH. Evaluation and management of cauda equina syndrome. *Am J Med* 2021;134:1483-9. DOI PubMed
 27. Bulloch L, Thompson K, Spector L. Cauda equina syndrome. *Orthop Clin North Am* 2022;53:247-54. DOI PubMed
 28. Issack PS, Cunningham ME, Pumberger M, Hughes AP, Cammisa FP Jr. Degenerative lumbar spinal stenosis: evaluation and management. *J Am Acad Orthop Surg* 2012;20:527-35. DOI PubMed
 29. Fritz JM, Delitto A, Welch WC, Erhard RE. Lumbar spinal stenosis: a review of current concepts in evaluation, management, and outcome measurements. *Arch Phys Med Rehabil* 1998;79:700-8. DOI PubMed
 30. Goacher E, Sanders MI, Ivanov M. Safety and feasibility of same-day discharge following lumbar decompression surgery: a systematic review. *Brain Spine* 2022;2:100888. DOI PubMed PMC
 31. Degen T, Fischer K, Theiler R, et al. Outcomes after spinal stenosis surgery by type of surgery in adults aged 60 years and older. *Swiss Med Wkly* 2020;150:w20325. DOI PubMed
 32. Matsumoto K, Shah A, Kelkar A, et al. Biomechanical evaluation of a novel decompression surgery: transforaminal full-endoscopic lateral recess decompression (TE-LRD). *N Am Spine Soc J* 2021;5:100045. DOI PubMed PMC
 33. Hathi K, Bigney E, Richardson E, et al. Minimally invasive vs. open surgery for lumbar spinal stenosis in patients with diabetes - a Canadian spine outcomes and research network study. *Global Spine J* 2023;13:1602-11. DOI PubMed PMC
 34. Goldstein CL, Macwan K, Sundararajan K, Rampersaud YR. Perioperative outcomes and adverse events of minimally invasive versus open posterior lumbar fusion: meta-analysis and systematic review. *J Neurosurg Spine* 2016;24:416-27. DOI PubMed
 35. Skovrlj B, Belton P, Zarzour H, Qureshi SA. Perioperative outcomes in minimally invasive lumbar spine surgery: a systematic review. *World J Orthop* 2015;6:996-1005. DOI PubMed PMC
 36. Chalhoub R, Mouawad A, Aoun M, et al. Will ChatGPT be able to replace a spine surgeon in the clinical setting? *World Neurosurg* 2024;185:e648-52. DOI PubMed
 37. Mejia MR, Arroyave JS, Saturno M, et al. Use of ChatGPT for determining clinical and surgical treatment of lumbar disc herniation with radiculopathy: a North American spine society guideline comparison. *Neurospine* 2024;21:149-58. DOI PubMed PMC
 38. Rajjoub R, Arroyave JS, Zaidat B, et al. ChatGPT and its role in the decision-making for the diagnosis and treatment of lumbar spinal stenosis: a comparative analysis and narrative review. *Global Spine J* 2024;14:998-1017. DOI PubMed PMC

Review

Open Access



Machine learning applications in adult spinal deformity corrective surgery: a narrative review

Nader Toossi¹ , Ozhan Jerry²

¹Musculoskeletal Education and Research Center, Clinical Research, Audubon, PA 19403, USA.

²University of New England, College of Dental Medicine, Portland, ME 04103, USA.

Correspondence to: Dr. Nader Toossi, Musculoskeletal Education and Research Center, Clinical Research, 2560 General Armistead Avenue, Audubon, PA 19403, USA. E-mail: nt392@drexel.edu

How to cite this article: Toossi N, Jerry O. Machine learning applications in adult spinal deformity corrective surgery: a narrative review. *Art Int Surg* 2024;4:258-66. <https://dx.doi.org/10.20517/ais.2024.27>

Received: 6 May 2024 **First Decision:** 25 Jul 2024 **Revised:** 24 Aug 2024 **Accepted:** 9 Sep 2024 **Published:** 13 Sep 2024

Academic Editor: Andrew A. Gumbs **Copy Editor:** Dong-Li Li **Production Editor:** Dong-Li Li

Abstract

Adult spinal deformity (ASD) poses significant challenges in spinal surgery, requiring precise planning and execution for successful correction. Additionally, optimization of outcomes and reducing the high complication rates of ASD surgeries are additional challenges facing spinal deformity surgeons. The advent of machine learning (ML) has revolutionized various aspects of healthcare, including spinal surgery. This review provides a comprehensive overview of the current state of ML applications in spinal deformity corrective surgery, highlighting its potential benefits and challenges.

Keywords: Machine learning, adult spinal deformity, predictive modeling, artificial intelligence

INTRODUCTION

With the aging population, the incidence and prevalence of adult spinal deformity (ASD) are on the rise^[1], affecting millions worldwide and significantly impacting their quality of life. This often leads to the necessity of complex surgical interventions. Planning ASD surgery involves evaluating not only the entire spinal column but also the entirety of the skeleton to ensure appropriate radiographic alignment. ASD patients present with a variety of heterogeneous clinical manifestations, and there is a vast array of surgical methods available for their treatment, making the treatment algorithm quite complex. Additionally, ASD



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



surgery is associated with high complication rates in both the short and long term. These observations make ASD an ideal candidate for leveraging the significant potential offered by artificial intelligence and machine learning (ML).

Computational techniques have been used in the past several years to process large datasets and create complex mathematical models to determine the relationship between different variables affecting the outcomes of surgery. The idea behind ML, a subset of artificial intelligence, is to develop a system similar to the human brain to learn from clinical and radiographic data and apply the knowledge to new situations. In other words, ML employs computer algorithms to learn from data and past experiences, enabling the creation of intelligent models. These algorithms enable computers to identify patterns in datasets without relying on predefined rules, allowing them to learn relationships from the data and make predictions or decisions based on that knowledge. It has been shown that validated ML risk calculators can provide more accurate and objective prognoses to adjust patient expectations during patient care than expert surgeons' perception of risks in ASD surgery^[2]. The development of predictive models via ML algorithms for prognosticating patient outcomes following ASD surgery represents a significant advancement over traditional statistical models, which are more adept at identifying statistical associations between variables rather than providing predictive value^[3].

ML has shown promise in enhancing the accuracy and efficiency of various medical procedures, including spinal surgery. By taking advantage of large datasets and advanced algorithms, ML can assist surgeons in preoperative planning, intraoperative decision-making, and postoperative care, leading to improved patient outcomes.

The aim of this narrative review is to provide an overview of the current status of ML in enhancing spinal deformity correction surgery and its applicability in preplanning, intraoperative guidance, predictive modeling, and postoperative risk assessment.

METHODOLOGY

The authors conducted a non-systematic review of recent literature to support their perspectives on the applicability of ML in corrective spine surgery for adult ASD. This narrative review addresses three key stages in surgical practice [Figure 1] where ML can be impactful, and concludes by discussing the major challenges and future directions in the field.

Preoperative planning

Appropriate preoperative patient selection significantly impacts patient satisfaction, individualized decision-making by surgeons, and hospital resource utilization. Identifying patients with favorable outcomes preoperatively is a challenging task. Traditional statistical methods, such as multiple regression analyses, are better suited for hypothesis testing rather than predicting individual patient outcomes. In contrast, ML algorithms can readily identify patterns within large datasets without the need to test a specific hypothesis^[4]. However, this advantage of ML algorithms comes at the cost of interpretability. Predictive models generated by ML are more difficult to interpret than risk factors identified by traditional statistical tests^[3].

ASD patients exhibit significant heterogeneity in demographics, comorbidities, spinal pathologies, and genetic factors. Traditional outcome predictive models often overlook these individual variabilities, leading to suboptimal predictions. However, ML models excel in accounting for these differences. By analyzing comprehensive datasets that include detailed individual patient profiles, these models can generate personalized predictions, enhancing clinical decision-making and patient outcomes.

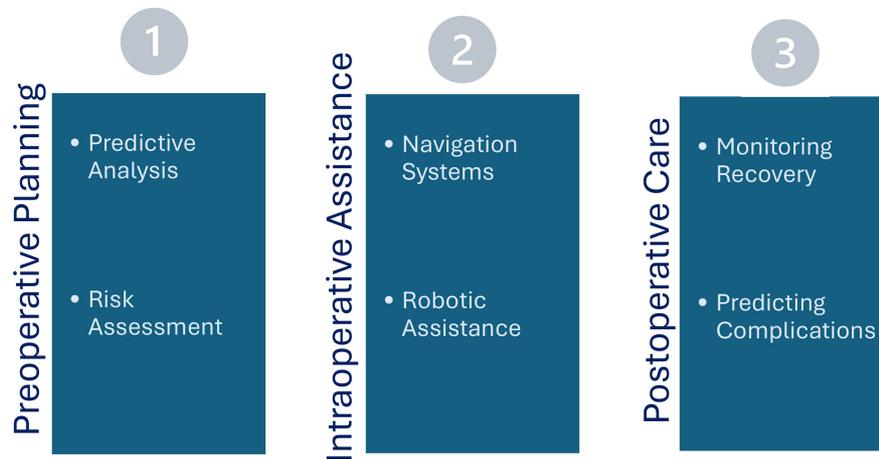


Figure 1. An overview of the most common applications of ML in ASD. ML: Machine learning; ASD: adult spinal deformity.

ML algorithms can analyze preoperative imaging studies, such as X-rays, CT scans, and MRI scans, to provide detailed insights into the patients' spinal alignments^[5-7]. This includes assessing the degree of deformity, identifying critical structures, and predicting the optimal surgical approach^[8]. ML models can also assist in selecting the appropriate implants (like pre-bent patient-specific rods) and predicting the postoperative spinal alignment, helping surgeons customize their surgical plan for each patient^[9,10]. Using ML algorithms, a group of investigators could accurately predict spinopelvic parameters and thoracic kyphosis after deformity correction surgery in 20 adult patients with spinal deformity^[11]. ML models can preoperatively be used to estimate the likelihood of extended length of stay following ASD surgery^[12,13]. Thus, the surgeon can optimize modifiable risk factors, enhance preoperative planning, and manage patients' expectations. Other investigators have developed predictive models to estimate the risk of rehabilitation discharge for adult patients undergoing elective surgeries, including ASD patients^[14].

Lafage *et al.* used artificial neural network based on preoperative data and alignment goals to accurately (81%) predict the upper instrumented vertebra (UIV) in a series of 143 ASD patients. This study showed how “to employ a neural network to mimic surgeon decision-making for UIV selection”^[15]. A neural network is a type of ML model inspired by the human brain's structure and functioning. It consists of interconnected nodes or neurons organized into layers: an input layer, one or more hidden layers, and an output layer. Each connection between neurons has a weight that adjusts during training to minimize error.

Additionally, prognosis can be predicted by using ML algorithms to identify different patient phenotypes preoperatively. In a recent prospective multi-center study on 570 ASD patients conducted by European and US-based Spine Study Group, investigators could identify three different qualitative preoperative phenotypes in ASD patients based on demographics, surgical history, frailty, radiographic measures, and patient-reported outcome measures. These phenotypes had been identified through unsupervised machine-based clustering. Based on these phenotypes, one can augment preoperative decision-making, predict the clinical outcome of deformity surgery (prognostic values), and tailor treatment approaches^[16].

An international team of researchers used a predictive ML model preoperatively to predict the individual answers to the Scoliosis Research Society-22R (SRS-22R) questionnaire at 1 and 2 years after ASD surgery in 561 patients. This prediction provides the patients with reasonable preoperative counseling based on their

expectations and perceptions of the corrective surgery clinical outcomes^[17]. Mekhael *et al.* used a random forest ML model to accurately predict health-related quality of life outcomes after ASD surgery in 173 patients. They found that three-dimensional movement assessment of ASD patients can better predict clinical outcomes than stand-alone radiographic parameters, not only for physical but also for mental scores^[18]. Random forest is an ensemble learning method that constructs multiple decision trees during training and merges their results to improve accuracy and control overfitting.

In another study, the researchers used a “ML model based on random forest regression and a systematic decision tree-like approach” to predict health-related quality of life scores, gait kinematics, and spatial-temporal parameters based on radiographic global alignment parameters^[19]. They found that Global Sagittal Angle^[20] and T9 tilt^[21] were the best predictors of joint kinematics and health-related quality of life scores based on the results from 127 primary ASD patients and 47 controls.

Conditional inference tree run ML analysis was used to identify the baseline threshold for different radiographic parameters to achieve a good outcome following ASD surgery in 431 patients. These parameters were: sagittal vertical axis, pelvic incidence-lumbar lordosis mismatch, pelvic tilt, T1 pelvic angle, L1 pelvic angle, L4-S1 lordosis, C2-C7 sagittal vertical axis, C2-T3, C2 slope^[22]. Conditional inference tree is a type of decision tree used in ML that addresses some of the biases found in traditional decision trees. It helps to make decisions by asking questions about data and splitting it accordingly, aiming to improve predictions and analyses.

Aiming to preoperatively predict proximal junctional kyphosis (PJK) after ASD corrective surgery in 191 patients, a team of researchers included preoperative thoracic T1 MRIs in a deep learning ML model (convolutional neural network) to increase the accuracy of the prediction^[23]. Using a large prospective multi-center database, a group of investigators constructed a supervised ensemble of decision trees to preoperatively predict the risk of pseudarthrosis at 2 years after ASD surgery in 336 patients with 91% accuracy^[24].

Intraoperative guidance

During surgery, ML algorithms can provide surgeons with real-time guidance, enhancing the accuracy of instrument placement and overall surgical technique. By integrating with navigation systems, ML can track the position of surgical instruments relative to the patient’s anatomy, ensuring precise correction of spinal deformity. Using ML methodology, Burström *et al.* were able to accurately place pedicle screws during CT-based navigation^[25]. Preplanning the pedicle screw trajectory using the ML system has yielded highly accurate results^[26]. ML can also be used to analyze intraoperative data, such as neuromonitoring signals, to alert surgeons of potential complications, such as nerve injury, enabling prompt intervention. Real-time automated decision-making systems regularly integrate inputs from intraoperative neuromonitoring and the operating room environment, utilizing predictive models to generate instructions or warnings for the surgical team. These systems continuously update their predictive models and decision-making processes based on new data and feedback from the surgeon and neurophysiologist, ensuring adaptive and accurate responses during surgery^[27].

Using perioperative data, ML-based risk calculators can predict the 30-day complication and mortality risk following ASD corrective surgery in 9,143 patients from The American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) database^[28]. Kim *et al.* could use ML algorithms to predict mortality and medical complications following ASD surgery. Using the data of 4,073 patients

queried from the ACS-NSQIP database, they found that ML algorithms outperformed American Society of Anesthesiologists score in predicting individual risk prognosis^[29].

Using conditional inference tree analysis, a team of investigators could predict blood loss and perioperative blood transfusion in 909 ASD patients undergoing surgery^[30]. The artificial neural network was used to predict perioperative blood transfusion after ASD corrective surgery in 1,173 cases identified from the NSQIP database between 2017 and 2019, with 81% accuracy^[31]. Furthermore, another group of researchers found no difference between random forest and tree-based ML models to predict blood transfusion following ASD corrective surgery in 1,029 patients^[32].

A team of researchers used ML-based predictive models to estimate the likelihood of overall improvement and surpassing the minimal clinically important difference (MCID) following adult spinal deformity (ASD) surgery, testing their models with eight patient-reported outcome measure instruments. The models could predict accurately and consistently whether a procedure would achieve MCID for a given patient using a given outcome instrument across a given time interval^[33,34].

Postoperative care

After ASD surgery, ML can aid in monitoring patients' recovery and predicting potential complications. By analyzing postoperative imaging^[35] and clinical data, ML models can identify early signs of implant failure, infection, or other complications, allowing for timely intervention^[36]. Since ASD surgery is fraught with complications postoperatively, many different characterizations have been developed to predict the complications after the surgery or determine risk profiles for the development of complications following deformity correction. The success of computer vision, large language models, and genome-wide association (incorporating advanced ML technologies) in predicting various complications in a cohort of ASD patients has been shown recently by a group of investigators^[37]. Major medical complications, discharge to a facility, and 90-day readmission were predicted using ML methods with decent accuracy^[38,39].

ML can also assist in predicting long-term complications, such as the risk of adjacent segment degeneration and PJK, and help surgeons and patients make informed decisions about follow-up care. Korean investigators recently developed and verified an online calculator for predicting PJK risk following ASD surgery using a ML model. They based their study on the radiographic outcomes obtained from 16 surgical centers^[40]. Moreover, to predict mechanical complications following ASD surgery, some investigators tried different ML models and found that random forest had the best prediction accuracy of 73.2%^[41]. Additionally, in a postoperatively well-aligned group of 244 patients following ASD surgery, some researchers could predict the mechanical complications with moderate accuracy (74%) using extreme gradient boosting ML algorithms. The mechanical complications investigated were: proximal junctional kyphosis and failure, distal junctional kyphosis and failure, rod breakage, and implant-related complications^[42]. Extreme gradient boosting is a technique that builds a strong predictive model by combining several weaker models, learning from mistakes, and doing so in a very efficient way.

Lovecchio *et al.* used decision tree analysis to predict the risk of proximal junctional failure and PJK by studying pre-discharge standing radiographs of 117 ASD patients^[43]. A group of Korean investigators could identify risk factors for unplanned readmission after ASD in 210 patients and predict it using a ML model^[44].

Some researchers used a conditionally unbiased regression tree and random forest algorithm to predict cost outliers in ASD correction up to 2 years after the index surgery in 210 patients^[45]. Conditionally unbiased

regression tree is a type of decision tree used for making predictions, particularly when numbers or continuous data are studied. It avoids bias by splitting data into different groups based on certain conditions.

Challenges and future directions

Despite its potential benefits, the integration of ML into spinal deformity correction surgery faces several challenges. These include the need for large, high-quality datasets, the interpretability of ML models, and the ethical and regulatory implications of ML algorithmic decision-making^[46,47]. Some investigators have suggested using biological samples (muscle and bone sampling, assessment of circulating biomarkers,...) to improve the accuracy of ML predictions in the future^[48,49]. Furthermore, most current studies employ a random split approach, in which the majority (70%-90%) of the available data are used for training the model, while the remaining 10%-30% for testing its performance. This approach is not generally deemed sufficient for the aim of “external” validation^[50]. Moreover, the extent to which ML-based predictions meaningfully affect clinical decisions and practices in real life has yet to be investigated. Future research should focus on addressing these challenges, as well as exploring new applications of ML, such as personalized surgical planning and robotic-assisted surgery, to further improve patient outcomes.

For spine surgeons embarking on ML collaborations, key considerations include:

1. Data quality and privacy: ensure high-quality, well-annotated data while adhering to patient privacy regulations like HIPAA.
2. Interdisciplinary communication: foster clear communication between clinicians and data scientists to bridge the gap between medical expertise and technical execution.
3. Clinical relevance: focus on models that address specific clinical challenges, ensuring they provide actionable insights in surgical planning, outcome prediction, or patient monitoring.
4. Validation and bias: rigorously validate ML models in diverse clinical settings to avoid biases and ensure generalizability.
5. Regulatory compliance: stay informed about the evolving regulatory landscape for AI in healthcare to ensure compliance with FDA or other relevant bodies.

These considerations are crucial for developing impactful and ethical ML solutions in spine surgery.

CONCLUSION

ML has the potential to revolutionize spinal deformity correction surgery by enhancing preoperative planning, intraoperative guidance, and postoperative care. By leveraging the power of large datasets and advanced algorithms, ML can assist surgeons in achieving more precise and personalized surgical outcomes, ultimately benefiting patients with spinal deformities.

DECLARATIONS

Authors' contributions

Made substantial contributions to the conception and design of the study and drafted the manuscript: Toossi N

Performed literature search, as well as providing administrative, technical, and material support: Jerry O

Availability of data and materials

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

Both authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2024.

REFERENCES

- Schwab F, Dubea A, Gamez L, et al. Adult scoliosis: prevalence, SF-36, and nutritional parameters in an elderly volunteer population. *Spine* 2005;30:1082-5. [DOI](#) [PubMed](#)
- Pellisé F, Vila-Casademunt A, Núñez-Pereira S, et al; European Spine Study Group, International Spine Study Group. Surgeons' risk perception in ASD surgery: the value of objective risk assessment on decision making and patient counselling. *Eur Spine J* 2022;31:1174-83. [DOI](#) [PubMed](#)
- Joshi RS, Lau D, Scheer JK, et al. State-of-the-art reviews predictive modeling in adult spinal deformity: applications of advanced analytics. *Spine Deform* 2021;9:1223-39. [DOI](#) [PubMed](#) [PMC](#)
- Akosman I, Lovecchio F, Lyons K, et al. The emerging role of artificial intelligence in adult spinal deformity. *Semin Spine Surg* 2022;34:100986. [DOI](#)
- Haselhuhn JJ, Soriano PBO, Grover P, et al. Spine surgeon versus AI algorithm full-length radiographic measurements: a validation study of complex adult spinal deformity patients. *Spine Deform* 2024;12:755-61. [DOI](#)
- Harake ES, Linzey JR, Jiang C, et al. Development and validation of an artificial intelligence model to accurately predict spinopelvic parameters. *J Neurosurg Spine* 2024;41:88-96. [DOI](#)
- Suri A, Tang S, Kargilis D, et al. Conquering the cobb angle: a deep learning algorithm for automated, hardware-invariant measurement of cobb angle on radiographs in patients with scoliosis. *Radiol Artif Intell* 2023;5:e220158. [DOI](#) [PubMed](#) [PMC](#)
- Galbusera F, Niemeyer F, Wilke HJ, et al. Fully automated radiological analysis of spinal disorders and deformities: a deep learning approach. *Eur Spine J* 2019;28:951-60. [DOI](#) [PubMed](#)
- Faulks CR, Biddau DT, Munday NR, McKenzie DP, Malham GM. Patient-specific spinal rods in adult spinal deformity surgery reduce proximal junctional failure: a review of patient outcomes and surgical technique in a prospective observational cohort. *J Spine Surg* 2023;9:409-21. [DOI](#) [PubMed](#) [PMC](#)
- Picton B, Stone LE, Liang J, et al. Patient-specific rods in adult spinal deformity: a systematic review. *Spine Deform* 2024;12:577-85. [DOI](#) [PubMed](#) [PMC](#)
- Lee NJ, Sardar ZM, Boddapati V, et al. Can machine learning accurately predict postoperative compensation for the uninstrumented thoracic spine and pelvis after fusion from the lower thoracic spine to the sacrum? *Global Spine J* 2022;12:559-66. [DOI](#) [PubMed](#) [PMC](#)
- Arora A, Demb J, Cummins DD, Callahan M, Clark AJ, Theologis AA. Predictive models to assess risk of extended length of stay in adults with spinal deformity and lumbar degenerative pathology: development and internal validation. *Spine J* 2023;23:457-66. [DOI](#) [PubMed](#)
- Li W, Zhang Y, Zhou X, et al. Ensemble learning-assisted prediction of prolonged hospital length of stay after spine correction surgery: a multi-center cohort study. *J Orthop Surg Res* 2024;19:112. [DOI](#) [PubMed](#) [PMC](#)
- Arora A, Lituiev D, Jain D, et al. Predictive models for length of stay and discharge disposition in elective spine surgery: development, validation, and comparison to the ACS NSQIP risk calculator. *Spine* 2023;48:E1-13. [DOI](#) [PubMed](#) [PMC](#)
- Lafage R, Ang B, Alshabab BS, et al. Predictive model for selection of upper treated vertebra using a machine learning approach. *World Neurosurg* 2021;146:e225-32. [DOI](#) [PubMed](#)
- 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. [DOI](#) [PubMed](#)
- Ames CP, Smith JS, Pellisé F, et al; European Spine Study Group; International Spine Study Group. Development of predictive models for all individual questions of SRS-22R after adult spinal deformity surgery: a step toward individualized medicine. *Eur Spine J* 2019;28:1998-2011. [DOI](#) [PubMed](#)

18. Mekhael E, El Rachkidi R, Saliby RM, et al. Functional assessment using 3D movement analysis can better predict health-related quality of life outcomes in patients with adult spinal deformity: a machine learning approach. *Front Surg* 2023;10:1166734. DOI PubMed PMC
19. Nassim N, Mekhael E, Rachkidi RE, et al. Global Sagittal Angle and T9-tilt seem to be the most clinically and functionally relevant global alignment parameters in patients with Adult Spinal Deformity. *Brain Spine* 2024;4:102805. DOI PubMed PMC
20. Diebo BG, Oren JH, Challier V, et al. Global sagittal axis: a step toward full-body assessment of sagittal plane deformity in the human body. *J Neurosurg Spine* 2016;25:494-9. DOI PubMed
21. Legaye J, Duval-Beaupere G. Sagittal plane alignment of the spine and gravity: a radiological and clinical evaluation. *Acta Orthop Belg* 2005;71:213-20. PubMed
22. Williamson TK, Krol O, Tretiakov P, et al. Crossing the bridge from degeneration to deformity: when does sagittal correction impact outcomes in adult spinal deformity surgery? *Spine* 2023;48:E25-32. DOI PubMed
23. Johnson GW, Chanbour H, Ali MA, et al. Artificial intelligence to preoperatively predict proximal junction kyphosis following adult spinal deformity surgery: soft tissue imaging may be necessary for accurate models. *Spine* 2023;48:1688-95. DOI PubMed PMC
24. Scheer JK, Oh T, Smith JS, et al; International Spine Study Group. Development of a validated computer-based preoperative predictive model for pseudarthrosis with 91% accuracy in 336 adult spinal deformity patients. *Neurosurg Focus* 2018;45:E11. DOI PubMed
25. Burström G, Buerger C, Hoppenbrouwers J, et al. Machine learning for automated 3-dimensional segmentation of the spine and suggested placement of pedicle screws based on intraoperative cone-beam computer tomography. *J Neurosurg Spine* 2019;31:147-54. DOI PubMed
26. Siemionow KB, Forsthoefel CW, Foy MP, Gawel D, Luciano CJ. Autonomous lumbar spine pedicle screw planning using machine learning: A validation study. *J Craniovertebr Junction Spine* 2021;12:223-7. DOI PubMed PMC
27. Wilson JP Jr, Kumbhare D, Kandregula S, Oderhowho A, Guthikonda B, Hoang S. Proposed applications of machine learning to intraoperative neuromonitoring during spine surgeries. *Neurosci Inf* 2023;3:100143. DOI
28. Pierce KE, Kapadia BH, Naessig S, et al. Validation of the ACS-NSQIP risk calculator: a machine-learning risk tool for predicting complications and mortality following adult spinal deformity corrective surgery. *Int J Spine Surg* 2021;15:1210-6. DOI PubMed PMC
29. Kim JS, Arvind V, Oermann EK, et al. Predicting surgical complications in patients undergoing elective adult spinal deformity procedures using machine learning. *Spine Deform* 2018;6:762-70. DOI PubMed
30. Raman T, Vasquez-Montes D, Varlotta C, Passias PG, Errico TJ. Decision tree-based modelling for identification of predictors of blood loss and transfusion requirement after adult spinal deformity surgery. *Int J Spine Surg* 2020;14:87-95. DOI PubMed PMC
31. De la Garza Ramos R, Hamad MK, Ryvlin J, et al. An artificial neural network model for the prediction of perioperative blood transfusion in adult spinal deformity surgery. *J Clin Med* 2022;11:4436. DOI PubMed PMC
32. Durand WM, DePasse JM, Daniels AH. Predictive modeling for blood transfusion after adult spinal deformity surgery: a tree-based machine learning approach. *Spine* 2018;43:1058-66. DOI PubMed
33. Ames CP, Smith JS, Pellisé F, et al; European Spine Study Group, International Spine Study Group. Development of deployable predictive models for minimal clinically important difference achievement across the commonly used health-related quality of life instruments in adult spinal deformity surgery. *Spine* 2019;44:1144-53. DOI PubMed
34. Patel AV, White CA, Schwartz JT, et al. Emerging technologies in the treatment of adult spinal deformity. *Neurospine* 2021;18:417-27. DOI PubMed PMC
35. Löchel J, Putzier M, Dreischarf M, et al. Deep learning algorithm for fully automated measurement of sagittal balance in adult spinal deformity. *Eur Spine J* 2024;Online ahead of print. DOI PubMed
36. Hopkins BS, Mazmudar A, Driscoll C, et al. Using artificial intelligence (AI) to predict postoperative surgical site infection: a retrospective cohort of 4046 posterior spinal fusions. *Clin Neurol Neurosurg* 2020;192:105718. DOI PubMed
37. Schonfeld E, Pant A, Shah A, et al. Evaluating computer vision, large language, and genome-wide association models in a limited sized patient cohort for pre-operative risk stratification in adult spinal deformity surgery. *J Clin Med* 2024;13:656. DOI PubMed PMC
38. Jain D, Durand W, Burch S, Daniels A, Berven S. Machine learning for predictive modeling of 90-day readmission, major medical complication, and discharge to a facility in patients undergoing long segment posterior lumbar spine fusion. *Spine* 2020;45:1151-60. DOI PubMed
39. Arora A, Demb J, Cummins DD, et al. Development and internal validation of predictive models to assess risk of post-acute care facility discharge in adults undergoing multi-level instrumented fusions for lumbar degenerative pathology and spinal deformity. *Spine Deform* 2023;11:163-73. DOI PubMed PMC
40. 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. DOI PubMed PMC
41. Noh SH, Lee HS, Park GE, et al. Predicting mechanical complications after adult spinal deformity operation using a machine learning based on modified global alignment and proportion scoring with body mass index and bone mineral density. *Neurospine* 2023;20:265-74. DOI PubMed PMC
42. Balaban B, Yilgor C, Yucekul A, et al. Building clinically actionable models for predicting mechanical complications in postoperatively well-aligned adult spinal deformity patients using XGBoost algorithm. *Inform Med Unlocked* 2023;37:101191. DOI
43. Lovecchio F, Lafage R, Sheikh Alshabab B, et al. Can discharge radiographs predict junctional complications? A decision tree

- analysis. *Global Spine J* 2024;14:970-7. DOI PubMed PMC
44. Ryu SJ, So JY, Ha Y, et al. Risk factors for unplanned reoperation after corrective surgery for adult spinal deformity. *Bone Joint Res* 2023;12:245-55. DOI PubMed PMC
 45. Ames CP, Smith JS, Gum JL, et al. Utilization of predictive modeling to determine episode of care costs and to accurately identify catastrophic cost nonwarranty outlier patients in adult spinal deformity surgery: a step toward bundled payments and risk sharing. *Spine* 2020;45:E252-65. DOI PubMed
 46. Wondra JP 2nd, Kelly MP, Greenberg J, et al. Validation of adult spinal deformity surgical outcome prediction tools in adult symptomatic lumbar scoliosis. *Spine* 2023;48:21-8. DOI PubMed PMC
 47. Yagi M, Yamanouchi K, Fujita N, Funao H, Ebata S. Revolutionizing spinal care: current applications and future directions of artificial intelligence and machine learning. *J Clin Med* 2023;12:4188. DOI PubMed PMC
 48. Haddad S, Pizones J, Raganato R, et al. Future data points to implement in adult spinal deformity assessment for artificial intelligence modeling prediction: the importance of the biological dimension. *Int J Spine Surg* 2023;17:S34-44. DOI PubMed PMC
 49. Choy W, Azad TD, Scheer JK, Safaee MM, Ames CP. Biomarkers in adult spinal deformity surgery. *Semin Spine Surg* 2023;35:101058. DOI
 50. Galbusera F, Cina A, Samartzis D. Commentary on “Predicting mechanical complications after adult spinal deformity operation using a machine learning based on modified global alignment and proportion scoring with body mass index and bone mineral density”. *Neurospine* 2023;20:275-7. DOI PubMed PMC

Original Article

Open Access



Patient perspectives on AI: a pilot study comparing large language model and physician-generated responses to routine cervical spine surgery questions

Ezra T. Yoseph¹, Anaysis D. Gonzalez-Suarez¹, Siegmund Lang^{1,2}, Atman Desai¹, Serena S. Hu³, Corinna C. Zygourakis¹

¹Department of Neurosurgery, Stanford University School of Medicine, Stanford, CA 94304, USA.

²Department of Trauma Surgery, University Hospital Regensburg, Regensburg 93053, Germany.

³Department of Orthopedic Surgery, Stanford University School of Medicine, Stanford, CA 94063, USA.

Correspondence to: Dr. Ezra T. Yoseph, Department of Neurosurgery, Stanford University School of Medicine, 300 Pasteur Dr, Palo Alto, Stanford, CA 94304, USA. E-mail: ezyoseph@stanford.edu

How to cite this article: Yoseph ET, Gonzalez-Suarez AD, Lang S, Desai A, Hu SS, Zygourakis CC. Patient perspectives on AI: a pilot study comparing large language model and physician-generated responses to routine cervical spine surgery questions. *Art Int Surg* 2024;4:267-77. <https://dx.doi.org/10.20517/ais.2024.38>

Received: 4 Jun 2024 **First Decision:** 2 Sep 2024 **Revised:** 11 Sep 2024 **Accepted:** 25 Sep 2024 **Published:** 29 Sep 2024

Academic Editor: Andrew A. Gumbs **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Aim: The purpose of this study was to elucidate differences in patient perspectives on large language model (LLM) vs. physician-generated responses to frequently asked questions about anterior cervical discectomy and fusion (ACDF) surgery.

Methods: This cross-sectional study had three phases: In phase 1, we generated 10 common questions about ACDF surgery using ChatGPT-3.5, ChatGPT-4.0, and Google search. Phase 2 involved obtaining answers to these questions from two spine surgeons, ChatGPT-3.5, and Gemini. In phase 3, we recruited 5 cervical spine surgery patients and 5 age-matched controls to assess the clarity and completeness of the responses.

Results: LLM-generated responses were significantly shorter, on average, than physician-generated responses (30.0 +/- 23.5 vs. 153.7 +/- 86.7 words, $P < 0.001$). Study participants were more likely to rate LLM-generated responses with more positive clarity ratings ($H = 6.25$, $P = 0.012$), with no significant difference in completeness ratings ($H = 0.695$, $P = 0.404$). On an individual question basis, there were no significant differences in ratings



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



given to LLM vs. physician-generated responses. Compared with age-matched controls, cervical spine surgery patients were more likely to rate physician-generated responses as higher in clarity ($H = 6.42$, $P = 0.011$) and completeness ($H = 7.65$, $P = 0.006$).

Conclusion: Despite a small sample size, our findings indicate that LLMs offer comparable, and occasionally preferred, information in terms of clarity and comprehensiveness of responses to common ACDF questions. It is particularly striking that ratings were similar, considering LLM-generated responses were, on average, 80% shorter than physician responses. Further studies are needed to determine how LLMs can be integrated into spine surgery education in the future.

Keywords: Anterior cervical discectomy and fusion (ACDF), large language model (LLM), ChatGPT, Gemini, patient education, health information quality, patient perspectives

INTRODUCTION

Anterior cervical discectomy and fusion (ACDF) is a common surgical intervention for the management of cervical spinal pathologies, including degenerative disc disease (central and paracentral disc herniations, and cervical stenosis), traumatic injuries, infection, and tumors^[1]. The procedure's technical aspects have undergone significant evolution, enhancing surgical outcomes and patient recovery trajectories^[2]. Despite the procedure's high prevalence, the complexity of ACDF, the heterogeneous pathologies for which it is performed, and the varying surgical techniques pose challenges for patients attempting to understand the surgery's risks, benefits, and postoperative recovery process^[3].

Studies have shown that a significant proportion of patients rely on online resources to gather information about surgeries, and that facilitating access to online health information can bolster patient compliance, postoperative plan adherence, and support the patient-physician relationship. However, this reliance on digital health resources can prove problematic, as outdated, contradictory, or highly technical information can complicate the patient's decision making^[4]. In this context, Langford *et al.* emphasized the importance of integrating high-quality online information into medical consultations, significantly impacting patient care and the dialogue between patients and physicians^[5]. Thus, patient-focused online educational resources must be precise, accessible, and importantly, accurate.

Studies have reported on the capability of large language models (LLMs), such as OpenAI's ChatGPT and Google's Gemini (formerly known as Bard), to parse through vast datasets and online surgical information to generate patient-specific responses that are coherent, comprehensive, and concise^[6-8]. Nonetheless, the accuracy, clarity, and completeness with which LLMs navigate complex medical domains, interpret clinical nuances, and subsequently deliver patient-friendly explanations warrants validation and continuous refinement. While LLMs have the potential to enhance patient comprehension of their medical conditions and treatment options, thereby increasing transparency and trust in surgical decision making, they may also carry the risk of disseminating inaccurate or biased information that could mislead patients and adversely affect their decision making and health outcomes^[9]. In this study, we evaluate the clarity and comprehensiveness of ChatGPT, Gemini, and two spine surgeons' responses to ten frequently asked patient questions by comparing how cervical spine surgery patients and their age-matched non-surgical patient counterparts rated these answers in terms of clarity and completeness.

METHODS

This cross-sectional study was approved by the Stanford Institutional Review Board (IRB-eProtocol #73097), and informed consent was obtained from all study participants.

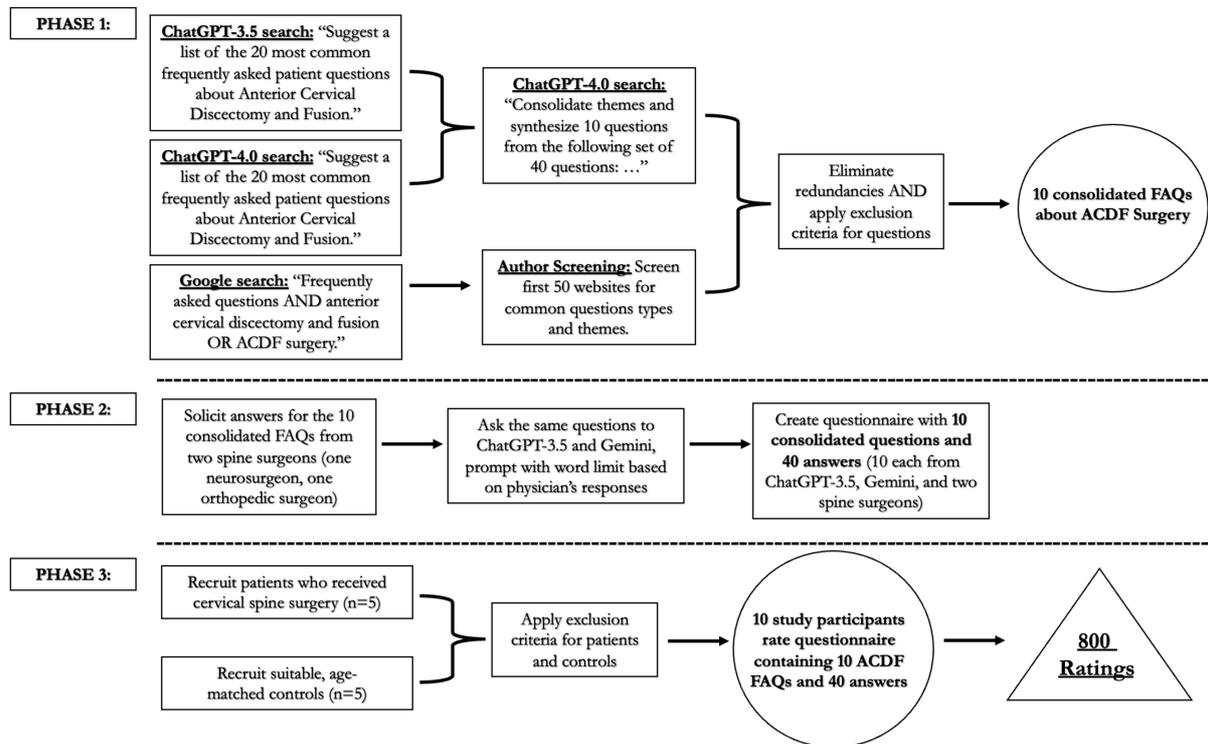


Figure 1. Methodology schematic detailing the three phases of the study design.

Study design and participants

Figure 1 shows the three phases of the study design. The goal of phase 1 was to craft ten commonly asked questions regarding ACDF surgery. To accomplish this goal, we utilized the following search engines: ChatGPT-3.5, ChatGPT-4.0, and Google. For both ChatGPT-3.5 and ChatGPT-4.0, the following prompt was submitted: “Suggest a list of the 20 most common frequently asked patient questions about anterior cervical discectomy and fusion”. The newly generated 40 questions were again submitted to ChatGPT-4.0 with the following prompt: “Consolidate themes and synthesize 10 questions from the following set of 40 questions: ...”. Concurrently, a Google search for “frequently asked questions AND anterior cervical discectomy and fusion OR ACDF surgery” was submitted, and the first 50 websites meeting our website inclusion criteria [Table 1] were surveyed. After eliminating redundancies and applying exclusion criteria, we were able to generate 10 frequently asked questions about ACDF surgery [Table 2]. All searches for phase 1 occurred on November 6, 2023.

Phase 2 involved soliciting responses to our 10 commonly asked ACDF surgery questions from spine surgeons and LLMs. In this effort, we conducted interviews with two attending spine surgeons, including one neurosurgeon and one orthopedic surgeon. Both surgeons were given clear instructions to answer the questions as if they were answering questions from a patient. Notably, both surgeons were blinded to all aspects of our study’s design, including the questions themselves, prior to the day of the interview. Following the interview, a transcript of their answers was produced. We next asked the same questions to ChatGPT-3.5 and Gemini with the following prompt: “Speak as an expert spine surgeon who is up to date with the latest scientific research and has years of experience counseling patients with empathy and clarity. Provide a comprehensive and easily understandable answer to the following question about cervical spine fusion surgery. Limit your answer to 250 words and focus on the most important aspects to ensure clarity: ...”. The word limit was determined based on the average length and range of our physician-generated

Table 1. Inclusion and exclusion criteria for websites containing frequently asked questions regarding ACDF surgery

Inclusion criteria	Exclusion criteria
Patient-focused, relevant questions and answers	Nongeneralizable physician anecdotes and physician-specific inquiries
Evidenced-based medical websites	Proprietary surgical techniques and devices which are not widely available
Information presented in the form of questions and answers	Research articles, non-patient-centered information

ACDF: Anterior cervical discectomy and fusion.

Table 2. Ten consolidated frequently asked questions administered to ChatGPT-3.5, Gemini, and doctors

Question 1	What is ACDF surgery, and how is it performed?
Question 2	How long is the typical recovery and fusion period, and when can I expect to return to work and daily activities?
Question 3	What materials are used for fusion in ACDF surgery, and what are the complications associated with both the procedure and the materials?
Question 4	Why is ACDF surgery recommended, and what are its potential risks and benefits?
Question 5	What are the long-term outcomes, success rates, and potential long-term effects or risks associated with ACDF surgery?
Question 6	What restrictions or precautions should I be aware of during my recovery, including wearing a neck brace and certain activities to avoid?
Question 7	What should I expect post-surgery in terms of incisions, scars, pain management, and potential discomfort?
Question 8	How long will I need to stay in the hospital post-surgery, and will I require physical therapy or rehabilitation?
Question 9	How long does ACDF surgery typically take?
Question 10	Are there any alternative treatments to ACDF, and under what circumstances might this surgery be repeated for other disc issues?

ACDF: Anterior cervical discectomy and fusion.

responses. At the end of phase 2, we had a total of 40 answers (10 each from two spine surgeons, ChatGPT-3.5, and Gemini) to our 10 commonly asked questions [Supplementary Table 1]. All searches for phase 2 occurred on November 23, 2023.

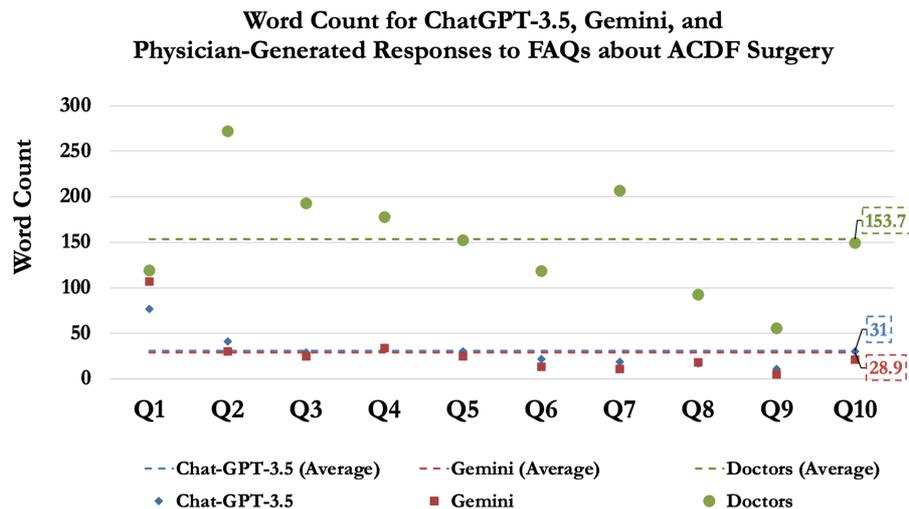
Phase 3 involved recruiting study participants according to the inclusion and exclusion criteria shown in Table 3. In total, there were 10 participants including 5 patients who had previously had cervical spine surgery and 5 gender- and age-matched controls. All study participants were given the same questionnaire with the 10 commonly asked ACDF surgery questions and 40 answers from phase 2. Participants were asked to use a 5-point Likert scale (Strongly Disagree, Disagree, Neutral, Agree, and Strongly Agree) to rate every response on both clarity and completeness. The exact prompts in the questionnaire were “This answer is clear and easy to understand” and “This answer completely answers the question”. All participants were blinded to LLMs or physicians being involved in the generation of responses, and the responses for every question appeared in a random order for each participant. Phase 3 produced a total of 800 data points for subsequent analysis.

Statistical methods

Data normality was assessed using the Shapiro-Wilk test. For non-normal data, the Kruskal-Wallis test and the Mann-Whitney U test were utilized, as appropriate, to assess differences in word counts by LLMs vs. physicians and overall ratings provided by patients and controls. Dunn-Bonferroni post-hoc test was conducted to pinpoint specific differences. Categories for the 100% stacked bar charts were set as follows: Likert ratings are converted to numbers (strongly disagree = 1, disagree = 2, neutral = 3, agree = 4, and strongly agree = 5), and then positive and negative feelings are combined to create categories (1 and 2 = disagree, 3 = neutral, 4 and 5 = agree). For our analyses, physician-generated response ratings were

Table 3. Inclusion and exclusion criteria used to select cervical spine surgery patients and age-matched controls

Inclusion criteria	Exclusion criteria
Speaks and reads English at native level proficiency	Participants who did not fill out the study questionnaire
Age > 18 years old	Participants who did not consent to participate in the study
Patients must have a history of cervical spine surgery at Stanford from 2019 to 2023	Age-matched controls cannot have a history of spine surgery

**Figure 2.** Line graph showing total and average word count for answers to the 10 frequently asked questions regarding ACDF surgery generated by ChatGPT-3.5, Gemini, and doctors. ACDF: Anterior cervical discectomy and fusion.

averaged. For a subset of the analysis, ChatGPT-3.5 and Gemini ratings were also averaged to assess patient perspectives on the two different LLMs. Independent two-sample t-test was used to compare ratings for each question. Inter-rater reliability was assessed using Fleiss' Kappa. Kappa values of > 0.80 indicate excellent reliability; 0.61 to 0.80, substantial reliability; 0.41 to 0.60, moderate reliability; 0.21 to 0.40, fair reliability; and ≤ 0.20 , poor reliability^[10]. The level of statistical significance was set at $P < 0.05$ or a specifically listed P -value when a conservative Bonferroni correction was applied in instances of analyses for multiple comparisons. All statistical analyses were executed using R Studio (version 4.1.2) or Python (version 3.8; Python Software Foundation).

RESULTS

The Shapiro-Wilk test indicated that the data were not normally distributed ($W = 0.825$, $P < 0.001$). This finding justified the use of non-parametric statistical methods for subsequent analyses.

Word count analysis

Compared to physician-generated responses, ChatGPT-3.5 and Gemini produced markedly shorter responses to every question (LLM avg = 30.0 +/- 23.5 vs. doctors avg = 153.7 +/- 86.7 words; $P < 0.01$; Figure 2). Despite being asked to limit responses to 250 words, the longest responses produced by ChatGPT-3.5 and Gemini were 77 and 107 words, respectively, while the average LLM responses were 31 and 28.9 words, respectively. Responses from physicians were significantly longer, with an average of 153.7 words per question [Table 4]. Overall, LLMs produced significantly shorter responses than physician-generated responses ($P < 0.001$). Comparisons of individual LLM platforms also revealed shorter responses produced by ChatGPT-3.5 vs. physicians ($P < 0.001$) and shorter responses produced by Gemini vs.

Table 4. Word counts for answers by ChatGPT-3.5, Gemini, and doctors

	ChatGPT-3.5	Gemini	LLM (average)	Doctors (average)
Question 1	77	107	92	119
Question 2	41	30	36	272
Question 3	29	25	27	193
Question 4	34	34	34	178
Question 5	30	25	28	153
Question 6	22	13	18	119
Question 7	19	11	15	207
Question 8	17	18	18	93
Question 9	11	5	8	56
Question 10	30	21	26	149
Average	31	28.9	30.0	153.7
STDEV	18.4	28.8	23.4	86.7

LLM: Large language model; STDEV: standard deviation.

physicians ($P < 0.001$), but no statistical difference in word count between ChatGPT-3.5 vs. Gemini ($P = 0.383$).

Aggregate ratings for LLM vs. physician-generated responses

Analysis of overall clarity ratings for LLM responses from study participants ($n = 10$) revealed that 75% agreed that responses were clear, while 6.5% disagreed and 18.5% were neutral. Clarity ratings for physician-generated responses showed that a statistically significantly lower 62.5% agreed that responses were clear, while 20% disagreed and 17.5% were neutral [Figure 3A]. Analysis of completeness ratings for Chatbot responses revealed that 63% agreed that responses were complete, while 18.5% disagreed and 18.5% were neutral. Completeness ratings for physician-generated responses showed that 64.5% agreed that responses were complete, while 12% disagreed and 23.5% were neutral [Figure 3B].

Overall, study participants were more likely to agree that responses generated by LLMs were clearer compared to responses generated by physicians ($H = 6.25$, $P = 0.012$). Despite the differences seen in the word count analysis, findings from study participants' ratings do not support differences in ratings for completeness between LLM vs. physician-generated responses ($H = 0.695$, $P = 0.404$). When comparing responses to each individual question, there were no significant differences between clarity or completeness ratings for LLM vs. physician-generated responses [Figure 4A and B].

Perspectives of cervical spine patients vs. controls

Ratings from cervical spine surgery patients ($n = 5$) were compared to those of gender- and age-matched controls ($n = 5$). There was an overall trend of patients being more likely to agree with statements about clarity and completeness compared to age-matched controls. Compared to controls, cervical spine surgery patients were more likely to give higher ratings for clarity ($H = 6.42$, $P = 0.011$) and completeness ($H = 7.65$, $P = 0.006$) for the physician-generated answers. Patients also showed a trend of rating LLM responses higher on clarity ($H = 3.04$, $P = 0.081$) and completeness ($H = 2.79$, $P = 0.09$) compared to the control group, but these differences did not reach statistical significance [Figure 5A and B]. Next, we separated the two LLM platforms to see if there were differences in patients vs. controls rating ChatGPT and Gemini responses. Compared to controls, spine surgery patients gave ChatGPT responses higher clarity ratings ($H = 9.06$, $P = 0.003$), with no significant differences in clarity ratings for Gemini responses ($H = 0.01$, $P = 0.930$). There

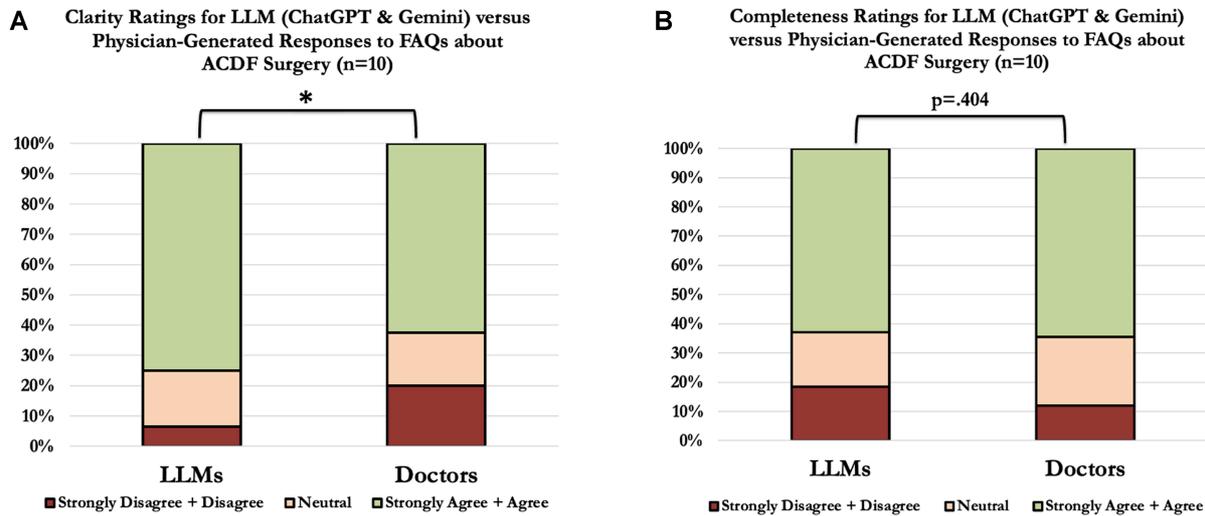


Figure 3. (A) Aggregate clarity and (B) completeness ratings, expressed in percentages, from all study participants ($n = 10$) comparing LLM vs. physician-generated responses. * $P < 0.05$. LLM: Large language model.

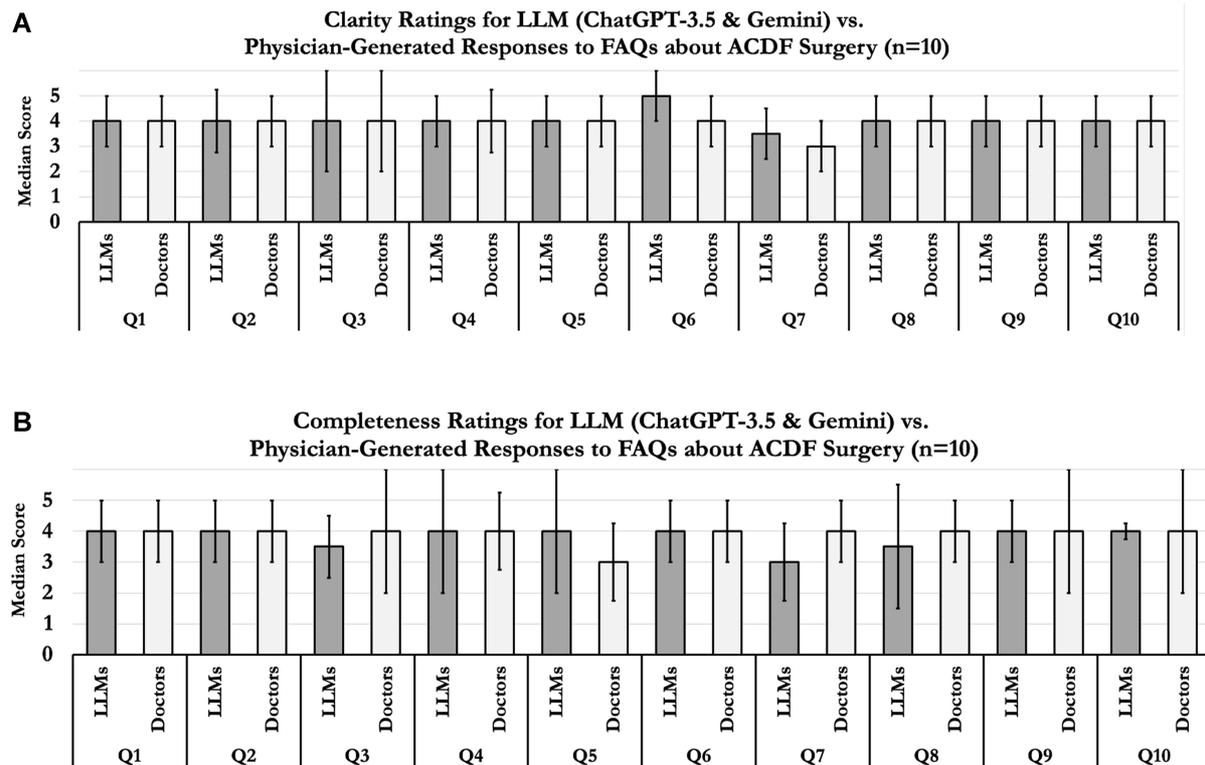


Figure 4. (A) Median clarity and (B) completeness ratings for individual questions from all study participants ($n = 10$) comparing LLM vs. physician-generated responses. Error bars represent IQR from the 25th through the 75th percentile. LLM: Large language model; IQR: interquartile range.

were no significant differences between patients and controls on completeness ratings for ChatGPT ($H = 5.36, P = 0.206$) or Gemini separately ($H = 1.61, P = 0.204$) [Supplementary Figure 1A and B].

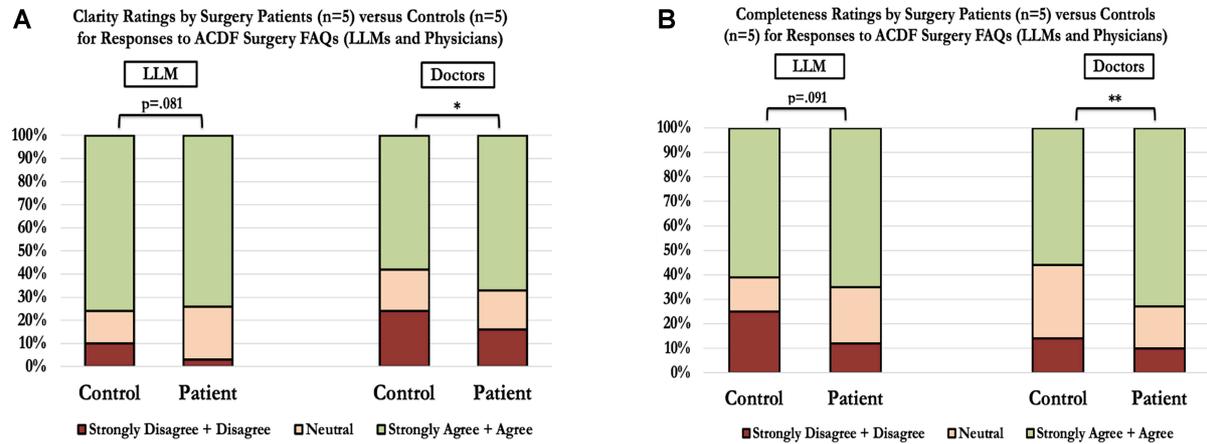


Figure 5. (A) Clarity and (B) completeness ratings, expressed in percentages, from cervical spine surgery patients ($n = 5$) vs. age-matched controls ($n = 5$) comparing answers generated by LLMs and physicians. * $P < 0.05$. ** $P < 0.01$. LLMs: Large language models.

Comparisons of individual questions revealed no statistically significant differences in clarity or completeness between patients and controls for responses by ChatGPT-3.5, Gemini, and physician-generated answers [Supplementary Figure 2A and B, Supplementary Table 2]. Overall, study participants exhibited poor to fair inter-rater reliability in ratings for LLM vs. physician-generated responses with regards to clarity (LLMs: $k = 0.16$, $P < 0.001$; Physicians: $k = 0.24$, $P < 0.001$) and completeness (LLMs: $k = 0.23$, $P < 0.001$; Physicians: $k = 0.12$, $P < 0.001$).

DISCUSSION

Several recent studies have demonstrated the potential of LLMs to deliver precise medical information and educate patients across various medical specialties^[11-13]. This study used patients who had undergone cervical spine surgery and gender- and age-matched controls to investigate perspectives on LLMs vs. physician-generated answers to commonly asked questions regarding ACDF surgery. We found that study participants were more likely to rate LLMs than physician-generated responses with positive ratings for clarity. Despite LLM responses being much shorter than physician-generated responses, they received equal ratings on completeness. This finding is exciting as it demonstrates that LLMs can provide short, concise responses to complex medical questions that are both clear and complete, appealing to patients and controls alike.

We also found that, when compared to age-matched controls, patients were more likely to rate physician-generated responses as clear and complete. This could potentially be explained by the patients having recently undergone spine surgery and spine surgery education (from the surgeon and surgical team), so that they are more familiar with medical terminology regarding ACDF surgery. This is further supported by patients also showing a trend of giving higher clarity and completeness ratings to LLM responses, potentially reflecting their familiarity with the subject matter. The familiarity with spine surgery likely introduces a bias for these patients, leading to a preference for responses that align with their prior knowledge. While this effect is evident in our study, it could potentially be generalized to other medical contexts, particularly where patients have prior experience or familiarity with a specific procedure. However, more research with larger sample sizes is needed to confirm this effect across different medical questions and procedures. For patients without prior surgery experience, LLMs could offer a more neutral perspective, potentially leveling the playing field between LLM and physician-generated responses. To better meet the needs of such patients, LLMs could be tailored with explanations that build foundational

understanding and make complex medical information more accessible to those with less familiarity. Practically, this can be accomplished with more specific LLM prompting based on one's prior understanding (or lack thereof) of the medical intervention. The lack of significant differences in individual question responses in our study is important because it validates that our findings are not skewed by any particular question, ultimately reinforcing the reliability of our findings.

As LLMs become more advanced, including faster and better at responding to complex medical questions clearly and completely, it may become prudent for physicians to employ LLMs as tools to improve practice efficiency and patient education. A recent study by Jahanshahi *et al.* assessed AI and machine learning techniques to process online messages between doctors and patients and to generate multiple automatic responses^[14]. Their machine learning model "BERT" was able to achieve an accuracy rate of 85.41% when suggesting the top 3 doctor responses. Worldwide, other studies have employed LLMs in telemedicine to reduce barriers to healthcare access and receive quick consultations in the setting of a pandemic^[15-17]. Collectively, these studies suggest that LLMs show great potential for quickly addressing medical questions from patients. Building upon this research, our study found that both spine patients and non-spine patient controls were satisfied with the clarity and completeness of LLM, as compared to physician-generated responses, and that LLMs outperformed physicians in some respects including brevity and clarity.

Our study is limited by its small sample size and poor to fair inter-rater reliability. The uniformly low to fair interrater reliability across all questions is likely due to differences in participants' background knowledge and potential ambiguities in our questions. Our initial intent was to capture the participant's gut reaction and initial response to the educational material, which is why we did not provide in-depth training. It is likely (and has been shown here) that these "gut reactions" or impulse responses are less reliable than ones that are given with systematic criteria. To improve reliability in future studies, we could provide rater training to ensure raters are aligned in their understanding of evaluation criteria. This study is also limited in that we used the free, more easily accessible ChatGPT-3.5, instead of paying for the newest version ChatGPT-4.0 which is - at the time of writing - OpenAI's most advanced system featuring the most safe and useful responses^[18]. It is important to consider the differences between these models since advancements in models' abilities can significantly enhance their performance. Specifically, ChatGPT-4.0 boasts significant improvements in understanding and generating human-like text, likely resulting in higher accuracy and a deeper comprehension of complex topics. If ChatGPT-4.0 had been used in our study, the responses might have been clearer and more closely aligned with expert-level answers, potentially influencing our assessment of AI's utility in this study^[19]. We expect that as the models continue to be refined, the capabilities of LLMs in this space will only improve.

We nevertheless feel that our study is significant in that it is the first of its kind to specifically evaluate LLM vs. physician-generated responses regarding ACDF surgery and the first to look for differences between patient and non-patient populations. Future studies examining patient perspectives on LLM vs. physician-generated responses should explore multiple other dimensions associated with patient satisfaction, including empathy and perceived trustworthiness of the response. Prior research has shown that physicians are more likely to rate LLM-produced responses as higher in empathy compared to physician-generated responses^[20]. Another study revealed that ChatGPT-4.0 shows the capacity for empathy when used to answer USMLE Step 2 Clinical Skills questions which are known to forecast performance in key residency domains, such as patient care, teamwork, professionalism, and communication^[21,22]. These studies both beg the question of whether the empathy, imparted by artificial intelligence, is felt by patients scouring through LLMs for answers to their healthcare queries. The impact of significantly shorter responses associated with LLMs vs. physicians is also an avenue worth exploring as a measure of patient satisfaction in future studies.

While our study provides valuable insights, it also raises several important research questions that warrant further exploration. Future studies could investigate how LLMs perform across various medical specialties and how they manage more complex or sensitive patient inquiries. Understanding the impact of LLM-generated responses on patient decision making is another exciting area for future research. There is also great potential in determining how a combined model that integrates LLMs with physician oversight changes the surgical decision-making process, potentially in terms of increasing or decreasing the number of patients who opt for surgical intervention, and/or improving their comfort level and understanding of the risks/benefits.

One final exciting application of AI in surgery is its significant role in enhancing the surgical consent process in spinal surgery. Recent studies have demonstrated that AI can effectively simplify complex medical information, improving readability and comprehension for patients. For example, Ali *et al.* showed that ChatGPT-4.0 could generate procedure-specific consent forms at an average 6th-grade reading level, significantly enhancing patient understanding without sacrificing important medical details^[23]. This AI-human expert collaborative approach not only improves patient education but also addresses medico-legal concerns by ensuring that consent forms meet both medical and legal standards. Given the litigious nature of spinal surgery, it is critical for future work to address the medico-legal implications of incorporating AI into this field. LLMs can still produce errors or “hallucinations”, making it essential to implement strict validation processes to ensure that only the most accurate information is conveyed to patients^[24]. This is particularly important as AI-generated content becomes more integrated into surgical decision making, where the stakes are highest.

Notably, ChatGPT provided the following statement after answering the last question: “Always consult with your spine surgeon for personalized advice and to address specific concerns regarding your ACDF surgery”. As LLMs evolve and enhance their precision, clarity, and comprehensiveness, it is important for physicians and medical researchers to evaluate and study the best way to incorporate these tools into the routine care of our patients moving forward.

DECLARATIONS

Authors' contributions

Made substantial contributions to the conception and design of the study: Yoseph ET, Lang S, Zygourakis CC

Performed data analysis and interpretation: Yoseph ET, Gonzalez-Suarez AD

Supported data acquisition: Yoseph ET, Zygourakis CC, Hu SS, Desai A

Wrote the manuscript: Yoseph ET, Gonzalez-Suarez AD, Zygourakis CC

Critically revised and approved final manuscript: Yoseph ET, Gonzalez-Suarez AD, Lang S, Desai A, Hu SS, Zygourakis CC

Availability of data and materials

Data supporting the findings of this study are available in [Supplementary Tables](#).

Financial support and sponsorship

None.

Conflicts of interest

Zygourakis CC is a consultant for Stryker and Amgen; Desai A is a consultant for Stryker and Carlsmed. While the other authors have declared that they have no conflicts of interest.

Ethical approval and consent to participate

Informed consent was gathered from each participant and this study was approved by the Stanford University Institutional Review Board (IRB-eProtocol #73097).

Consent for publication

Not Applicable.

Copyright

© The Author(s) 2024.

REFERENCES

1. Rhee JM, Ju KL. Anterior cervical discectomy and fusion. *JBJS Essent Surg Tech* 2016;6:e37. DOI PubMed PMC
2. Gould H, Sohal OA, Haines CM. Anterior cervical discectomy and fusion: techniques, complications, and future directives. *Semin Spine Surg* 2020;32:100772. DOI
3. Gaudin D, Krafcik BM, Mansour TR, Alnemari A. Considerations in spinal fusion surgery for chronic lumbar pain: psychosocial factors, rating scales, and perioperative patient education - a review of the literature. *World Neurosurg* 2017;98:21-7. DOI PubMed
4. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. *Health Educ Res* 2001;16:671-92. DOI PubMed
5. Langford AT, Roberts T, Gupta J, Orellana KT, Loeb S. Impact of the Internet on patient-physician communication. *Eur Urol Focus* 2020;6:440-4. DOI PubMed
6. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med* 2023;29:1930-40. DOI PubMed
7. Hung YC, Chaker SC, Sigel M, Saad M, Slater ED. Comparison of patient education materials generated by chat generative pre-trained transformer versus experts: an innovative way to increase readability of patient education materials. *Ann Plast Surg* 2023;91:409-12. DOI PubMed
8. Lang SP, Yoseph ET, Gonzalez-Suarez AD, et al. Analyzing large language models' responses to common lumbar spine fusion surgery questions: a comparison between ChatGPT and Bard. *Neurospine* 2024;21:633-41. DOI PubMed PMC
9. Blease C, Bernstein MH, Gaab J, et al. Computerization and the future of primary care: a survey of general practitioners in the UK. *PLoS One* 2018;13:e0207418. DOI PubMed PMC
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74. PubMed
11. Goodman RS, Patrinely JR, Stone CA Jr, et al. Accuracy and reliability of chatbot responses to physician questions. *JAMA Netw Open* 2023;6:e2336483. DOI PubMed PMC
12. Subramanian T, Shahi P, Araghi K, et al. Using artificial intelligence to answer common patient-focused questions in minimally invasive spine surgery. *J Bone Joint Surg Am* 2023;105:1649-53. DOI PubMed
13. Mika AP, Martin JR, Engstrom SM, Polkowski GG, Wilson JM. Assessing ChatGPT responses to common patient questions regarding total hip arthroplasty. *J Bone Joint Surg Am* 2023;105:1519-26. DOI PubMed
14. Jahanshahi H, Kazmi S, Cevik M. Auto response generation in online medical chat services. *J Healthc Inform Res* 2022;6:344-74. DOI PubMed PMC
15. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. *J Med Internet Res* 2023;25:e40789. DOI PubMed PMC
16. Bharti U, Bajaj D, Batra H, Lalit S, Lalit S, Gangwani A. Medbot: conversational artificial intelligence powered chatbot for delivering tele-health after COVID-19. In: 2020 5th International Conference on Communication and Electronics Systems (ICCES); 2020 Jun 10-12; Coimbatore, India. IEEE; 2020. pp. 870-5. DOI
17. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc* 2018;25:1248-58. DOI PubMed PMC
18. OpenAI. GPT-4. Available from: <https://openai.com/gpt-4>. [Last accessed on 27 Sep 2024].
19. OpenAI. GPT-4 Research. Available from: <https://openai.com/index/gpt-4-research/>. [Last accessed on 27 Sep 2024].
20. Brin D, Sorin V, Vaid A, et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. *Sci Rep* 2023;13:16492. DOI PubMed PMC
21. Sharma A, Schauer DP, Kelleher M, Kinnear B, Sall D, Warm E. USMLE step 2 CK: best predictor of multimodal performance in an internal medicine residency. *J Grad Med Educ* 2019;11:412-9. DOI PubMed PMC
22. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med* 2023;183:589-96. DOI PubMed PMC
23. Ali R, Connolly ID, Tang OY, et al. Bridging the literacy gap for surgical consents: an AI-human expert collaborative approach. *NPJ Digit Med* 2024;7:63. DOI PubMed PMC
24. Emsley R. ChatGPT: these are not hallucinations - they're fabrications and falsifications. *Schizophrenia* 2023;9:52. DOI PubMed PMC

Perspective

Open Access



Navigating artificial intelligence in spine surgery: implementation and optimization across the care continuum

Antony A. Fuleihan^{1,2} , Arjun K. Menta¹ , Tej D. Azad¹ , Kelly Jiang¹ , Carly Weber-Levine¹ ,
A. Daniel Davidar¹ , Andrew M. Hersh¹ , Nicholas Theodore¹ 

¹Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA.

²Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA 19107, USA.

Correspondence to: Dr. Nicholas Theodore, Department of Neurosurgery, Johns Hopkins University School of Medicine, 600 N. Wolfe St., Meyer 7-113, Baltimore, MD 21287, USA. E-mail: theodore@jhmi.edu

How to cite this article: Fuleihan AA, Menta AK, Azad TD, Jiang K, Weber-Levine C, Davidar AD, Hersh AM, Theodore N. Navigating artificial intelligence in spine surgery: implementation and optimization across the care continuum. *Art Int Surg* 2024;4:288-95. <https://dx.doi.org/10.20517/ais.2024.39>

Received: 8 Jun 2024 **First Decision:** 4 Sep 2024 **Revised:** 24 Sep 2024 **Accepted:** 8 Oct 2024 **Published:** 12 Oct 2024

Academic Editor: Andrew Gumbs **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

The field of spine surgery has long been characterized by innovations and technological advancements. The integration of artificial intelligence (AI) into spine surgery represents one of the latest technical developments in the field. The ability of AI to rapidly analyze datasets improves decision making, risk assessment, intraoperative precision, and postoperative management, all of which contribute to increasing personalized spine care and improving outcomes. However, the successful implementation of AI faces regulatory and privacy challenges that must be addressed before its full potential can be realized. Here, we provide a detailed analysis of the current applications and future prospects of AI in spine surgery, highlighting both the opportunities and challenges in this evolving field.

Keywords: Artificial intelligence, spine, machine learning, personalized medicine, education, imaging, patient safety, healthcare technology



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



INTRODUCTION

Artificial intelligence (AI) is increasingly being adapted for healthcare purposes, including analyzing complex data, identifying patterns, and making predictions and decisions [Supplementary Table 1]^[1,2]. A new frontier in AI has emerged with the genesis of generative AI models, which can create new content, including text and images, based on input data. These models could be leveraged to generate tailored surgical plans, create patient education packets, and assist in clinical documentation. Estimates project that AI solutions could potentially save up to \$360 billion dollars of US healthcare expenditure (5%-10%) annually^[3]. Here, we provide a comprehensive perspective on the existing applications as well as the frontiers and challenges for AI integration in spine surgery, including preoperative planning, intraoperative care, and postoperative management.

PREOPERATIVE PLANNING AND PATIENT SELECTION

Imaging analysis

AI algorithms excel at accurately analyzing spine imaging data, enabling the detection and characterization of pathologies with remarkable precision. For example, deep learning models have been developed for the automated detection of vertebral compression fractures on computed tomography (CT) or magnetic resonance imaging (MRI) scans^[4]. Al Arif *et al.* used a training set of 138 X-rays and a test set of 172 images to identify vertebral centers and outlines with an average error of only 1.81 and 1.69 mm, respectively^[5]. Doerr *et al.* used a region-based convolutional neural network to train and validate a deep learning model that can predict and classify a patient's thoracolumbar trauma based on CT imaging alone, reducing the need to pursue additional costly and time-consuming MRI imaging for assessment^[6]. AI can thereby help rapidly identify and triage patients in emergent settings and expedite the time to surgical intervention^[7].

Risk stratification and surgical planning

Machine learning algorithms can predict the risk of complications, such as surgical site infections, venous thromboembolism, and reoperation, during and after spinal procedures^[8-13]. Pellisé *et al.* utilized data from 1,612 patients across two independent prospective databases on adult spinal deformity to develop prognostic models for major complications, readmissions, and reoperations. The models can be used preoperatively to identify patients at greatest risk of postoperative complications and improve the patient counseling process^[14].

While AI is still in its early stages, it has shown significant potential when trained on robust and extensive retrospective data. For example, machine learning algorithms have shown that they can outperform surgeons' gestalt in predicting the risk of complications after emergency general surgery, including mortality, bleeding, and pneumonia^[15]. AI-driven predictive risk models can also incorporate biopsychosocial patient factors including demographics, comorbidities, frailty, laboratory values, and imaging data, as well as surgical details including approach, spinal levels, and instrumentation - all of which are critical cues in spine surgery^[16-18]. For instance, Goedmakers *et al.* developed a deep learning algorithm to predict adjacent segment disease following anterior cervical discectomy and fusion surgery, using only preoperative cervical MRI scans. The algorithm achieved a 95% accuracy rate, significantly outperforming expert neurosurgeons and neuroradiologists, who achieved only 58% accuracy^[19].

Patient engagement and education

Large language models can simplify the reading levels of consent forms from a collegiate level to a seventh-grade level, allowing for more patient accessibility and understanding^[20,21]. Moreover, AI-powered virtual assistants and chatbots can be harnessed as valuable care companions capable of providing knowledge to patients on demand^[22,23]. Boczar *et al.* created an AI-powered virtual assistant that correctly answered 92% of patient questions regarding plastic surgery in a sample of 30 participants and 294 questions^[24]. AI systems

can also provide personalized education materials, track patient-reported outcomes, and monitor adherence to preoperative instructions, without requiring additional appointments or an advanced understanding of medical terminology^[22,23]. However, caution must be taken to ensure that patients are not being misinformed by these systems.

Furthermore, AI systems, leveraging tools such as natural language processing, could also be integrated into clinic settings to automate patient intake^[25]. Such “virtual scribes” can automatically generate comprehensive clinical notes and summarize patient-provider interactions, reducing administrative burden. Such safeguards include regularly updating the AI’s knowledge base to ensure it reflects the latest clinical guidelines, implementing clear disclaimers that patients are interacting with a virtual assistant rather than a clinician, and ensuring the AI defers to human oversight in complex or ambiguous cases.

Expanding the diagnostic armamentarium

Beyond the analysis of static imaging data, AI holds significant potential to enhance spinal diagnostics by incorporating dynamic and longitudinal patient characteristics. Machine learning models and video capture tools have been used to identify abnormal gait and compensation patterns and estimate biomechanical variables, such as joint loading and range of motion, that are not easily discernible through traditional clinical evaluation^[26,27]. By integrating this dynamic data, clinicians can gain deeper insights into the underlying causes of spinal disorders and tailor interventions accordingly.

INTRAOPERATIVE SPINE CARE

Navigation and surgical accuracy

AI technologies can significantly enhance the spine operative experience. For example, surgical navigation systems powered by AI can seamlessly integrate with robotics to offer real-time guidance during complex spinal procedures^[28]. While the regulatory burden remains high, this is due to the need for rigorous validation to ensure patient safety and efficacy in high-stakes environments. The justification for adopting these technologies lies in their potential to greatly improve surgical precision, reduce complications, and enhance patient outcomes, which outweighs the hurdles posed by regulatory requirements. Future applications of AI can enhance existing navigation systems, allowing for minimized intraoperative errors and surgical risk via real-time corrections to unexpected changes^[29].

AI can also simulate procedures for educational training and create individualized models based on patient imaging. These models allow trainees to practice surgery on complex anatomical variations, serving as a powerful training tool^[30-33]. In addition to benefitting trainees and health professionals, these models help patients better understand their conditions, facilitating more informed discussions about treatment options and fostering engagement in the decision-making process.

Intraoperative documentation

In the operating room, AI techniques can streamline documentation. In plastic surgery, ChatGPT templates have been shown to generate operative notes over 42 times faster than traditional methods^[34]. Furthermore, with integration into billing and insurance information, AI can streamline the prior authorization process by extracting and organizing patient information, reducing delays in patient care. Zaidat *et al.* demonstrated a class-by-class accuracy between 77%-87% using a dataset of 902 operative notes of spine patients to generate billing codes^[35]. While AI has made tremendous progress in improving administrative efficiency, it still faces challenges with redundancy, inaccuracy, and hallucinations. Regular audits and human oversight are essential to prevent these errors.

POSTOPERATIVE MONITORING AND REHABILITATION

AI continues to play a vital role in the postoperative phase, facilitating efficient recovery and optimizing patient outcomes through solutions such as personalized rehabilitation plans^[36]. Lee *et al.* showed that an AI-based real-time motion feedback system improved strength and engagement during rehabilitation in spinal cord injury patients^[37]. Similarly, models have been applied to identify patients who may need prolonged postoperative opioid prescriptions. Karhade *et al.* trained numerous models on a database of 5,413 patients and accurately predicted sustained postoperative opioid dependence between 90 and 180 days^[38].

Leveraging longitudinal patient data, including clinical outcomes, activity levels, and patient-reported measures, AI can predict the trajectory of spinal conditions as well as the risk of complications or disease progression^[39,40]. These predictive models can help identify high-risk patients, optimize surgical indications, and guide proactive management strategies^[41].

FUTURE FRONTIERS IN SPINE CARE DATA OPTIMIZATION AND ANNOTATION

With patients generating gigabytes of data, the sheer volume presents challenges to clinicians. AI technologies can facilitate the interpretation of high-quality, structured data from diverse sources within the clinical environment, making them readily available for further analysis^[42]. AI tools are also particularly adept at extracting relevant data from large, unstructured datasets, a common challenge in medical settings. When discussing the performance of AI models, metrics such as precision, recall, and specificity are vital for evaluating their effectiveness in various tasks. These metrics help quantify how well an AI model identifies relevant data and minimizes errors.

Additionally, AI significantly enhances dataset annotation by automatically labeling imaging datasets, surgical videos, and other medical data with high accuracy^[43]. This capability accelerates the training process for retrospective analyses, thereby increasing research efficiency and identifying areas for improvement in the field.

Looking forward, the development of virtual scribes or “co-pilots” opens exciting possibilities. For patients, AI-powered co-pilots can serve as personalized guides through the care continuum, providing education and answering questions in real time. For surgeons, AI co-pilots can augment the surgical process by providing robust decision support, analyzing intraoperative metrics, and suggesting surgical approaches^[44]. We foresee a future where AI co-pilots integrate into the existing architecture of the spine surgery ecosystem [Figure 1]. In addition to offering real-time information to surgeons, co-pilots could assist with elevating critical non-technical roles, including improving communication, aiding with surgical team efficiency, and maintaining situational awareness^[45,46].

CHALLENGES

While the implementation of AI in spine surgery holds immense promise, several challenges must be addressed to fully realize its potential. A primary concern is the reliance on high-quality, standardized data. High-quality data are essential for training accurate AI models, and standardization ensures that these models can be applied broadly and effectively across different clinical settings^[2,47]. With the paramount importance of privacy and data security, compliance with regulations such as the Health Insurance Portability and Accountability Act (HIPAA) of 1996 is necessary to safeguard patient information.

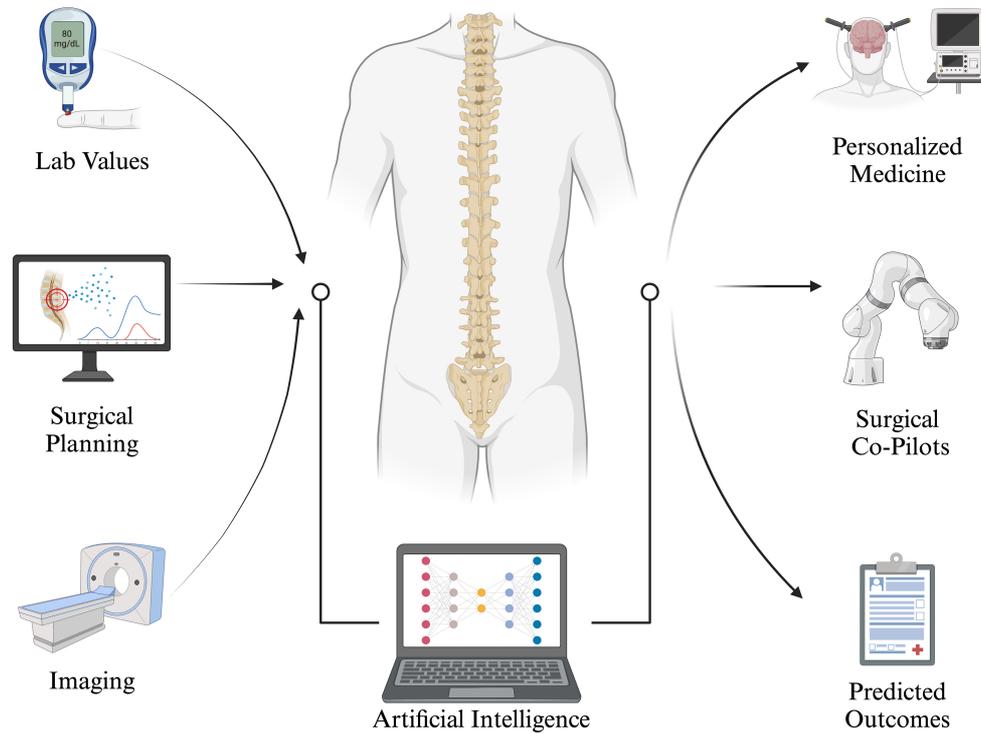


Figure 1. Data integration in AI-driven spine surgery. This figure illustrates the flow of multiple data collection points into advanced AI systems. The integrated AI processes these diverse inputs to generate valuable clinical outputs that inform patient care. Figure created with [BioRender.com](https://www.biorender.com). AI: Artificial intelligence.

Furthermore, maintaining trust in AI technology requires transparency and accountability. Skepticism of AI is often centered around the “black box” nature of its operations, where users cannot see how decisions are made. The decision-making process of AI systems should be clear and understandable to clinicians and patients alike. By familiarizing themselves with how these technologies function, they can better grasp the decision-making processes of AI systems. For example, machine learning allows computers to learn from data and improve over time, while deep learning, a subset of machine learning, mimics how the human brain processes information using layers of algorithms to analyze complex data.

A critical issue in the development of AI algorithms is addressing and mitigating biases that could lead to disparities in care. AI models should be trained on diverse and representative datasets to ensure they are applicable to a wide range of clinical and patient populations. To detect and address potential biases, it is essential to implement robust validation techniques, including bias audits and continuous monitoring during deployment. Engaging diverse stakeholders in the design and review processes can further help identify and rectify biases, ensuring fairness and equity in AI-driven treatments.

CONCLUSION

AI’s potential is vast and multifaceted, ranging from enhancing diagnostic accuracy to optimizing postoperative care. Its applications can lead to significant cost reductions, improved therapeutic outcomes, and enhanced quality of patient care. However, realizing this potential requires addressing challenges in data quality, standardization, and ethical implementation. By advancing and actively engaging in the ongoing discourse surrounding AI technologies, we can ensure that AI serves as a transformative force in spine surgery, ushering in a new era of personalized, precise, and proactive spine care.

DECLARATIONS

Authors' contributions

Conceptualization, design, synthesis, writing, and editing: Fuleihan AA, Menta AK, Azad TD, Theodore N
Writing and editing: Jiang K, Weber-Levine C, Davidar AD, Hersh AM

Availability of data and materials

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

Theodore N receives royalties from and owns stock in Globus Medical. He is a consultant for Globus Medical and has served on the scientific advisory board/other office for Globus Medical. While the other authors have declared that they have no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2024.

REFERENCES

1. Varghese C, Harrison EM, O'Grady G, Topol EJ. Artificial intelligence in surgery. *Nat Med* 2024;30:1257-68. DOI PubMed
2. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov* 2019;9:e1312. DOI PubMed PMC
3. Sahni NR, Stein G, Zimmel R, Cutler D. The potential impact of artificial intelligence on healthcare spending. In: The economics of artificial intelligence. University of Chicago Press; 2023. Available from: <https://www.degruyter.com/document/doi/10.7208/chicago/9780226833125-004/pdf?licenseType=restricted>. [Last accessed on 10 Oct 2024].
4. Burns JE, Yao J, Summers RM. Vertebral body compression fractures and bone density: automated detection and classification on CT images. *Radiology* 2017;284:788-97. DOI PubMed PMC
5. Al Arif SMMR, Knapp K, Slabaugh G. Fully automatic cervical vertebrae segmentation framework for X-ray images. *Comput Methods Programs Biomed* 2018;157:95-111. DOI PubMed
6. Doerr SA, Weber-Levine C, Hersh AM, et al. Automated prediction of the thoracolumbar injury classification and severity score from CT using a novel deep learning algorithm. *Neurosurg Focus* 2022;52:E5. DOI PubMed
7. Maier-Hein L, Vedula SS, Speidel S, et al. Surgical data science for next-generation interventions. *Nat Biomed Eng* 2017;1:691-6. DOI PubMed
8. Kim JS, Merrill RK, Arvind V, et al. Examining the ability of artificial neural networks machine learning models to accurately predict complications following posterior lumbar spine fusion. *Spine* 2018;43:853-60. DOI PubMed PMC
9. Karhade AV, Bongers MER, Groot OQ, et al. Development of machine learning and natural language processing algorithms for preoperative prediction and automated identification of intraoperative vascular injury in anterior lumbar spine surgery. *Spine J* 2021;21:1635-42. DOI PubMed
10. Wijnberge M, Geerts BF, Hol L, et al. Effect of a machine learning-derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the HYPE randomized clinical trial. *JAMA* 2020;323:1052-60. DOI PubMed PMC
11. Kalidasan V, Yang X, Xiong Z, et al. Wirelessly operated bioelectronic sutures for the monitoring of deep surgical wounds. *Nat Biomed Eng* 2021;5:1217-27. DOI PubMed
12. Bihorac A, Ozragat-Baslanti T, Ebadi A, et al. MySurgeryRisk: development and validation of a machine-learning risk algorithm for major complications and death after surgery. *Ann Surg* 2019;269:652-62. DOI PubMed PMC
13. Hu Z, Simon GJ, Arsoniadis EG, Wang Y, Kwaan MR, Melton GB. Automated detection of postoperative surgical site infections

- using supervised methods with electronic health record data. *Stud Health Technol Inform* 2015;216:706-10. [PubMed](#) [PMC](#)
14. Pellisé F, Serra-Burriel M, Smith JS, et al; International Spine Study Group, European Spine Study Group. Development and validation of risk stratification models for adult spinal deformity surgery. *J Neurosurg Spine* 2019;31:587-99. [DOI](#) [PubMed](#)
 15. El Moheb M, Gebran A, Maurer LR, et al. Artificial intelligence versus surgeon gestalt in predicting risk of emergency general surgery. *J Trauma Acute Care Surg* 2023;95:565-72. [DOI](#) [PubMed](#)
 16. Ames CP, Smith JS, Pellisé F, et al; European Spine Study Group, International Spine Study Group. Artificial intelligence based hierarchical clustering of patient types and intervention categories in adult spinal deformity surgery: towards a new classification scheme that predicts quality and value. *Spine* 2019;44:915-26. [DOI](#) [PubMed](#)
 17. Choy W, Azad TD, Scheer JK, Safaee MM, Ames CP. Biomarkers in adult spinal deformity surgery. *Semin Spine Surg* 2023;35:101058. [DOI](#)
 18. Topol EJ. As artificial intelligence goes multimodal, medical applications multiply. *Science* 2023;381:adk6139. [DOI](#) [PubMed](#)
 19. Goedmakers CMW, Lak AM, Ducey AH, et al. Deep learning for adjacent segment disease at preoperative MRI for cervical radiculopathy. *Radiology* 2021;301:664-71. [DOI](#) [PubMed](#)
 20. Mirza FN, Tang OY, Connolly ID, et al. Using ChatGPT to facilitate truly informed medical consent. *NEJM AI* 2024;1:Aics2300145. [DOI](#)
 21. Ali R, Connolly ID, Tang OY, et al. Bridging the literacy gap for surgical consents: an AI-human expert collaborative approach. *NPJ Digit Med* 2024;7:63. [DOI](#) [PubMed](#) [PMC](#)
 22. Solanki SL, Pandrowala S, Nayak A, Bhandare M, Ambulkar RP, Shrikhande SV. Artificial intelligence in perioperative management of major gastrointestinal surgeries. *World J Gastroenterol* 2021;27:2758-70. [DOI](#) [PubMed](#) [PMC](#)
 23. Dave M, Patel N. Artificial intelligence in healthcare and education. *Br Dent J* 2023;234:761-4. [DOI](#) [PubMed](#) [PMC](#)
 24. Boczar D, Sisti A, Oliver JD, et al. Artificial intelligent virtual assistant for plastic surgery patient's frequently asked questions: a pilot study. *Ann Plast Surg* 2020;84:e16-21. [DOI](#) [PubMed](#)
 25. Ni L, Lu C, Liu N, Liu J. MANDY: towards a smart primary care chatbot application. *Commun Comput Inform Sci* 2017;780:38-52. [DOI](#)
 26. Ino T, Samukawa M, Ishida T, et al. Validity of AI-based gait analysis for simultaneous measurement of bilateral lower limb kinematics using a single video camera. *Sensors* 2023;23:9799. [DOI](#) [PubMed](#) [PMC](#)
 27. Keller AV, Torres-Espin A, Peterson TA, et al. Unsupervised machine learning on motion capture data uncovers movement strategies in low back pain. *Front Bioeng Biotechnol* 2022;10:868684. [DOI](#) [PubMed](#) [PMC](#)
 28. Rasouli JJ, Shao J, Neifert S, et al. Artificial intelligence and robotics in spine surgery. *Global Spine J* 2021;11:556-64. [DOI](#) [PubMed](#) [PMC](#)
 29. Mascagni P, Alapatt D, Sestini L, et al. Computer vision in surgery: from potential to clinical value. *NPJ Digit Med* 2022;5:163. [DOI](#) [PubMed](#) [PMC](#)
 30. Goodman ED, Patel KK, Zhang Y, et al. Analyzing surgical technique in diverse open surgical videos with multitask machine learning. *JAMA Surg* 2024;159:185-92. [DOI](#) [PubMed](#) [PMC](#)
 31. Lavanchy JL, Zindel J, Kirtac K, et al. Automation of surgical skill assessment using a three-stage machine learning algorithm. *Sci Rep* 2021;11:5197. [DOI](#) [PubMed](#) [PMC](#)
 32. Levin M, McKechnie T, Khalid S, Grantcharov TP, Goldenberg M. Automated methods of technical skill assessment in surgery: a systematic review. *J Surg Educ* 2019;76:1629-39. [DOI](#) [PubMed](#)
 33. Lam K, Chen J, Wang Z, et al. Machine learning for technical skill assessment in surgery: a systematic review. *NPJ Digit Med* 2022;5:24. [DOI](#) [PubMed](#) [PMC](#)
 34. Abdelhady AM, Davis CR. Plastic surgery and artificial intelligence: how ChatGPT improved operation note accuracy, time, and education. *Mayo Clin Proc* 2023;1:299-308. [DOI](#)
 35. Zaidat B, Tang J, Arvind V, et al. Can a novel natural language processing model and artificial intelligence automatically generate billing codes from spine surgical operative notes? *Global Spine J* 2024;14:2022-30. [DOI](#) [PubMed](#) [PMC](#)
 36. Alsareii SA, Raza M, Alamri AM, et al. Machine learning and internet of things enabled monitoring of post-surgery patients: a pilot study. *Sensors* 2022;22:1420. [DOI](#) [PubMed](#) [PMC](#)
 37. Lee HJ, Jin SM, Kim SJ, et al. Development and validation of an artificial intelligence-based motion analysis system for upper extremity rehabilitation exercises in patients with spinal cord injury: a randomized controlled trial. *Healthcare* 2023;12:7. [DOI](#) [PubMed](#) [PMC](#)
 38. Karhade AV, Ogink PT, Thio QCBS, et al. Development of machine learning algorithms for prediction of prolonged opioid prescription after surgery for lumbar disc herniation. *Spine J* 2019;19:1764-71. [DOI](#) [PubMed](#)
 39. Wells CI, Xu W, Penfold JA, et al. Wearable devices to monitor recovery after abdominal surgery: scoping review. *BJS Open* 2022;6:zrac031. [DOI](#) [PubMed](#) [PMC](#)
 40. Dawes AJ, Lin AY, Varghese C, Russell MM, Lin AY. Mobile health technology for remote home monitoring after surgery: a meta-analysis. *Br J Surg* 2021;108:1304-14. [DOI](#) [PubMed](#)
 41. McGirt MJ, Sivaganesan A, Asher AL, Devin CJ. Prediction model for outcome after low-back surgery: individualized likelihood of complication, hospital readmission, return to work, and 12-month improvement in functional disability. *Neurosurg Focus* 2015;39:E13. [DOI](#) [PubMed](#)
 42. Vedula SS, Hager GD. Surgical data science: the new knowledge domain. *Innov Surg Sci* 2017;2:109-21. [DOI](#) [PubMed](#) [PMC](#)

43. Zuluaga L, Rich JM, Gupta R, et al. AI-powered real-time annotations during urologic surgery: the future of training and quality metrics. *Urol Oncol* 2024;42:57-66. [DOI](#) [PubMed](#)
44. 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. [DOI](#) [PubMed](#)
45. Gillespie BM, Harbeck E, Kang E, Steel C, Fairweather N, Chaboyer W. Correlates of non-technical skills in surgery: a prospective study. *BMJ Open* 2017;7:e014480. [DOI](#) [PubMed](#) [PMC](#)
46. Agha RA, Fowler AJ, Sevdalis N. The role of non-technical skills in surgery. *Ann Med Surg* 2015;4:422-7. [DOI](#) [PubMed](#) [PMC](#)
47. Li XT, Huang RY. Standardization of imaging methods for machine learning in neuro-oncology. *Neurooncol Adv* 2020;2:iv49-55. [DOI](#) [PubMed](#) [PMC](#)

Perspective

Open Access



Redefining precision: the current and future roles of artificial intelligence in spine surgery

Ryan W. Turlip, Harmon S. Khela, Mert Marcel Dagli, Daksh Chauhan, Yohannes Ghenbot, Hasan S. Ahmad, Jang W. Yoon

Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

Correspondence to: Dr. Jang W. Yoon, Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, 800 Spruce Street, Philadelphia, PA, 19107 USA. E-mail: jang.yoon@penmedicine.upenn.edu

How to cite this article: Turlip RW, Khela HS, Dagli MM, Chauhan D, Ghenbot Y, Ahmad HS, Yoon JW. Redefining precision: the current and future roles of artificial intelligence in spine surgery. *Art Int Surg* 2024;4:324-30. <https://dx.doi.org/10.20517/ais.2024.29>

Received: 17 May 2024 **First Decision:** 23 Sep 2024 **Revised:** 2 Oct 2024 **Accepted:** 15 Oct 2024 **Published:** 24 Oct 2024

Academic Editor: Rafael D. de la Garza-Ramos **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

The integration of artificial intelligence (AI) into spine surgery presents a transformative approach to preoperative and postoperative care paradigms. This paper explores the application of AI within spine surgery, focusing on diagnostic and predictive applications. AI-driven analysis of radiographic images, facilitated by machine learning (ML) algorithms such as convolutional neural networks (CNNs), potentially promises enhanced accuracy in identifying spinal pathologies and deformities; by combining these techniques with patient-specific data, predictive modeling can guide and inform diagnosis, prognosis, surgery selection, and treatment. Postoperatively, AI can leverage data from digital wearable technology to enhance the quantity and quality of outcome measures surgeons use to define and understand treatment success or failure. Still, challenges such as internal and external validation of AI models remain pertinent. Future directions include incorporating continuous variables from digital biomarkers and standardizing reporting metrics for AI studies in spine surgery. As AI continues to evolve, transparent validation frameworks and adherence to reporting guidelines will be crucial for its successful integration into clinical practice.

Keywords: Artificial intelligence, adult spinal deformity, radiographic imaging, machine learning, predictive modeling, objective outcomes



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



INTRODUCTION

The integration of artificial intelligence (AI) into spine surgery has given rise to significant improvements in patient safety, peri-operative decision making, and clinical outcomes^[1]. As new technological innovations herald faster, more efficient, and more accurate AI models, it is imperative for surgeons to understand the impact of AI on current treatment paradigms and where spine surgeons' focus should lie as we assist in the development of AI-enabled personalized and precision medicine.

At the cornerstone of clinical advancement with AI are machine learning (ML) models, capable of identifying and extracting patterns from large datasets and making predictions based on learned trends. As the availability of data grows, ML model performance continues to improve; therefore, the advancement of AI in medicine is uniquely tied to our ability to provide these models with accurate and pertinent datapoints. In this perspective, we provide a brief historical outline of current ML and AI applications in spine surgery. We then offer our thoughts on where the future of AI and spine surgery lies, and how the unique relationship between model accuracy and data volume will shape the future of how AI is implemented in clinical contexts.

CURRENT AI APPLICATIONS IN SPINE SURGERY

One of the earliest and most compelling uses of ML in spine surgery has been the use of models to automatically decipher radiographic images. For example, the classification of lumbar disc degeneration from 2-dimensional magnetic resonance image (MRI) using ML has now reached levels comparable to expert radiologists^[1-3]. The morphology of the discs is first described according to their pathological features and classified according to the standardized grading system proposed by Pfirrmann *et al.*^[4]. A convolutional neural network (CNN) is then used to extract image features from the training data set to make predictions based on the radiologists' interpretations. CNNs, a specialized subtype of deep learning (DL) algorithms, parallel the architecture of human visual cortex processing and rely on unsupervised pattern recognition to classify images. CNN-based models for image classification are typically validated through a combination of k-fold cross-validation on training data and then tested on independent and external datasets to ensure generalizability. Other groups have also explored the use of generative models to create image-to-image translations of the musculoskeletal system^[5,6]. Clinically, this can provide a means to correct poor image resolution or blurriness due to patient motion during image acquisition.

As DL algorithms became more prevalent, they have gradually been implemented to automatically determine spinal landmarks to calculate deformity parameters. DL models are trained on large datasets to identify and classify complex phenomena through non-linear analysis in artificial neurons, similar in structure to the mammalian brain^[7]. The automated analysis of the Cobb angle to describe the severity of scoliotic curvature has been addressed through several DL techniques^[8-10]. Korez *et al.* also used DL to identify anatomical landmarks in X-ray images and measure spinopelvic parameters, finding no difference between DL and manual identification^[11].

The transformative capability of AI can expedite diagnosis and treatment planning, and has the potential to standardize surgical treatment strategies for various spinal pathologies after taking patient-specific factors into account. Widespread implementation, however, faces substantial ethical challenges as the prospect of removing human interpretation may lead to more patient distrust in conclusions. It is unlikely, then, that human radiologists will be replaced by AI technology; instead, their diagnostic accuracy will be improved as models continue to advance.

The advent of AI-powered predictive modeling also holds immense promise in the realm of personalized precision medicine. By assimilating vast repositories of patient data, including demographic information, comorbidities, and procedural specifics, AI algorithms can generate prognostic models tailored to individual patients, ushering in a new era where therapeutic decisions are guided by each patient's unique physiology. This is particularly important for patient risk stratification, where clinical variables can be used as inputs (predictors) for the potential of operative complications. Pellisé *et al.* trained a random forest algorithm with clinical variables from 1,612 patients with adult spinal deformity (ASD) and identified age, surgical invasiveness, and deformity magnitude as potential risk factors for major complications^[12]. Predictive models, such as random forest algorithms for complication risk stratification, undergo internal validation through cross-validation and are, at times, externally validated using datasets from different clinical settings to evaluate model transferability. In the study by Pellisé *et al.*, internal validation was performed with an 80%/20% split between training/testing groups, measuring model performance through the observed area under the receiver operating characteristic curve (AUC) and the Brier score^[12]. Ames *et al.* augmented this approach by applying unsupervised hierarchical clustering to classify ASD based on patient demographics and radiographic measurements with the goal of constructing a risk-benefit grid as a preoperative tool for decision making^[13].

Current work continues to build upon existing outcomes prediction and postoperative prognostication. ML has been implemented to assess the likelihood of surgical site infection, major intra-operative complications, hospital length of stay, or the necessity of blood transfusion after surgery^[14-17]. This has led to the development of universal prediction models trained retrospectively on large patient registries, such as the American College of Surgeons National Surgical Quality Improvement Project (ACS-NSQIP) database. The ACS-NSQIP developed an online calculator for morbidity and mortality risk, but reports demonstrated poor predictive performance^[18]. Other groups have used the available ACS-NSQIP patient data as a resource to train their own models, with early indications of clinical efficacy at predicting outcomes^[19,20]. Fully unsupervised models have extensive utility to revolutionize personalized care and stratify risk; however, deploying under-validated AI tools can lead to inaccurate diagnoses or inappropriate treatment recommendations, so caution is needed to ensure patient safety.

Lastly, an emerging implementation of ML and AI has been in the realm of outcomes assessment. Traditionally, evaluation of surgical outcomes relies on physician interpretation of radiographic imaging combined with patient questionnaires assessing changes in patient mobility, pain, and quality of life. These patient-reported outcome measures (PROMs) offer valuable insight into patients' own interpretation of their health status and physical function. However, these methods contain inherent subjectivity and often lack the precision and reliability needed for precise and actionable insights^[21,22]. More recently, there has also been a trend toward utilizing digital biomarkers and data-driven outcomes measurements in conjunction with traditional PROMs. Objective measurements of patient mobility obtained from patient smartphones, smartwatches, or other biometric wearables can add additional unbiased insight into patient function^[23-26]. The quantitative and continuous features of these data are well suited for integration with data-driven statistical and ML techniques, and they have enabled surgeons to better quantify changes in patient mobility after surgery and to predict which patients may be better suited to recover from a particular pathology^[24,25].

FUTURE DIRECTIONS

The use of accelerometer and GPS information is a relatively novel concept, and more complex ML predictive models have yet to be applied. The incorporation of such models could significantly improve the accuracy of patient assessments by providing real-time, continuous data that captures a patient's functional mobility in their everyday life. This can lead to a more detailed understanding of a patient's functional

baseline status and postoperative recovery, resulting in tailored personalized medicine. While many analyses of mobility data have been retrospective in nature, upon the growth of adequate datasets, predictive models may be able to accurately identify subtle changes in mobility-related complications or improvements earlier than would be possible with traditional assessments.

Further, advanced mobility metrics can add potential value for patient prognostication. As previously mentioned, groups are beginning to engineer universal prognostic models for outcome prediction trained on large data registries^[19,20]. Although still in their infancy, accurate prognostic models could transform patient management by offering more realistic recovery trajectories, customizing patient care, or identifying high risk for adverse outcomes. There are still challenges that limit the widespread implementation of such models, ranging from access to generalizable datasets, cost-effectiveness for stable implementation, or ethical concerns.

Mobility metrics are not the only AI application that is challenged with limited data availability. Access to high-quality, standardized data sets is one of the greatest challenges to overall AI and ML implementation, especially within spine surgery, given the varied and nuanced model inputs spanning complex patient presentations, operative courses, and radiographic imaging. To address this challenge, there is a growing movement toward the creation of standardized, multi-center datasets that include patients from several geographic areas and socioeconomic groups. Other groups such as the ACS are refining their existing patient registries to integrate additional data from the electronic health record. Together, these datasets and registries aim to provide a foundation for training more accurate and generalizable AI models that can be deployed across various clinical settings.

Patient selection is another area of current clinical practice that stands to benefit from future AI and ML integration. The art of understanding which patients will benefit from certain procedures is not easily replicated with frameworks and rules that can be directly input into computerized programs. However, as CNNs and ML algorithms continue to grow in computational ability, they can potentially identify relationships between datapoints that are otherwise unnoticeable to the un-aided human mind; in this way, future AI and ML models can augment surgeons' clinical practice and assist in identifying certain patient characteristics that are indicative of patients likely to benefit from specific surgical interventions.

While AI technologies like predictive modeling and image analysis hold promise in decision making, their potential intra-operative impact is already apparent^[1,7]. AI-assisted intra-operative tools, such as robotics, navigation systems, and mixed reality, have the potential to significantly enhance the surgeon's ability to execute procedures with high precision, particularly in minimally invasive and percutaneous surgeries. These technologies allow for real-time guidance and adjustment during complex procedures, reducing the margin of error. However, while AI can minimize the risk of intra-operative errors, it cannot fully replace the human element of adaptability and judgment. Surgeons must remain vigilant in managing unforeseen intra-operative variables and complications, as AI systems, though highly advanced, still require human oversight to ensure patient safety and the proper handling of unexpected challenges.

Although surgeon experience is regarded as a significant factor in decision making, there have been attempts to apply mathematical and data-driven approaches to surgical decision making^[27]. Lewandrowski *et al.* recently used the Rasch model to determine the choice of procedure for endoscopic lumbar decompression^[27]. The Rasch model is a logistic function analyzing categorical data, such as questionnaire responses, to find the relative difficulty of a task, and it has been widely established in education, marketing, and health economics^[28]. However, it was found that there was still disagreement among surgeons regarding

the ability to achieve adequate clinical outcomes, indicating that increased granularity through additional metrics is needed to overcome the disordered responses^[27].

Despite the promising advancements of AI in spine surgery, a significant limitation in the current literature is the lack of external validation of many studies. Most models are only internally validated on the same data from which they were derived, raising concerns about model generalizability to larger patient populations or different clinical settings. It was estimated that only 5% of published articles on prognostic models included an external validation framework^[29]. Without external validation, it is difficult to ensure that these AI models will perform reliably in diverse environments, further limiting their clinical application. This issue is compounded by the scarcity of randomized controlled trials (RCTs) investigating AI in spine surgery, which are essential for evaluating long-term effectiveness and accuracy.

Due to the lack of standardized reporting metrics for AI studies, it is imperative to create clear guidelines through which the risk of bias and the potential utility of these models can be evaluated. AI studies that focus primarily on diagnostic applications using medical imaging should adhere to the Checklist for Artificial Intelligence in Medical Imaging (CLAIM)^[30]. The forthcoming Standards for Reporting of Diagnostic Accuracy Studies for AI (STARD-AI), an AI-specific adaptation of the established STARD guidelines, is also under development. Upon its release, it is expected to be indexed on the Enhancing the QUALity and Transparency Of health Research (EQUATOR) Network, addressing similar methodological issues as those covered by CLAIM^[31].

For ML multivariable prediction models, whether diagnostic or prognostic, the recently published Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis + Artificial Intelligence (TRIPOD + AI) provides a structured protocol for reporting predictive algorithms^[32]. Despite the advancements since the initial 2015 TRIPOD statement, which has shown promise in improving methodological transparency^[32,33], substantial gaps persist that hinder the broader integration of AI in clinical practice^[34]. As AI prediction algorithms become more pervasive in spine surgery, internal and external validation frameworks are necessary to appraise model performance, ensuring the variability in different patient populations is reflected to enhance surgical precision.

CONCLUSION

The integration of AI and ML into spine surgery represents a transformative shift toward precision medicine, offering enhanced diagnostic and prognostic capabilities. With the advances in automated radiographic imaging, patient risk stratification, outcomes prediction, and personalized medicine, future work promises to tailor treatment to individual patients more accurately. Despite the promising achievements so far, the field must address challenges in data accuracy by expanding training datasets and implementing robust validation frameworks. As AI becomes more prevalent in spine surgery, successful integration has the power to refine surgical decision making and improve patient outcomes.

DECLARATIONS

Authors' contributions

Original draft preparation, methodology, conceptualization: Turlip RW

Original draft presentation, conceptualization: Khela HS

Review and editing, supervision: Dagli MM, Ghenbot Y, Ahmad HS

Review and editing, validation: Chauhan D

Review and editing, supervision, conceptualization: Yoon JW

Critical writing: Turlip RW, Khela HS, Dagli MM, Chauhan D, Ghenbot Y, Ahmad HS, Yoon JW

Availability of data and materials

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2024.

REFERENCES

1. Galbusera F, Casaroli G, Bassani T. Artificial intelligence and machine learning in spine research. *JOR Spine* 2019;2:e1044. DOI PubMed PMC
2. Ruiz-España S, Arana E, Moratal D. Semiautomatic computer-aided classification of degenerative lumbar spine disease in magnetic resonance imaging. *Comput Biol Med* 2015;62:196-205. DOI PubMed
3. Jamaludin A, Lootus M, Kadir T, et al; Genodisc Consortium. ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. *Eur Spine J* 2017;26:1374-83. DOI PubMed
4. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine* 2001;26:1873-8. DOI PubMed
5. Johnson PM, Drangova M. Conditional generative adversarial network for 3D rigid-body motion correction in MRI. *Magn Reson Med* 2019;82:901-10. DOI PubMed
6. Galbusera F, Bassani T, Casaroli G, et al. Generative models: an upcoming innovation in musculoskeletal radiology? A preliminary test in spine imaging. *Eur Radiol Exp* 2018;2:29. DOI PubMed PMC
7. Charles YP, Lamas V, Ntilikina Y. Artificial intelligence and treatment algorithms in spine surgery. *Orthop Traumatol Surg Res* 2023;109:103456. DOI PubMed
8. Sun H, Zhen X, Bailey C, Rasoulinejad P, Yin Y, Li S. Direct estimation of spinal Cobb angles by structured multi-output regression. In: Niethammer M, Styner M, Aylward S, Zhu H, Oguz I, Yap P, Shen D, editors. *Information processing in medical imaging*. Cham: Springer International Publishing; 2017. pp. 529-40. DOI
9. Zhang J, Li H, Lv L, Zhang Y. Computer-aided Cobb measurement based on automatic detection of vertebral slopes using deep neural network. *Int J Biomed Imaging* 2017;2017:9083916. DOI PubMed PMC
10. Wu H, Bailey C, Rasoulinejad P, Li S. Automated comprehensive adolescent idiopathic scoliosis assessment using MVC-Net. *Med Image Anal* 2018;48:1-11. DOI PubMed
11. Korez R, Putzier M, Vrtovec T. A deep learning tool for fully automated measurements of sagittal spinopelvic balance from X-ray images: performance evaluation. *Eur Spine J* 2020;29:2295-305. DOI PubMed
12. Pellisé F, Serra-Burriel M, Smith JS, et al; International Spine Study Group, European Spine Study Group. Development and validation of risk stratification models for adult spinal deformity surgery. *J Neurosurg Spine* 2019;31:587-99. DOI PubMed
13. Ames CP, Smith JS, Pellisé F, et al; European Spine Study Group, International Spine Study Group. Artificial intelligence based hierarchical clustering of patient types and intervention categories in adult spinal deformity surgery: towards a new classification scheme that predicts quality and value. *Spine* 2019;44:915-26. DOI
14. Lee MJ, Cizik AM, Hamilton D, Chapman JR. Predicting surgical site infection after spine surgery: a validated model using a prospective surgical registry. *Spine J* 2014;14:2112-7. DOI PubMed
15. Scheer JK, Smith JS, Schwab F, et al; International Spine Study Group. Development of a preoperative predictive model for major complications following adult spinal deformity surgery. *J Neurosurg Spine* 2017;26:736-43. DOI PubMed

16. Durand WM, DePasse JM, Daniels AH. Predictive modeling for blood transfusion after adult spinal deformity surgery: a tree-based machine learning approach. *Spine* 2018;43:1058-66. [DOI PubMed](#)
17. Wang SK, Wang P, Li ZE, et al. Development and external validation of a predictive model for prolonged length of hospital stay in elderly patients undergoing lumbar fusion surgery: comparison of three predictive models. *Eur Spine J* 2024;33:1044-54. [DOI PubMed](#)
18. Sebastian A, Goyal A, Alvi MA, et al. Assessing the performance of national surgical quality improvement program surgical risk calculator in elective spine surgery: insights from patients undergoing single-level posterior lumbar fusion. *World Neurosurg* 2019;126:e323-9. [DOI PubMed](#)
19. Goyal A, Ngufor C, Kerezoudis P, McCutcheon B, Storlie C, Bydon M. Can machine learning algorithms accurately predict discharge to nonhome facility and early unplanned readmissions following spinal fusion? Analysis of a national surgical registry. *J Neurosurg Spine* 2019;31:568-78. [DOI PubMed](#)
20. Broda A, Sanford Z, Turcotte J, Patton C. Development of a risk prediction model with improved clinical utility in elective cervical and lumbar spine surgery. *Spine* 2020;45:E542-51. [DOI PubMed](#)
21. Churrua K, Pomare C, Ellis LA, et al. Patient-reported outcome measures (PROMs): a review of generic and condition-specific measures and a discussion of trends and issues. *Health Expect* 2021;24:1015-24. [DOI PubMed PMC](#)
22. Dansie EJ, Turk DC. Assessment of patients with chronic pain. *Br J Anaesth* 2013;111:19-25. [DOI PubMed PMC](#)
23. Mobbs RJ. From the subjective to the objective era of outcomes analysis: how the tools we use to measure outcomes must change to be reflective of the pathologies we treat in spinal surgery. *J Spine Surg* 2021;7:456-7. [DOI PubMed PMC](#)
24. Ahmad HS, Yang AI, Basil GW, et al. Developing a prediction model for identification of distinct perioperative clinical stages in spine surgery with smartphone-based mobility data. *Neurosurgery* 2022;90:588-96. [DOI PubMed](#)
25. Chauhan D, Ahmad HS, Subtirelu R, et al. Defining the minimal clinically important difference in smartphone-based mobility after spine surgery: correlation of survey questionnaire to mobility data. *J Neurosurg Spine* 2023;39:427-37. [DOI PubMed](#)
26. Boaro A, Leung J, Reeder HT, et al. Smartphone GPS signatures of patients undergoing spine surgery correlate with mobility and current gold standard outcome measures. *J Neurosurg Spine* 2021;35:796-806. [DOI PubMed PMC](#)
27. Lewandrowski KU, Alvim Fiorelli RK, Pereira MG, et al. Polytomous rasch analyses of surgeons' decision-making on choice of procedure in endoscopic lumbar spinal stenosis decompression surgeries. *Int J Spine Surg* 2024;18:164-77. [DOI PubMed PMC](#)
28. Lorio M, Martinson M, Ferrara L. Paired comparison survey analyses utilizing rasch methodology of the relative difficulty and estimated work relative value units of CPT[®] code 27279. *Int J Spine Surg* 2016;10:40. [DOI PubMed PMC](#)
29. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* 2021;14:49-58. [DOI PubMed PMC](#)
30. Mongan J, Moy L, Kahn CE Jr. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell* 2020;2:e200029. [DOI PubMed PMC](#)
31. 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. [DOI PubMed PMC](#)
32. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385:e078378. [DOI PubMed PMC](#)
33. Collins GS, Reitsma JB, Altman DG, Moons KG; TRIPOD Group. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group. *Circulation* 2015;131:211-9. [DOI PubMed PMC](#)
34. Zamanipoor Najafabadi AH, Ramspek CL, Dekker FW, et al. TRIPOD statement: a preliminary pre-post analysis of reporting and methods of prediction models. *BMJ Open* 2020;10:e041537. [DOI PubMed PMC](#)

Review

Open Access



Beyond AI and robotics: the dawn of surgical automation in spine surgery

Nishanth S. Sadagopan, Dillan Prasad, Rishi Jain, Christopher Ahuja, Nader S. Dahdaleh, Najib E. El Tecle 

Department of Neurological Surgery, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA.

Correspondence to: Prof. Najib E. El Tecle, Department of Neurological Surgery, Northwestern University Feinberg School of Medicine, 676 N. St. Clair Street, Suite 2210, Chicago, IL 60611, USA. E-mail: najib.eltecle@nm.org

How to cite this article: Sadagopan NS, Prasad D, Jain R, Ahuja C, Dahdaleh NS, El Tecle NE. Beyond AI and robotics: the dawn of surgical automation in spine surgery. *Art Int Surg* 2024;4:387-400. <https://dx.doi.org/10.20517/ais.2024.34>

Received: 31 May 2024 **First Decision:** 17 Oct 2024 **Revised:** 21 Oct 2024 **Accepted:** 29 Oct 2024 **Published:** 11 Nov 2024

Academic Editors: Andrew Gumbs, Peter Passias **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Artificial intelligence (AI), deep learning (DL), and machine learning (ML) algorithms are revolutionizing spine surgery. Soon, these technologies may allow the integration of automated devices into clinical practice. The roles of such devices are yet to be imagined and then developed, but one could assume that automated surgical devices can assist spine surgeons in a variety of ways, such as contextual guidance, precise screw placements, or intraoperative monitoring. In the not-too-distant future, such devices may be able to perform entire surgeries autonomously. Current literature suggests that advancements toward autonomous robotic surgery may improve surgical approaches and reduce negative clinical variation in spine surgery outcomes. This review aims to examine the current trends, practices, and advancements in surgical automation and provide an overview of the stages of automation of devices currently employed within spine surgery.

Keywords: Neurological surgery, spine surgery, robotic surgery, artificial intelligence, machine learning, deep learning, surgical automation

INTRODUCTION

Artificial intelligence, machine learning, deep learning, and surgical automation

Artificial Intelligence (AI) is revolutionizing healthcare by arguably mimicking human intelligence in machines, enabling them to perform complex tasks autonomously. At the forefront of this movement is



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



deep learning (DL), a subset of AI that uses neural networks structured in multiple layers of interconnected nodes^[1]. Convolutional neural networks (CNNs), a type of DL algorithm, have shown exceptional proficiency in analyzing medical images, and more recently, transformers have further pushed the capabilities of AI models. By using a self-attention mechanism to understand context, they enable parallel processing of data and incorporate positional encoding to maintain the order of input sequences. Transformers power state-of-the-art models like GPT, Llama, and Gemini, significantly advancing applications in text generation, classification, translation, and even extending to computer vision and speech recognition.

Nearly 30% of the world's data are produced by the healthcare sector, with 80% of the data being unstructured. It is estimated that the average American hospital produces 50 petabytes of data every year, double the size of the Library of Congress^[2]. This data-rich environment presents itself as a unique opportunity for AI.

Surgical automation aims to develop devices capable of performing surgery with varying degrees of autonomy without the intervention of humans^[3]. DL algorithms optimize surgical strategies based on pre- and intraoperative patient data, leveraging predictive models to anticipate complications and adapt surgical plans dynamically. The application of these technologies in surgical robotics includes systems for image-guided navigation, autonomous instrument control, and real-time decision support.

While the integration of AI-based frameworks in surgical robotics progresses at a guarded pace - as self-learning systems are still striving to achieve clinically acceptable confidence levels - mechanical advances in surgical devices have facilitated the integration of automation in the operating room (OR).

Current robotic developments focus on creating a streamlined operating space; guidance cameras, robotic arms, and attachments can be stored in a central console and manipulated from a single site. Robotic arms now offer movement up to 6 or 7 degrees of freedom (DOF) and are integrated with imaging derived from preoperative annotations, 3D field mapping cameras, and traditional O-arms to precisely guide operative trajectories^[4,5]. These innovations have enabled high-accuracy implant placement and corrections while reducing radiation exposure and the need to transport patients between different pieces of equipment. Haptic feedback and self-stabilizing arms have further improved safety outcomes for patients. In addition to reducing peak force applied, instrumentation collision, and risk of undesired tissue penetration, tactile sensation coupled with self-stabilization serves to reduce surgeon fatigue^[6-8]. As automated systems work toward more complex procedures, advances in both the hardware and software layers are necessary for surgical automation to materialize.

History and stages of surgical automation

The concepts of autonomous and robotic surgery go hand-in-hand and have significantly progressed together since their inception 40 years ago. The first surgical robot used in the OR was the programmable universal machine for assembly (PUMA) 200 robotic surgical arm in 1985, which performed precise neurosurgical biopsies^[9]. The 1990s introduced the PROBOT, designed for prostate surgery, and the Robodoc Surgical System, which enhanced hip replacement procedures^[10,11]. With the continued introduction of robotics in the surgical setting, emphasis was placed on developing a robot capable of implementing the "master-slave" framework; this framework utilizes an operator controlling the movements of the machine from a remote location. The first iteration of the teleoperation framework was the ZEUS robotic system and SOCRATES, allowing for a cholecystectomy to be done on a patient located in France by a surgeon in New York^[12]. Expanding on the capabilities of the ZEUS robotic system, Intuitive

Surgical Inc. created the da Vinci Surgical System, which received Food and Drug Administration (FDA) approval in 2000, revolutionizing minimally invasive surgery with unparalleled precision and control^[13]. Robotic-assisted surgery continues to evolve with innovations such as improved haptic feedback, real-time imaging, and augmented reality, significantly enhancing surgical outcomes and patient safety.

The integration of DL into robotic systems has led to automated surgical systems capable of enhancing surgical precision and efficiency and making decisions intraoperatively. These systems utilize sensor fusion, combining data from multiple sources such as cameras, force sensors, and navigation systems to provide comprehensive situational awareness, while advanced control algorithms enable real-time adjustments to the robotic instruments^[14]. The advancements in machine learning (ML) and DL have shifted the paradigm from developing robots governed by a master-slave framework, to autonomous surgical systems capable of aiding and making intraoperative decisions. The first autonomous surgery was conducted with the smart tissue autonomous robot (STAR) in 2016 for bowel anastomosis^[15]. This system conducts the surgery autonomously, only needing a human surgeon to approve its plan at the start of the procedure, and throughout its duration if correction was needed. TSolution one is another autonomous surgical system that can drill and carve bone for knee replacement surgery according to a predetermined plan but cannot distinguish between types of tissue^[16]. Thus, a human surgeon must clear a path for the device to access the bone by mobilizing the skin and fascia superficial to the bone.

Autonomous surgical systems currently employed in the OR exhibit a range of autonomous capabilities. The varying levels of autonomy in surgical devices necessitated a classification system to identify their operative capabilities. In 2017, Yang *et al.* proposed a framework outlining the stages of automation in surgical procedures based on the taxonomy of automated self-driving vehicles^[17,18]. These six stages delineate the progression from manual to fully autonomous surgical devices, encompassing various levels of human involvement and, conversely, machine autonomy. This framework provides a comprehensive roadmap for understanding the evolution of surgical automation, delineating distinct stages of technological advancement and human-machine collaboration. The stages are outlined below [Table 1].

Stage 0 - no autonomy

In this initial stage, surgeons perform procedures manually with minimal technological assistance. Surgeons rely solely on their skills and expertise to execute surgical tasks without the aid of automation technologies. This stage also includes teleoperated devices that respond directly to the surgeon's command even from a distance.

Stage 1 - robot assistance

This stage provides mechanical assistance during the procedure to aid the surgeon's skills. Surgeons operate consoles equipped with haptic feedback, enabling them to control robotic arms with precision while visualizing the surgical site through advanced automated imaging modalities.

Stage 2 - task autonomy

In stage 2, the operator maintains control of the system and the robot can perform surgeon-defined tasks autonomously. These semi-autonomous systems incorporate AI-driven algorithms to assist surgeons during specific phases of surgery. These systems analyze intraoperative data and provide contextual guidance, enhancing surgical precision and safety. Surgeons retain control over critical decision-making aspects while leveraging automation for assistance.

Table 1. Stages of surgical automation proposed by Yang *et al.*^[17]

Stage	Description
0	No autonomy
1	Robot assistance
2	Task autonomy
3	Conditional autonomy
4	High autonomy
5	Full autonomy

Stage 3 - conditional autonomy

Conditional autonomy represents a partnership between surgeons and robotic systems, where both entities contribute to surgical tasks. Surgeons choose the surgical plan, but then the robot implements the plan with predefined constraints under the surgeon's oversight. This stage fosters synergy between human expertise and machine capabilities.

Stage 4 - high autonomy

In stage 4, robotic systems assume greater responsibility for executing surgical tasks and can make decisions while under the surveillance of a human operator. Surgeons would oversee the procedure and intervene when necessary, ensuring patient safety and procedural integrity. Robotic systems leverage AI algorithms to adapt to dynamic surgical environments and make operative decisions, enhancing adaptability and responsiveness.

Stage 5 - full autonomy

Full autonomy represents the pinnacle of surgical automation, where robotic systems perform entire surgical procedures independently without direct human intervention. Currently, there are no devices approved for operative use with stage 5 autonomy.

This stratification of surgical autonomy has been utilized extensively to analyze the levels of automation of robotic devices in both pre-clinical and clinical phases of FDA approval^[19]. However, this framework has not been utilized to thoroughly analyze autonomous systems employed in spine surgery. This review aims to examine the current trends, practices, and advancements in surgical automation and serves as an overview of the stages of automation of devices currently employed within spine surgery.

METHODS

A comprehensive literature review was conducted using PubMed, Web of Science, and Google Scholar, focusing on autonomous robotic systems in spine surgery. The inclusion criteria encompassed research articles on robotic automation in spine surgery. Data were extracted regarding robot names, manufacturers, purposes, FDA status, automation stages, methods, results, and significance.

RESULTS OF THE LITERATURE REVIEW

Emerging autonomous technologies

The following technologies are currently not FDA-approved but are at the cutting edge of autonomous spine surgery [Table 2]. In some cases, companies have developed the discussed robotic platforms for explicitly surgical use; in other cases, research groups have adapted existing robotic platforms and integrated them with other components - such as sophisticated imaging modalities, augmented reality systems, and more - to achieve new degrees of autonomy and efficiency.

Table 2. Experimental autonomous spine surgery devices

Device or system	Manufacturer	Function	Automation stage
DLR Light-Weight Robot LWR-II	KUKA	Robotic drilling and milling for pedicle screw placement	Stage 2
LBR iiwa 7 R800	KUKA	Autonomous spinal sonography using a robotic ultrasound probe guided by a shadow-aware dual-agent framework	Stage 3
AOSRV	Shenzen Futuretec	Autonomous vertebral puncture and bone cement injection for PVP	Stage 4
7-DOF robotic manipulator	Politecnico di Milano and IRCCS Humanitas Research Hospital	Autonomous control of exoscope	Stage 3
RONNA	University of Zagreb	Frameless stereotactic neurosurgery for precise navigation	Stage 2
KUKA light weight robot 4+ and BTS smart-D motion capture system	KUKA	Pedicle screw fixation	Stage 2
Hand-held bone-cutting tool	University of Tokyo	Autonomous detection of bone penetration	Stage 2
6D-PKM surgical robot	Homi Bhabha National Institute	Autonomous registration improving overall accuracy in robot-based neurosurgery	Stage 3
AUBO-i5 robot with SRI force sensor	AUBO Robotic Technology and Shanghai Yuli Industrial	Autonomous laminectomy procedures	Stage 4
minaroHD	RWTH Aachen University	On-site teleoperated milling with haptic assistance for precise bone surgery	Stage 1

AOSRV: Autonomous orthopaedic surgical robotic for vertebroplasty; PVP: percutaneous vertebroplasty; DOF: degrees of freedom; RONNA: robotic neuronavigation; 6D-PKM: 6 degrees of freedom parallel kinematic mechanism.

Robots for pedicle cannulation or screw insertion

KUKA is a major German company that has developed numerous autonomous and assistive platforms over the years, primarily for commercial manufacturing and surgery. The DLR Light-Weight Robot LWR-II, developed by KUKA, was introduced in 2006 as an advanced robotic system for spinal surgeries, specifically for pedicle screw placement via robotic drilling and milling^[20]. Pedicle screw fixation requires high precision. Since 2006, several iterations of the LWR have come to market, with the iiwa and LBR Med being the most recent products. Although not FDA-approved, it represents a stage 2 automation in surgical robotics as it provides autonomy over strict tasks previously defined by the controlling surgeon. The system features a navigation system integrated with the robotic arm, ensuring precise control during the surgical procedure. Ortmaier *et al.* validated robotic performance in artificial bone and bovine spine models via quantitative comparison of drill-hole diameters, showing that the milling process performed by the LWR-II is superior to traditional drilling, providing enhanced accuracy and reduced surgical errors^[20].

Another group developed a semi-autonomous “shared control” pedicle screw fixation system that is deployed during the tapping phase of pedicle screw insertion. Lauretti *et al.* adapted the KUKA LWR 4+ such that the surgeon maintains full control over the procedure by maneuvering the robot’s end-effector using a control interface, aligning it along a pre-planned trajectory, and continuously monitoring the forces exerted on the patient’s spine during tapping^[21]. This system displays level 2 autonomy, similar to the DLR LWR-II. This method was found to enhance comfort, improve ergonomic postures, and reduce fatigue for the surgeon after testing on an anthropomorphic model.

Robots for laminectomy or bone drilling

While the primary role of robots in spine surgery is to aid in pedicle screw placement, autonomous surgical platforms are promising for other spinal conditions such as percutaneous vertebroplasty (PVP) as well. The autonomous orthopaedic surgical robotic for vertebroplasty (AOSRV) surgical system, developed by

Shenzen Futuretec and introduced in 2022, represents an experimental stage 4 automation in robotic vertebroplasty^[22], as it is capable of making intraoperative decisions based on preoperative planning. This system is designed for autonomous vertebral puncture and bone cement injection, specifically targeting procedures for spinal stabilization. The base robotic platform is the Orthobot XGK-6508A. It integrates preoperative planning with real-time intraoperative guidance using fused CT and C-arm fluoroscopic images. The system operates with a robotic arm capable of 6 DOF, ensuring precise movements and positioning. Key components include a bone drill and an injection propulsion unit, both equipped with force sensors for real-time pressure feedback, which enhance safety and real-time adjustment control during the procedure. The AOSRV demonstrated superior performance in comparative studies using a pig spine model, significantly reducing operation time, puncture adjustments, and intraoperative fluoroscopies while achieving high accuracy and lower bone cement leakage rates.

There is a notable absence of mature systems for robot-assisted laminectomy. A Chinese group at Peking University in 2023 proposed a novel integrative system for automated laminectomy based on a 6-DOF AUBO-i5 robotic arm which relies on preoperative CT to operate fully autonomously, a stage 4 innovation^[23]. The arm is further equipped with a force sensor and an ultrasonic osteotome, providing real-time feedback and precise control. The study involved 40 vertebrae from four cadavers. Robotic performance, as evaluated by mean deviation from the cutting path, was 0.67 mm at the superior point and 0.73 mm at the inferior point, with 83% of the laminectomy planes rated as grade A for accuracy and 81% considered safe. The system demonstrated high accuracy and efficiency, with no significant differences in deviation between thoracic and lumbar procedures.

Assistive robots for visualization

The KUKA LBR iiwa 7 R800 is a robotic arm introduced in 2022 that has been adapted for use in autonomous spinal sonography^[24]. Manual ultrasound acquisitions are costly and time-consuming as they require operation by trained sonographers. Groups have built the LBR iiwa 7 into a broader system to implement a “dual agent” framework (real-time reinforcement and DL) to autonomously guide an ultrasound probe in such a way that mimics the decision making of an expert sonographer by using view-specific acoustic shadowing as a robotic guidance marker^[24]. With the integration of DL software into the robotic arm, this system may have stage 3 autonomous capabilities as it would function autonomously within a set of parameters set by the surgeon. The validated system demonstrated high navigational accuracy, a promising finding for future autonomous surgical systems that will need to adjust in real time for micromovements and other intraoperative positional changes.

Further innovation in autonomous imaging includes autonomous neuro-registration. The Robotic neuronavigation (RONNA) system, developed by the University of Zagreb, was introduced in 2018 for frameless stereotactic neurosurgery for spinal applications^[25]. This stage 2 automation robot focuses on providing precise navigation for spinal surgeries without the need for invasive frames. The system is designed to be mounted on any robotic arm (the researchers used a KUKA robotic arm) and relies on several fiducial markers. RONNA has shown substantial accuracy in its applications; evaluations using different localization strategies revealed application errors in the sub-millimeter range, indicating a high level of precision in navigational tasks. RONNA’s ability to perform stereotactic procedures with minimal error could significantly improve the efficiency and safety of applicable procedures, making it a reliable alternative to traditional methods.

Exoscopes provide neurosurgeons with enhanced visualization and ergonomics compared to traditional surgical microscopes by projecting the surgical field onto a 2D or 3D monitor, thus offering a closer view

and better access to the surgical site without the need for the surgeon to contort themselves to maintain a clear perspective. However, as conventional exoscopes like the Aesculap Aeos require manual or foot joystick repositioning, groups have proposed designs for autonomous exoscope control, aiming to improve ergonomics and reduce the surgeon's physical and cognitive burden compared to joystick control. One group has developed a markerless method that uses visual data from the operating field to control and adjust the robotic arm of the exoscope in real time^[26]. Validation was conducted using a 7-DOF robotic manipulator with a stereo camera in an eye-in-hand setup. The system achieved 89% accuracy in target detection and tracking, enhanced efficiency with a significantly shorter operation time compared to that required for foot-joystick control, and a lower overall time that the instrument spent out of the field of view relative to joystick control.

Another promising system for autonomous neuro-registration is the 6 DOF parallel kinematic mechanism (6D-PKM) robot, a 2018 stage 3 innovation developed by an Indian group^[27]. This system autonomously navigates and measures fiducial marker coordinates in the patient's real space, based on preoperative imaging, eliminating the need for manual marker placement and reducing line-of-sight issues. Validation experiments using various phantoms, including a PVC skull model and acrylic blocks, demonstrated successful registration with a tracking error ranging from 0.50 ± 0.17 cm for low-speed movements to 1.38 ± 0.73 cm for high-speed movements. The proposed system also reduced overall registration time and minimized the cognitive and physical load on surgeons. The system's precision and repeatability were confirmed through experiments that consistently demonstrated high accuracy, indicating substantial potential benefits.

FDA-approved automated surgical systems

Current FDA-approved surgical systems predominantly fall between stages 0 and 2 proposed by Yang *et al.*, providing mechanical assistance or improving pre- and intraoperative visualizations to streamline the execution of procedural tasks^[17]. Nonetheless, their applications are wide-ranging and a testament to the potential of integrating robotic innovation into treatment paradigms^[28]. We provide an overview of six major approved systems and their performances in spine surgery [Table 3].

The da Vinci Surgical System (stage 0-1), developed by Intuitive Surgical, Inc., was first introduced for formal clinical use in 2000 and cleared to assist with minimally invasive protocols spanning general laparoscopic surgery. Now on its fifth iteration, the system consists of a surgeon console (with a high-resolution patient imaging system and master controls), a patient-side cart holding both robotic arms and EndoWrist instruments [with increased range of motion (ROM) and embedded force feedback], and a vision system for monitoring of the surgical area. Albeit limited, applications in human spine surgery have been very promising in controlled clinical settings. Molteni *et al.* noted the merits of the system in reaching benign, anterior C1-C2 lesions, evasion of extensive cervical dissection, greater freedom of movement, and more efficient oro/rhino-pharyngeal suturing^[29]. Perez-Cruet *et al.* employed an off-label, anterior approach to the resection of paraspinal tumors presenting with intrathoracic extension in two patients with the da Vinci^[30]. Additionally, Sadagopan *et al.* presented a resection of a sciatic notch lipoma with the da Vinci Machine, demonstrating superior visualization and preservation of critical paraspinal and pelvic structures^[31].

The Mazor X Stealth Edition (stage 1) by Medtronic was cleared for spine surgery application in 2018 and serves to streamline operative navigation. The system brings together preoperative and intraoperative 3D visualization/implant trajectory tracking (MRI/CT guided) software, a linear optic camera, high-speed drill systems (Midas Rex and Stealth-Midas), graft inserters (Catalyft PL Expandable Interbody System), and

Table 3. FDA-approved robotic spine surgery systems

Device or system	Manufacturer	Function	Automation stage
da Vinci Surgical System	Intuitive Surgical Inc.	Minimally invasive laparoscopic surgery	Stage 0-1
Mazor X Stealth Edition	Medtronic	Pedicle screw insertion, TLIF, and MIDLF	Stage 1
ROSA Spine	Medtech	Pedicle screw insertion and TLIF	Stage 2
ExcelsiusGPS	Globus Medical	Fluoroscopy-guided pedicle screw insertion	Stage 2
Renaissance	Mazor Robotics	Pedicle screw insertion	Stage 1
SpineAssist	Mazor Robotics	Pedicle screw insertion	Stage 1

FDA: Food and Drug Administration; TLIF: transforaminal lumbar interbody fusion; MIDLF: midline lumbar interbody fusion.

additional custom attachments to support transforaminal lumbar interbody fusion (TLIF), midline lumbar interbody fusion (MIDLF), and deformity-correction procedures^[32]. The Mazor X system has particularly been cited as facilitating accurate pedicle screw placement (although relatively difficult at certain levels and challenging registration with complex deformities) with high degrees of safety^[33].

The ROSA Spine by Medtech (stage 2) gained FDA clearance in 2015; similar to the Mazor X system, it enables pre- and intraoperative CT-based implant trajectory planning and screw insertion over guidewires. Its most recent version - the ROSA ONE - allows for complete integration with the ROSA ecosystem for robotic arm accessory attachment. Applications of the ROSA Spine system are not limited but have been most commonly cited for arthrodesis (e.g., TLIF) with high accuracy^[34,35].

Introduced by Globus Medical in 2017, the ExcelsiusGPS (stage 2) improves pre- and intraoperative planning similar to the Mazor X and ROSA frameworks, bringing together a rigid robotic arm, surveillance markers and sensors, and visualization platforms. Relative to fluoroscopy-guided insertion, a GLOBUS-led study demonstrated accuracy and safety improvements (0% Grade 0 breaches) with significant reductions in time for placement and exposure to radiation and without the need for Kirschner guidewire placement^[36]. The system enables the execution of common fusion procedures with better overall alignment and minimized postoperative complications.

The Renaissance system (stage 1) by Mazor Robotics (acquired by Medtronic) was approved in 2011 and enables minimally invasive correction of back pain, degenerative pathologies (slipped disk, scoliosis, nerve impingement), and a range of fusion procedures. The Renaissance consists of a CT/MRI-guided preoperative image station with trajectory mapping/auto-alignment software, multiple framework arms, and a robot with 6 DOF; while screw misalignment and skiving are cited concerns with this model, the Renaissance demonstrates notably low breach rates with some studies citing a 1.1% rate comparable to the Mazor X - reduced time to procedure completion, and minimal learning curve for users^[37,38].

The SpineAssist (stage 1), another innovation by Mazor Robotics (acquired by Medtronic), is the formal predecessor to the Renaissance system and gained FDA approval in 2004. Similar to the Renaissance, the SpineAssist system lacks integrated navigation but provides preoperative imaging (CT) compatible with intraoperative fluoroscopy. As an earlier iteration, the SpineAssist is unable to flatten bone at screw entry points - compounding concerns of skiving - and stalls in processing speed comparisons^[39]. However, the system has achieved consistently clinically acceptable screw placement; Devito *et al.* reported 98.3% of placed screws as falling within a defined “safe zone”, with breaches exceeding 2 mm in only 1.7% of placements across 3,271 total pedicle screws^[40]. Notably, they cited no permanent peripheral nerve damage in their cohort and particular merits for percutaneous approaches lacking anatomical landmarks. Nonetheless, the SpineAssist has since been significantly improved in its successors.

DISCUSSION

Levels of automation specific for spine surgery

The development and implementation of autonomous devices in spine surgery are rapidly expanding due to the confluence of mechanistic, robotic, and AI innovations. Though the stratification of Yang *et al.* can be utilized to classify the stages of automated surgical devices in the broader surgical community, there remains a need for a spine-specific classification encompassing the unique challenges and future directions of the field^[17]. Here, we propose our own classification for different levels of surgical automation currently employed, or soon to be developed, in spine surgery [Table 4].

Level 0 - manual

All surgical tasks are performed manually by the surgeon without any automated assistance. The surgeon relies on their expertise to navigate and execute the procedure. Procedures encompassed within this level would include freehand pedicle screw placement.

Level 1 - computer-assisted navigation

In level 1, the system provides passive support, such as enhanced imaging and navigation tools, to assist the surgeon in planning and executing the surgery through neuronavigation, endoscopic, or exoscopic visualization. The surgeon retains full control of the surgical instruments while receiving assistance in advanced visualization.

Level 2 - task-specific automation

In level 2, automation assists with specific tasks, such as pedicle screw placement or drilling, under the surgeon's direct supervision. The surgeon initiates these tasks and monitors their execution, intervening if necessary.

Level 3 - semi-autonomous spine surgery

In level 3, the system can autonomously perform more complex sequences of tasks, such as those required for decompression or arthrodesis procedures, but still requires human surgeon oversight. The surgeon supervises the procedure and can intervene to ensure precision and safety.

Level 4 - highly autonomous spine surgery

In level 4, the system performs the majority of the surgical tasks autonomously with minimal human intervention. The surgeon's role is primarily supervisory, stepping in only for unexpected situations or critical decision making. The robot at this stage does not technically need a human and should be able to reach a safe position without human intervention (e.g., packing a bleeding wound in preparation for an angiogram).

Level 5 - fully autonomous spine surgery

Level 5 represents the highest level of autonomy within spine surgery, where the system can conduct the entire spine surgery autonomously (regardless of complexity), from planning to execution, without human intervention. This would represent "true autonomy" in spine surgery.

Our classification presents a reinterpretation of the initial stratification by Yang *et al.* made specifically for autonomous advancements in spine surgery^[17]. The complicated anatomy and critical locations of paraspinal neurovasculature make implementing autonomy in spine surgery exceedingly difficult. As progress is made toward increasingly autonomous surgical devices in spine, there is a greater need for surgeons to identify the proper devices for their respective procedures and to understand the capabilities of

Table 4. Levels of automation in spine surgery

Level	Description	Example	Device
0	Manual	Freehand pedicle screw placement	Standard surgical tools
1	Computer-assisted navigation	Neuronavigation	RONNA
2	Task-specific automation	Robot-assisted pedicle screw placement	DLR Light-Weight Robot LWR-II
3	Semi-autonomous spine surgery	Autonomous laminectomy with surgeon oversight	AUBO-i5 robot with SRI force sensor
4	Highly autonomous spine surgery	Complete autonomy over all surgical steps (initial exposure, fusion, closure) with some surgeon oversight	Not yet developed
5	Fully autonomous spine surgery	No human surgeon intervention	Not yet developed

RONNA: Robotic neuronavigation.

these novel surgical systems. This classification can serve as a guideline for stratifying the emerging technologies that are specific to the challenges and complexities of spine surgery.

Benefits of surgical automation

The standardization of surgical techniques through automation and AI reduces variation in clinical outcomes and enhances precision. By leveraging algorithms and robotic systems, surgical procedures can be executed with increased accuracy, leading to fewer errors and improved patient outcomes^[41,42]. Automation enables the execution of predefined strategies with consistency, minimizing the influence of human factors and ensuring reproducibility across different surgical settings.

The adoption of automation in surgery allows for the increased bandwidth of surgical staff to focus on human needs. By offloading repetitive and mundane tasks to automated systems, surgical teams can redirect their attention toward providing personalized care, communicating with patients and their families, and addressing the emotional and psychological aspects of the surgical experience. This shift in focus toward patient-centered care fosters a more holistic approach to healthcare delivery, promoting better overall patient satisfaction and well-being.

Incorporating preoperative and intraoperative monitoring enhances surgical precision and safety^[43]. AI-driven algorithms analyze imaging data in real time, providing surgeons with detailed insights into patient anatomy and pathology, facilitating informed decision making during surgery. Additionally, the integration of preoperative and intraoperative variables enables the early recognition and mitigation of postoperative complications, morbidity, and mortality.

In high-complexity scenarios, where surgeons' decision-making capacity may be compromised due to stress or cognitive overload, automation may be preferable. Automated systems can execute predefined surgical strategies and adapt to rapidly changing conditions, ensuring timely and effective interventions even in the most challenging circumstances^[10]. By augmenting surgeons' capabilities with AI-driven technologies, the risk of errors and adverse events can be minimized, ultimately improving success rates.

Integrating automation and AI with big data analytics holds the potential for advancing surgical practice^[44]. By harnessing vast amounts of patient data, including demographic information, clinical histories, and treatment outcomes, AI algorithms can identify patterns, predict patient responses to treatment, and optimize surgical strategies. This integration enables personalized healthcare delivery and informs

automated operations, leading to improved patient outcomes and enhanced efficiency in surgical practice.

Limitations of automated surgery

The integration of automation and AI into surgical practice holds immense promise for enhancing patient outcomes and optimizing healthcare delivery in spine surgery. However, several limitations and challenges must be addressed to ensure safe and effective implementation in clinical settings.

One significant limitation of automation in surgery lies in its ability to handle complex scenarios with potential unexpected intraoperative complications. While AI algorithms excel in analyzing structured data and predicting outcomes based on predefined parameters, they may struggle to adapt to unanticipated events or variations in patient anatomy^[45]. In such cases, human intervention and expertise remain indispensable for navigating unforeseen challenges and ensuring patient safety.

The high cost associated with robotic surgery, including equipment acquisition, maintenance, and training, presents a significant barrier for many healthcare institutions^[46]. As the level of autonomy in autonomous devices increases, the regulatory challenges also escalate. The FDA reviews and clears robotic-assisted devices via the 510(K) premarket notification process^[17]. However, higher-risk devices, such as those classified as stage 3 or higher, may face more stringent regulatory scrutiny, leading to significantly increased costs of bringing the device to market^[47]. The difference in costs for higher autonomy could be reflected in the cost to the patient. Thus, increases in levels of autonomy could further exacerbate healthcare disparities^[48].

The legal and ethical ramifications of autonomous surgery further compound the challenges surrounding its complete implementation. Analogous to the legal debates surrounding self-driving cars, questions regarding liability and accountability arise when patients are harmed, or complications occur during autonomous surgical procedures^[43]. The FDA approves devices but not the practice of medicine itself. Higher levels of automated devices, such as stage 4 or 5, will be making clinical decisions intraoperatively to the same level as a human physician^[49]. New regulatory bodies will likely need to be created to oversee the practice of highly autonomous devices to ensure that the safety of the patients is upheld.

CONCLUSION

The advancements in surgical automation and robotics in spine surgery signify a transformative shift in medical practice. From the initial introduction of robot-assisted systems to the development of semi-autonomous platforms, the field has witnessed significant technological progress. These systems enhance surgical precision and reduce operative time, offering potential benefits over traditional methods^[37]. The integration of AI, ML, and DL algorithms into these robotic systems may further optimize surgical planning and execution, allowing for real-time adjustments and improved outcomes.

Despite these advancements, the journey toward fully autonomous surgery is still in its early stages. With the increasing integration of real-time imaging with robotic platforms, surgical systems will become increasingly autonomous as computer vision improves unsupervised decision making. Better visualization will enable robotic systems to process more data and thus execute better movements in real time, from micro-corrections to serious changes of course when a complication may arise during surgery. Further, the autonomous spine surgery systems currently in the pipeline far exceed the scope of pedicle screw insertion alone: from neuro-registration to autonomous exoscopic guidance, groups have proposed innovative robotic approaches to many procedures and standing problems. These technologies promise to enhance the precision and safety of spinal surgeries, reducing the cognitive and physical load on surgeons.

The absence of FDA approval for many of these cutting-edge systems highlights the ongoing need for rigorous clinical validation and regulatory approval processes. While these systems have demonstrated promising accuracy and efficiency in early studies, they tend to fall prey to similar limitations; challenges such as high implementation cost, poor situational generalizability, regulatory hurdles, and the need for human oversight in complex scenarios remain significant barriers to widespread adoption.

The future of spine surgery lies in the continued integration of AI-driven technologies, which can analyze vast amounts of patient data to inform surgical decisions and predict outcomes. The potential for personalized surgical approaches, guided by big data analytics and real-time intraoperative monitoring, holds promise for improving patient care and reducing variability in surgical outcomes. Yet, the ethical and legal implications of autonomous surgical systems, including issues of liability and accountability, must be carefully addressed to ensure patient safety. While surgical automation in spine surgery is advancing rapidly, the full realization of its potential will require overcoming significant challenges. Our current trajectory suggests a future where autonomous systems will play an increasingly central role in spine surgery.

DECLARATIONS

Authors' contributions

Conceived the project: Sadagopan NS, El Teclé NE

Conducted the literature review, drafted the initial manuscript: Sadagopan NS, Prasad D, Jain R

Created the tables: Sadagopan NS

Reviewed and edited the manuscript: Sadagopan NS, Prasad D, Jain R, Ahuja C, Dahdaleh NS, El Teclé NE

All authors read and approved the final manuscript.

Availability of data and materials

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2024.

REFERENCES

1. Stahlschmidt SR, Ulfenborg B, Synnergren J. Multimodal deep learning for biomedical data fusion: a review. *Brief Bioinform* 2022;23:bbab569. DOI PubMed PMC
2. Nambiar J, Juyal A. Healthcare organizations must create a strong data foundation to fully benefit from generative AI. Available from: <https://www.cio.com/article/1293408/healthcare-organizations-must-create-a-strong-data-foundation-to-fully-benefit-from-generative-ai.html#>. [Last accessed on 5 Nov 2024].
3. Shademan A, Dumont MF, Leonard S, Krieger A, Kim PCW. Feasibility of near-infrared markers for guiding surgical robots. In:

- Optical Modeling and Performance Predictions VI; 2013. pp. 123-32. DOI
4. Buza JA 3rd, Good CR, Lehman RA Jr, et al. Robotic-assisted cortical bone trajectory (CBT) screws using the Mazor X Stealth Edition (MXSE) system: workflow and technical tips for safe and efficient use. *J Robot Surg* 2021;15:13-23. DOI PubMed
 5. Drossopoulos PN, Sharma A, Ononogbu-Uche FC, et al. Pushing the limits of minimally invasive spine surgery-from preoperative to intraoperative to postoperative management. *J Clin Med* 2024;13:2410. DOI PubMed PMC
 6. Nakazawa A, Nanri K, Harada K, et al. Feedback methods for collision avoidance using virtual fixtures for robotic neurosurgery in deep and narrow spaces. In: 2016 6th IEEE International Conference on Biomedical Robotics and Biomechanics (BioRob); 2016 Jun 26-29; Singapore. IEEE; 2016. pp. 247-52. DOI
 7. Bergholz M, Ferle M, Weber BM. The benefits of haptic feedback in robot assisted surgery and their moderators: a meta-analysis. *Sci Rep* 2023;13:19215. DOI PubMed PMC
 8. Kuo LJ, Ngu JC, Lin YK, Chen CC, Tang YH. A pilot study comparing ergonomics in laparoscopy and robotics: beyond anecdotes, and subjective claims. *J Surg Case Rep* 2020;2020:rjaa005. DOI PubMed PMC
 9. Kwoh YS, Hou J, Jonckheere EA, Hayati S. A robot with improved absolute positioning accuracy for CT guided stereotactic brain surgery. *IEEE Trans Biomed Eng* 1988;35:153-60. DOI
 10. Davies B. A review of robotics in surgery. *Proc Inst Mech Eng H* 2000;214:129-40. DOI PubMed
 11. Paul HA, Bargar WL, Mittlestadt B, et al. Development of a surgical robot for cementless total hip arthroplasty. *Clin Orthop Relat Res* 1992;57-66. PubMed
 12. Marescaux J, Leroy J, Gagner M, et al. Transatlantic robot-assisted telesurgery. *Nature* 2001;413:379-80. DOI PubMed
 13. Rosen J, Hannaford B, Satava RM. Surgical robotics: systems applications and visions. New York: Springer; 2011. DOI
 14. Ma R, Vanstrum EB, Lee R, Chen J, Hung AJ. Machine learning in the optimization of robotics in the operative field. *Curr Opin Urol* 2020;30:808-16. DOI PubMed PMC
 15. Saeidi H, Opfermann JD, Kam M, et al. Autonomous robotic laparoscopic surgery for intestinal anastomosis. *Sci Robot* 2022;7:eabj2908. DOI PubMed PMC
 16. Liow MHL, Chin PL, Pang HN, Tay DK, Yeo SJ. THINK surgical TSolution-One[®] (Robodoc) total knee arthroplasty. *SICOT J* 2017;3:63. DOI PubMed PMC
 17. Yang GZ, Cambias J, Cleary K, et al. Medical robotics-regulatory, ethical, and legal considerations for increasing levels of autonomy. *Sci Robot* 2017;2:eaaam8638. DOI PubMed
 18. Committee SO-RAVS. Taxonomy and definitions for terms related to on-road motor vehicle J3016_202104. Available from: https://www.sae.org/standards/content/j3016_202104/. [Last accessed on 5 Nov 2024].
 19. Rivero-Moreno Y, Rodriguez M, Losada-Muñoz P, et al. Autonomous robotic surgery: has the future arrived? *Cureus* 2024;16:e52243. DOI PubMed PMC
 20. Ortmaier T, Weiss H, Döbele S, Schreiber U. Experiments on robot-assisted navigated drilling and milling of bones for pedicle screw placement. *Int J Med Robot* 2006;2:350-63. DOI PubMed
 21. Lauretti C, Cordella F, Tamantini C, Gentile C, Luzio FSD, Zollo L. A surgeon-robot shared control for ergonomic pedicle screw fixation. *IEEE Robot Autom Lett* 2020;5:2554-61. DOI
 22. Huang J, Xing T, Cheng Z, et al. AOSRV: development and preliminary performance assessment of a new robotic system for autonomous percutaneous vertebroplasty. *Int J Med Robot* 2022;18:e2456. DOI PubMed
 23. 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. DOI PubMed
 24. Li K, Xu Y, Wang J, Ni D, Liu L, Meng MQ. Image-guided navigation of a robotic ultrasound probe for autonomous spinal sonography using a shadow-aware dual-agent framework. *IEEE Trans Med Robot Bionics* 2022;4:130-44. DOI
 25. Šuligoj F, Jerbić B, Šekoranja B, Vidaković J, Švaco M. Influence of the localization strategy on the accuracy of a neurosurgical robot system. *TFAMENA* 2018;42:27-38. DOI
 26. Iovene E, Casella A, Iordache AV, et al. Towards exoscope automation in neurosurgery: a markerless visual-servoing approach. *IEEE Trans Med Robot Bionics* 2023;5:411-20. DOI
 27. Kaushik A, Dwarakanath TA, Bhutani G. Autonomous neuro-registration for robot-based neurosurgery. *Int J Comput Assist Radiol Surg* 2018;13:1807-17. DOI PubMed
 28. D'Souza M, Gendreau J, Feng A, Kim LH, Ho AL, Veeravagu A. Robotic-assisted spine surgery: history, efficacy, cost, and future trends. *Robot Surg* 2019;6:9-23. DOI PubMed PMC
 29. Molteni G, Greco MG, Presutti L. Transoral robotic-assisted surgery for the approach to anterior cervical spine lesions. *Eur Arch Otorhinolaryngol* 2017;274:4011-6. DOI PubMed
 30. Perez-Cruet MJ, Welsh RJ, Hussain NS, Begun EM, Lin J, Park P. Use of the da Vinci minimally invasive robotic system for resection of a complicated paraspinal schwannoma with thoracic extension: case report. *Neurosurgery* 2012;71:209-14. DOI PubMed
 31. Sadagopan NS, Poylin VY, Jun C, El Tecle NE, Wolinsky JP. Robotic resection of a sciatic notch lipoma using the DaVinci Surgical System: 2-dimensional operative video. *Oper Neurosurg* 2024;27:381-2. DOI PubMed
 32. Medtronic. MazorTM robotic guidance platform. Available from: <https://www.medtronic.com/en-us/healthcare-professionals/products/surgical-robotics/robotic-systems/mazor-robotic-guidance-system.html>. [Last accessed on 5 Nov 2024].
 33. Lieberman IH, Kisinde S, Hesselbacher S. Robotic-assisted pedicle screw placement during spine surgery. *JBJS Essent Surg Tech* 2020;10:e0020. DOI PubMed PMC

34. Chenin L, Capel C, Fichten A, Peltier J, Lefranc M. Evaluation of screw placement accuracy in circumferential lumbar arthrodesis using robotic assistance and intraoperative flat-panel computed tomography. *World Neurosurg* 2017;105:86-94. [DOI PubMed](#)
35. Lefranc M, Peltier J. Evaluation of the ROSA™ Spine robot for minimally invasive surgical procedures. *Expert Rev Med Devices* 2016;13:899-906. [DOI PubMed](#)
36. Jiang B, Karim Ahmed A, Zygorakis CC, et al. Pedicle screw accuracy assessment in ExcelsiusGPS® robotic spine surgery: evaluation of deviation from pre-planned trajectory. *Chin Neurosurg J* 2018;4:23. [DOI PubMed PMC](#)
37. Lee NJ, Buchanan I, Boddapati V, et al. P16. Is there a difference in screw accuracy, robot time per screw, robot abandonment, and radiation exposure between the Mazor X and the renaissance? A propensity-matched analysis of 1,179 robot-assisted screws. *Spine J* 2021;21:S147-8. [DOI](#)
38. BÄcker HC, Freibott CE, Perka C, Weidenbaum M. Surgeons' learning curve of renaissance robotic surgical system. *Int J Spine Surg* 2020;14:818-23. [DOI PubMed PMC](#)
39. Alluri RK, Avrumova F, Sivaganesan A, Vaishnav AS, Lebl DR, Qureshi SA. Overview of robotic technology in spine surgery. *HSS J* 2021;17:308-16. [DOI PubMed PMC](#)
40. Devito DP, Kaplan L, Dietl R, et al. Clinical acceptance and accuracy assessment of spinal implants guided with SpineAssist surgical robot: retrospective study. *Spine* 2010;35:2109-15. [DOI](#)
41. Connor MJ, Dasgupta P, Ahmed HU, Raza A. Autonomous surgery in the era of robotic urology: friend or foe of the future surgeon? *Nat Rev Urol* 2020;17:643-9. [DOI PubMed](#)
42. Shademan A, Decker RS, Opfermann JD, Leonard S, Krieger A, Kim PC. Supervised autonomous robotic soft tissue surgery. *Sci Transl Med* 2016;8:337ra64. [DOI PubMed](#)
43. O'Sullivan S, Nevejans N, Allen C, et al. Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *Int J Med Robot* 2019;15:e1968. [DOI PubMed](#)
44. Bhandari M, Zeffiro T, Reddiboina M. Artificial intelligence and robotic surgery: current perspective and future directions. *Curr Opin Urol* 2020;30:48-54. [DOI PubMed](#)
45. Moustris GP, Hiridis SC, Deliparaschos KM, Konstantinidis KM. Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature. *Int J Med Robot* 2011;7:375-92. [DOI PubMed](#)
46. Attanasio A, Scaglioni B, De Momi E, Fiorini P, Valdastrì P. Autonomy in surgical robotics. *Annu Rev Control Robot Auton Syst* 2021;4:651-79. [DOI](#)
47. Lauer M, Barker JP, Solano M, Dubin J. FDA device regulation. *Mo Med* 2017;114:283-8. [PubMed PMC](#)
48. Gkegkes ID, Mamais IA, Iavazzo C. Robotics in general surgery: a systematic cost assessment. *J Minim Access Surg* 2017;13:243-55. [DOI PubMed PMC](#)
49. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg* 2018;268:70-6. [DOI PubMed PMC](#)

Original Article

Open Access



Pseudarthrosis following adult spinal deformity surgery may be predicted with preoperative MRI adipose tissue features: an artificial intelligence study on raw 3D imaging

Graham W. Johnson¹ , Hani Chanbour², Derek J. Doss¹, Ghassan S. Makhoul¹, Amir M. Abtahi^{2,3}, Byron F. Stephens^{2,3}, Scott L. Zuckerman^{2,3}

¹Department of Biomedical Engineering, Vanderbilt University, Nashville, TN 37212, USA.

²Department of Neurological Surgery, Vanderbilt University Medical Center, Nashville, TN 37212, USA.

³Department of Orthopedic Surgery, Vanderbilt University Medical Center, Nashville, TN 37212, USA.

Correspondence to: Dr. Scott L. Zuckerman, Department of Neurological Surgery, Vanderbilt University Medical Center, Medical Center North T-4224, Nashville, TN 37212, USA. E-mail: scott.zuckerman@vumc.org

How to cite this article: Johnson GW, Chanbour H, Doss DJ, Makhoul GS, Abtahi AM, Stephens BF, Zuckerman SL. Pseudarthrosis following adult spinal deformity surgery may be predicted with preoperative MRI adipose tissue features: an artificial intelligence study on raw 3D imaging. *Art Int Surg* 2024;4:401-10. <https://dx.doi.org/10.20517/ais.2024.40>

Received: 12 Jun 2024 **First Decision:** 10 Oct 2024 **Revised:** 26 Oct 2024 **Accepted:** 31 Oct 2024 **Published:** 1 Dec 2024

Academic Editor: Andrew A. Gumbs **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Aim: The purpose of this study is to investigate the utility of incorporating magnetic resonance imaging (MRI) into an artificial intelligence (AI) model to preoperatively predict pseudarthrosis for patients undergoing adult spinal deformity (ASD) surgery.

Methods: A retrospective cohort study was conducted on patients undergoing ASD surgery at Vanderbilt University Medical Center with at least 2 years of follow-up. We first collected demographic variables and measured traditional radiographic variables with Surgimap software. The primary outcome of interest was pseudarthrosis, defined as mechanical pain without evidence of bony union with or without a rod fracture. Next, cohort differences between patients diagnosed with and without pseudarthrosis were evaluated with *t*-tests for continuous variables and chi-squared tests for categorical variables using Bonferroni-Holm multiple comparison correction. Using a subpopulation of patients with preoperative thoracic MRI available, a three-dimensional convolutional neural network (3D-CNN) with five-fold nested cross-validation was developed to predict pseudarthrosis - accuracy was evaluated with the Youden index. Finally, class activation mapping (CAM) was



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



conducted to visualize the MRI features utilized by the model for accurate classifications.

Results: Of 191 patients undergoing ASD surgery, the demographic and traditional radiographic variables were collected, and only age was observed to be significantly different between the patients diagnosed with pseudarthrosis (69.9 ± 10.1 years old) and those without (60.9 ± 19.9), with a *t*-test *P*-value of 0.003. The 3D-CNN demonstrated an average Youden index of 0.49 ± 0.25 on the withheld data, with a *P*-value of $5.50e-3$ compared to an equivocal null model. Finally, CAM consistently revealed posterior adipose tissue to be most important in preoperatively predicting pseudarthrosis.

Conclusion: Adipose tissue features in MRI, independent of body mass index (BMI), may be useful for preoperatively predicting pseudarthrosis. Overall, this work demonstrates the capabilities of raw imaging AI in spine surgery and can serve as the basis for a deeper biological inquiry into the pathogenesis of pseudarthrosis.

Keywords: Adult spinal deformity, artificial intelligence, deep learning, machine learning, magnetic resonance imaging, pseudarthrosis

INTRODUCTION

Pseudarthrosis (or nonunion) is defined as the failure of bone to fuse following surgical fixation and is a common complication of adult spinal deformity (ASD) surgery, with incidence rates ranging from 5%-35%^[1,2]. Pseudarthrosis is associated with recurrent pain and neurologic symptoms, can be a reason for reoperation, and can occur with or without rod fracture^[3]. Despite its prevalence and contribution to patient morbidity, the risk factors for pseudarthrosis are difficult to characterize. A preoperative risk factor is thought to be age, with multiple studies suggesting that patients over the age of 55 experience higher rates of pseudarthrosis^[4-6]. Additionally, an intraoperative risk factor is thought to be fusion to the sacrum^[7]. However, there remains debate in the literature about these risk factors, and few validated tools are available for the surgeon to preoperatively prognosticate pseudarthrosis occurrence.

Due to the difficulty in prognostication, more advanced artificial intelligence (AI) modeling techniques have been developed to augment surgical decision workflows for ASD surgery^[8-16]. Specifically, Scheer *et al.* developed a decision tree model from 82 variables that achieved 91% accuracy in predicting pseudarthrosis following ASD surgery^[17]. This high level of accuracy demonstrates its promise for clinical application. An underutilized extension of this framework is to utilize raw imaging data to augment predictive models. Of interest, AI models that ingest raw imaging can be directly interpreted to gain insight into nuanced patient characteristics impossible to capture in demographic variables. One imaging modality of high interest is magnetic resonance imaging (MRI) due to the detailed soft tissue signal captured. Thus, these advanced imaging models can aid preoperative decision-making, but more importantly, they can provide insight into the biological variables that may drive pseudarthrosis pathogenesis.

With the above considerations, this work aims to characterize the raw preoperative MRI features that may predict the occurrence of pseudarthrosis. In a cohort of patients undergoing ASD surgery at the major academic medical center, we sought to: (1) develop an AI model that utilizes raw preoperative MRI to predict pseudarthrosis following ASD surgery; and (2) interpret the model with class activation mapping (CAM) to understand the imaging features used to classify pseudarthrosis.

METHODS

Patient population

The study included a population of 191 patients who underwent ASD surgery at a single institution from

Table 1. Demographic and surgical variables by pseudarthrosis

Demographic and surgical variables	Total cohort (N = 191)	No Pseud. (N = 143)	Pseud. (N = 48)	P-value
Age at surgery, mean ± SD	63.1 ± 18.4	60.9 ± 19.9	69.9 ± 10.1	0.003
BMI, mean ± SD	28.8 ± 7.0	28.8 ± 7.3	29.0 ± 8.2	0.874
Gender female, n (%)	146 (76.4)	108 (75.5)	38 (79.2)	0.607
Comorbidities, n (%)				
Diabetes	26 (13.6)	15 (10.5)	11 (22.9)	0.030
COPD	48 (25.1)	32 (22.4)	16 (33.3)	0.130
Heart failure	24 (12.6)	16 (11.2)	8 (16.7)	0.322
Hypertension	122 (63.9)	86 (60.1)	36 (75.0)	0.064
Osteoporosis	40 (20.9)	32 (22.4)	8 (16.7)	0.400
Surgical variables				
Previous fusion, n (%)	56 (29.3)	44 (30.8)	12 (25.0)	0.447
Pelvic fixation, n (%)	150 (78.5)	106 (74.1)	44 (91.7)	0.010
TIL, mean ± SD	10.6 ± 3.0	10.4 ± 3.1	10.2 ± 3.0	0.697
UIV Region, n (%)				
Upper thoracic	71	57	14	-
Thoracolumbar	120	86	34	0.185

P-values in bold passed Bonferroni-Holm multiple comparison correction. SD: Standard deviation; BMI: body mass index; COPD: chronic obstructive pulmonary disease; TIL: total instrumented levels; UIV: upper instrumented vertebra.

2009-2021 and had at least 2-year follow-up. A subpopulation of 59 patients had presurgical thoracic MRI available for raw imaging deep learning analysis. The electronic medical record was mined for demographic variables outlined in Table 1. Pseudarthrosis was defined with a combination of clinical semiology and radiographic evidence of fusion failure captured on coronal and sagittal computed tomography (CT) scan, with or without rod fracture. Every symptomatic rod fracture in our series was given a diagnosis of pseudarthrosis as well. Next, each patient's scoliosis radiographs were de-identified and processed with Surgimap v2.3.2.1 (Nemaris Inc, Methuen, Massachusetts, USA) to acquire traditional radiographic measurements [Table 2]. To evaluate any correlation between the demographic/radiographic variables and pseudarthrosis incidence, two-population *t*-tests for continuous variables and chi-squared tests for categorical variables were conducted with Bonferroni-Holm multiple comparison correction.

MRI deep learning

Next, a three-dimensional convolutional neural network (3D-CNN) was developed to input raw thoracic MRIs, demographic variables, and Surgimap measured variables [Figure 1]^[18]. Only patients with MRI available were included in this study. MRI images were resliced to the three dimensions of 256 × 256 × 20 voxels, histogram equalized, and augmented using random flips, noise, bias field, blur, and affine/elastic deformations to a total of 1,080 images. Five-fold nested cross-validation with a train/validate/test split ratio of 70%/20%/10% was used to prevent overfitting and evaluate the generalizability of the model^[19]. Importantly, all splits were conducted at the patient level. A Youden index (sensitivity + specificity - 100%) was calculated for all completely withheld test partitions for each fold. The Youden index reflects the true positive, true negative, false positive, and false negative rate of the model on completely withheld validation data. The mean Youden index of all folds was tested against an equivocal null model with a Youden index value of 0 using a single population Student's *t*-test^[20].

Imaging feature attention mapping

Finally, to interpret the model and elucidate MRI features used for correct classification, the CNN architecture was modified to accommodate gradient class activation mapping (Grad-CAM)^[21]. This

Table 2. Traditional radiographic measurements by pseudarthrosis

Radiographic measurements degrees mean \pm SD (unless otherwise noted)	Total cohort (N = 191)	No pseud. (N = 143)	Pseud. (N = 48)	P-value
Coronal measurements				
C7PL, mm	2.2 \pm 36.3	2.4 \pm 34.9	1.6 \pm 40.4	0.898
Major curve apex deviation, mm	-2.1 \pm 34.5	-1.8 \pm 34.6	-3.3 \pm 34.4	0.804
Major curve Cobb angle	-4.0 \pm 34.4	-3.8 \pm 34.2	-4.6 \pm 35.0	0.885
Small curve apex deviation, mm	0.3 \pm 18.0	-0.4 \pm 17.5	2.6 \pm 19.3	0.34
Small curve Cobb angle	3.6 \pm 24.7	3.4 \pm 24.4	4.2 \pm 25.8	0.848
T1 tilt	0.8 \pm 6.3	1.1 \pm 6.4	-0.4 \pm 6.1	0.156
Thoracic curve apex deviation, mm	-0.3 \pm 10.2	-0.0 \pm 9.5	-1.2 \pm 12.3	0.585
Thoracic curve Cobb angle	-0.9 \pm 16.0	-0.9 \pm 16.1	-1.0 \pm 15.4	0.979
Sagittal measurements				
C2 slope	18.1 \pm 13.8	17.9 \pm 13.0	18.8 \pm 15.9	0.688
CL	9.2 \pm 15.8	9.3 \pm 16.5	8.6 \pm 13.7	0.783
CPA	28.1 \pm 14.4	26.9 \pm 14.6	31.8 \pm 13.2	0.043
CTPA	2.9 \pm 1.6	3.0 \pm 1.5	2.7 \pm 2.0	0.312
L1-L4 angle	-7.0 \pm 19.0	-8.7 \pm 19.1	-2.0 \pm 17.9	0.033
L1-S1, mm	175.7 \pm 22.0	177.4 \pm 20.7	170.5 \pm 24.7	0.062
L1PA	12.0 \pm 11.5	11.5 \pm 11.7	13.9 \pm 10.7	0.251
L4-S1 angle	-30.7 \pm 15.6	-30.7 \pm 14.4	-30.7 \pm 18.8	0.995
LL	32.4 \pm 24.8	34.2 \pm 22.9	26.9 \pm 29.0	0.076
PI	53.5 \pm 16.7	53.4 \pm 16.5	53.9 \pm 17.0	0.847
PI-LL	20.4 \pm 22.8	19.2 \pm 23.2	24.0 \pm 21.2	0.208
PT	25.5 \pm 12.0	24.6 \pm 12.4	28.0 \pm 10.2	0.093
SS	28.0 \pm 14.1	28.7 \pm 13.7	25.9 \pm 14.8	0.225
C2-C7 cSVA, mm	29.2 \pm 14.7	29.2 \pm 13.4	29.1 \pm 18.1	0.965
C7-S1 SVA, mm	70.6 \pm 69.3	65.1 \pm 66.7	86.9 \pm 74.2	0.061
T1SPI	-0.3 \pm 6.8	-0.8 \pm 6.3	1.1 \pm 8.1	0.099
T1 slope	27.7 \pm 11.0	27.6 \pm 10.9	28.1 \pm 11.2	0.771
T1-CL	18.6 \pm 13.3	18.3 \pm 12.7	19.5 \pm 15.1	0.572
T1-L1, mm	306.9 \pm 34.5	306.0 \pm 35.0	309.6 \pm 32.6	0.54
T1-S1, mm	483.1 \pm 44.0	483.4 \pm 44.9	482.2 \pm 41.1	0.871
T2-T5 angle	9.8 \pm 10.2	10.5 \pm 10.0	7.6 \pm 10.4	0.085
T5-T12 angle	26.5 \pm 18.4	26.2 \pm 18.3	27.6 \pm 18.5	0.654
T9SPI	10.0 \pm 7.4	10.0 \pm 7.2	9.9 \pm 8.2	0.925
TK	30.7 \pm 19.7	30.1 \pm 19.3	32.5 \pm 20.8	0.469
TL	11.8 \pm 18.1	10.1 \pm 18.3	16.9 \pm 16.7	0.026
TPA	25.2 \pm 14.3	23.8 \pm 14.6	29.1 \pm 12.7	0.028

SD: Standard deviation; C7PL: C7 plumb line; CL: cervical lordosis; CPA: C2 pelvic angle; CTPA: cervico-thoracic pelvic angle; L1PA: L1 pelvic angle; LL: lumbar lordosis; PI: pelvic incidence; PT: pelvic tilt; SS: sacral slope; cSVA: cervical sagittal vertical axis; SPI: spinopelvic inclination; T1-CL: T1 slope - cervical lordosis; TK: thoracic kyphosis; TL: thoracolumbar alignment; TPA: T1 pelvic angle.

technique highlights regions of the raw image that were important for classification (“hotspots”, [Figure 1](#)). The MRI hotspots were then qualitatively analyzed across the cohort to infer important tissue types for accurate prediction of pseudarthrosis.

RESULTS

Demographic variables poorly correlate with pseudarthrosis

In our cohort of 191 patients who underwent ASD surgery, 48 (25.1%) had pseudarthrosis compared to 143

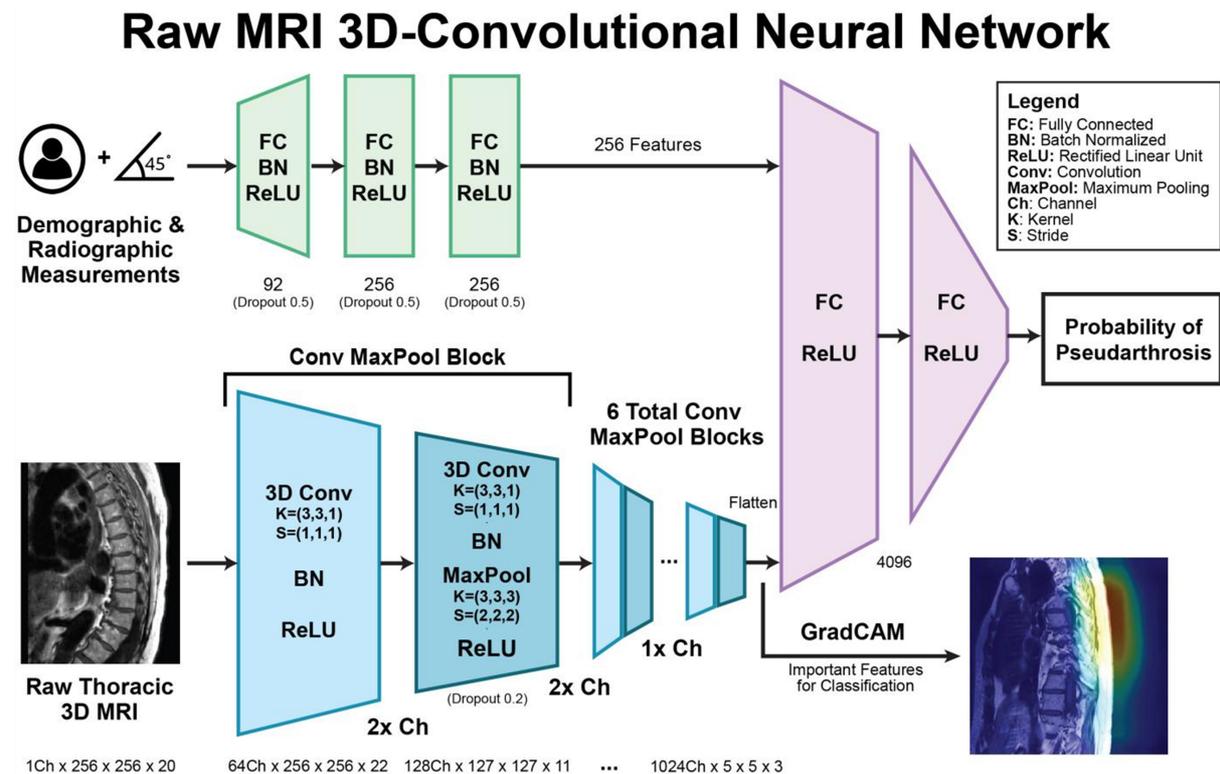


Figure 1. 3D-CNN for pseudarthrosis classification on raw thoracic MRI. 3D-CNN: Three-dimensional convolutional neural network; MRI: magnetic resonance imaging.

(74.9%) who did not, and 37 (19.4%) overall developed rod fracture - consistent with previously reported cohorts^[22]. Specifically, rod fractures were seen in 29 (60.4%) of the pseudarthrosis. All pseudarthrosis required reoperation. Upon demographic analysis, we only observed age to be significantly different between patients who were diagnosed with pseudarthrosis (69.9 ± 10.1 years old) versus those without (60.9 ± 19.9), with a P -value of 0.003. Furthermore, the radiographic variables captured by Surgimap were not observed to be significantly different between the groups after multiple comparison corrections. Please see [Tables 1](#) and [2](#) for the full extent of preoperative variables considered. Thus, in alignment with past literature, the occurrence of pseudarthrosis did not exhibit a distinct demographic or traditional radiographic signature.

Pseudarthrosis can be predicted with raw preoperative MRI

To explore the potential of AI to non-linearly utilize the millions of data points present in raw MRI, we implemented a custom 3D-CNN. Across the five-fold nested cross-validation, the completely withheld testing data were classified by the model with a Youden index ranging from 0.30 to 0.80 (mean 0.49, 95% confidence interval \pm 0.25, [Figure 2](#)). A single population t -test against a null hypothesis of a Youden index of 0.00, representing an equivocal model, was significant with a P -value of $5.50e-3$. These results indicate that the 3D-CNN model was accurate in predicting pseudarthrosis following ASD surgery with at least two years of follow-up and has the potential to generalize well to a larger population.

Superficial adipose tissue appears to be most important for classification

Of greatest interest to this work were the MRI features used by the 3D-CNN model to gain insight into the pathogenesis of pseudarthrosis. Upon model interrogation with GradCAM, the most important MRI features for classification of pseudarthrosis following ASD surgery appear to be posterior adipose tissue -

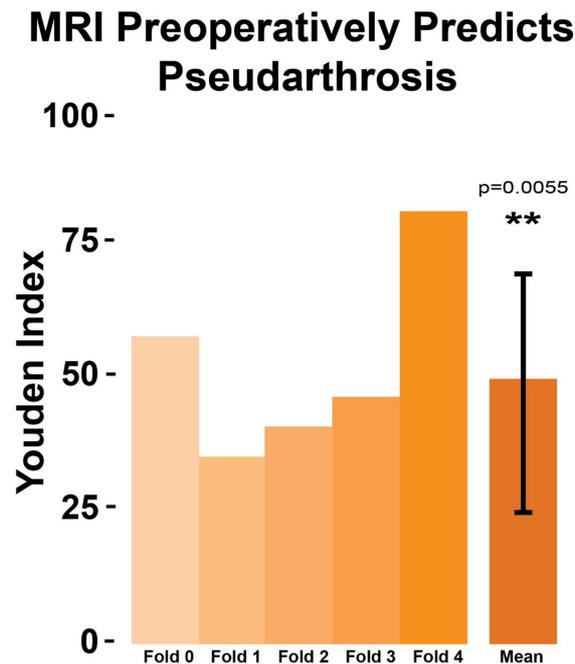


Figure 2. Youden index across the five-fold nested cross-validation. The values shown are only for the completely withheld testing partition for each fold. ** $P < 0.01$.

i.e., the majority of GradCAM feature maps highlight various aspects of superficial adipose tissue posterior to the spinous processes (example subjects in [Figure 3](#)). Notably, there is no significant difference in body mass index (BMI) between the pseudarthrosis cohort (28.8 ± 7.3) and the non-pseudarthrosis cohort (29.0 ± 8.2), with a t -test P -value of 0.874. Notably, of the 48 patients who developed pseudarthrosis, 24 (50.0%) also had proximal junctional kyphosis (PJK). However, in our past work, we found that posterior musculature was most predictive of PJK^[18]. Thus, these results indicate that there is an important radiologic signature within these adipose regions that enables the 3D-CNN model to accurately classify pseudarthrosis, independent of total adipose content estimated by BMI and independent of radiographic features that predict PJK.

DISCUSSION

The current study demonstrated the accuracy of using a 3D-CNN on raw thoracic MRI to predict pseudarthrosis following ASD surgery. More importantly, the imaging features associated with pseudarthrosis were elucidated to be mainly posterior adipose tissue - with a predominance of the upper thoracic region. Interestingly, except for age, our cohort did not demonstrate any demographic or traditional radiographic measurement difference between those who developed pseudarthrosis and those who did not. Thus, it is noteworthy that the 3D-CNN heavily utilized adipose tissue of the imaging to develop the classification despite the pseudarthrosis cohort not being significantly more overweight ($P = 0.874$). This observation leads the authors to surmise that there exists a subtle MRI signature in the adipose tissue that the model used for classification. Furthermore, the imaging hotspots are not consistently at a region of the largest adipose collection; thus, it is likely that the 3D-CNN model is detecting an intra-adipose or adipose-adjacent signal. Future work could focus on using image segmentation techniques to better quantify the exact types of tissue present within GradCAM hotspots. Finally, posterior upper thoracic adipose tissue is typically distant from the region of pseudarthrosis, which in our cohort was predominately in the low lumbar region. Thus, it can be surmised that the network learned a global signature of

MRI Features that Predict Pseudarthrosis

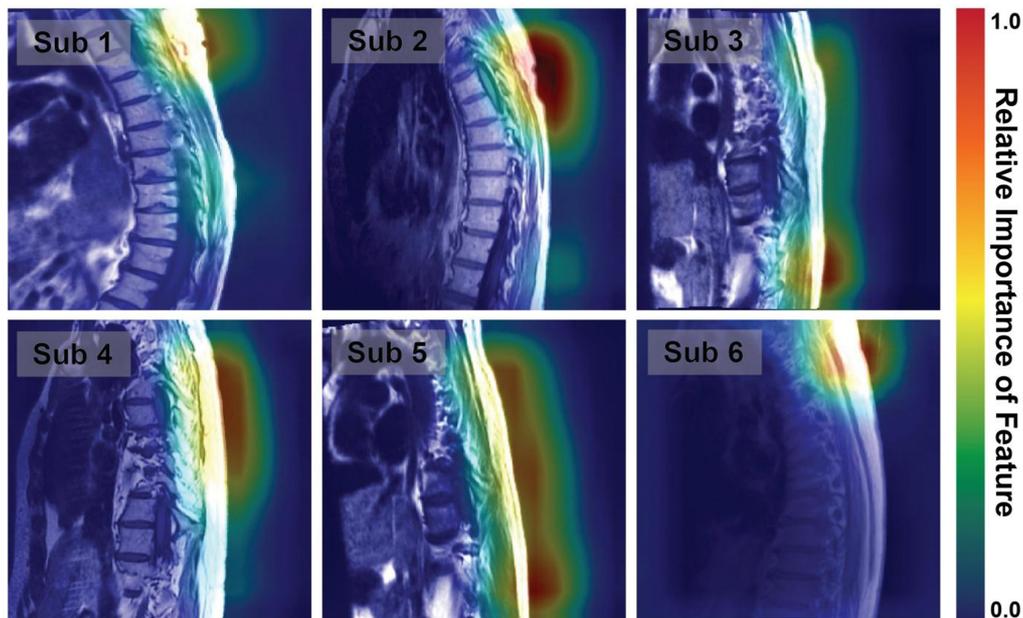


Figure 3. Posterior superficial adipose tissue appears to be important for the 3D-CNN model to classify pseudarthrosis. 3D-CNN: Three-dimensional convolutional neural network.

pseudarthrosis predilection, not captured by demographic variables, as opposed to a local tissue abnormality at the future site of pseudarthrosis.

Overall, this work serves as an augmentation to existing models to preoperatively predict pseudarthrosis^[17]. Previously identified risk factors to predict pseudarthrosis include alcoholism, smoking, fusion location, vertebral bone quality score, diabetes, sarcopenia, advanced age, and potentially graft material^[1,23-26]. Not all of these factors were readily available in our dataset. However, it is notable that our demographic analysis of available metrics did not reveal any significant differences between pseudarthrosis and non-pseudarthrosis cohorts except for age. Thus, by incorporating raw MRI, one can leverage high-dimensional data that are often readily available in clinical databases with minimal manual extraction required. With the incorporation of readily available imaging, this technique does not rely on meticulous database mining and is free from the problems of patient bias when reporting metrics such as alcohol use and smoking status. Furthermore, this work provides potential insight into the biological underpinnings of pseudarthrosis development and could serve as background evidence for future studies exploring the role of global adipose tissue characteristics in those who develop pseudarthrosis.

It is well-documented that posterior musculature and sarcopenia appear to be strongly correlated with mechanical complications following ASD surgery^[27]. Thus, it is notable that the current work, and our past work using similar techniques to predict PJK using MRI have both revealed imaging risk factors of soft tissue - adipose tissue and posterior musculature, respectively. This contributes to the growing body of literature focusing on soft tissue characteristics as driving factors for mechanical complications, as opposed to bony anatomy. The authors do not discount the importance of bone integrity for consideration of ASD surgery, but rather aim to outline the additional importance of soft tissue health when considering a large deformity operation.

Limitations to this technique include the ever-present potential for overfitting, the computing resources required, and the technical expertise required to run the analysis algorithms. An important next step would be to test these methods on an external cohort. Another consideration with this methodology is that the thoracic MRI did not capture the top of the implanted construct in a few subjects. This can be seen as both a potential weakness and potential strength of this study because the results were robust despite this consideration - this indicates that there is possibly a global imaging feature that the 3D-CNN detects to aid accurate classification. Finally, the proper de-identification of raw data is paramount to model creation to ensure patient privacy when deploying trained models.

Overall, the use of machine learning in medical imaging has garnered attention but has still been limited in scope compared to tabular data machine learning and large language models. We aimed to demonstrate the potential of a simple classification scheme on available 3D MRIs to predict the development of pseudarthrosis following ASD surgery. Beyond the cross-validated accuracy of the model, our approach has the benefit of providing a level of interpretation by outlining imaging features used by the model to make classification decisions. Overall, this work demonstrates the capabilities of raw imaging AI in spine surgery and can serve as the basis for a deeper biological inquiry into the pathogenesis of pseudarthrosis.

DECLARATIONS

Authors' contributions

Conceptualization, data acquisition, data analysis, results interpretation, and manuscript preparation: Johnson GW, Chanbour H

Conceptualization, data analysis, results interpretation, and manuscript preparation: Doss DJ

Conceptualization, results interpretation, and manuscript preparation: Makhoul GS

Conceptualization, data acquisition, data analysis, results interpretation, and manuscript preparation: Abtahi AM, Stephens BF, Zuckerman SL

Availability of data and materials

All data and code are available upon reasonable request to the corresponding author.

Financial support and sponsorship

None.

Conflicts of interest

Stephens BF receives institutional research support from Nuvasive and Stryker Spine. Zuckerman SL reports being an unaffiliated neurotrauma consultant for the National Football League. Abtahi AM receives institutional research support from Stryker Spine. The other authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. IRB: Approval Attained (#211290).

Consent for publication

Not applicable.

Copyright

© The Author(s) 2024.

REFERENCES

1. How NE, Street JT, Dvorak MF, et al. Pseudarthrosis in adult and pediatric spinal deformity surgery: a systematic review of the literature and meta-analysis of incidence, characteristics, and risk factors. *Neurosurg Rev* 2019;42:319-36. DOI PubMed
2. Deyo RA, Mirza SK, Martin BI, Kreuter W, Goodman DC, Jarvik JG. Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA* 2010;303:1259-65. DOI PubMed PMC
3. Buchholz AL, Quinn JC, Shaffrey CI. Postoperative spinal deformities: kyphosis, nonunion, and loss of motion segment. In: *Complications in neurosurgery*. 2019. pp. 325-30. Available from: <http://103.203.175.90:81/fdScript/RootOfEBooks/E%20Book%20Collection%202021%20-%20A/ENGLISH/Complications%20in%20Neurosurgery.pdf>. [Last accessed on 7 Nov 2024].
4. Kim YJ, Bridwell KH, Lenke LG, Rhim S, Cheh G. Pseudarthrosis in long adult spinal deformity instrumentation and fusion to the sacrum: prevalence and risk factor analysis of 144 cases. *Spine* 2006;31:2329-36. DOI PubMed
5. Kim YJ, Bridwell KH, Lenke LG, Rinella AS, Edwards C 2nd. Pseudarthrosis in primary fusions for adult idiopathic scoliosis: incidence, risk factors, and outcome analysis. *Spine* 2005;30:468-74. DOI PubMed
6. Kim YJ, Bridwell KH, Lenke LG, Cho KJ, Edwards CC 2nd, Rinella AS. Pseudarthrosis in adult spinal deformity following multisegmental instrumentation and arthrodesis. *J Bone Joint Surg Am* 2006;88:721-8. DOI PubMed
7. Pateder DB, Park YS, Kebaish KM, et al. Spinal fusion after revision surgery for pseudarthrosis in adult scoliosis. *Spine* 2006;31:E314-9. DOI PubMed
8. Joshi RS, Haddad AF, Lau D, Ames CP. Artificial intelligence for adult spinal deformity. *Neurospine* 2019;16:686-94. DOI PubMed PMC
9. Joshi RS, Lau D, Ames CP. Artificial intelligence for adult spinal deformity: current state and future directions. *Spine J* 2021;21:1626-34. DOI PubMed
10. Lenke LG. Commentary: artificial intelligence for adult spinal deformity. *Neurospine* 2019;16:695-6. DOI PubMed PMC
11. Perez-Breva L, Shin JH. Artificial intelligence in neurosurgery: a comment on the possibilities. *Neurospine* 2019;16:640-2. DOI PubMed PMC
12. Durand WM, Lafage R, Hamilton DK, et al; International Spine Study Group (ISSG). Artificial intelligence clustering of adult spinal deformity sagittal plane morphology predicts surgical characteristics, alignment, and outcomes. *Eur Spine J* 2021;30:2157-66. DOI PubMed
13. Durand WM, Daniels AH, Hamilton DK, et al; International Spine Study Group. Artificial intelligence models predict operative versus nonoperative management of patients with adult spinal deformity with 86% accuracy. *World Neurosurg* 2020;141:e239-53. DOI PubMed
14. Scheer JK, Smith JS, Schwab F, et al; International Spine Study Group. Development of a preoperative predictive model for major complications following adult spinal deformity surgery. *J Neurosurg Spine* 2017;26:736-43. DOI PubMed
15. Scheer JK, Osorio JA, Smith JS, et al; International Spine Study Group. Development of validated computer-based preoperative predictive model for proximal junction failure (PJF) or clinically significant PJK with 86% accuracy based on 510 ASD patients with 2-year follow-up. *Spine* 2016;41:E1328-35. DOI PubMed
16. Pellisé F, Serra-Burriel M, Smith JS, et al; International Spine Study Group, European Spine Study Group. Development and validation of risk stratification models for adult spinal deformity surgery. *J Neurosurg Spine* 2019;31:587-99. DOI PubMed
17. Scheer JK, Oh T, Smith JS, et al; International Spine Study Group. Development of a validated computer-based preoperative predictive model for pseudarthrosis with 91% accuracy in 336 adult spinal deformity patients. *Neurosurg Focus* 2018;45:E11. DOI PubMed
18. Johnson GW, Chanbour H, Ali MA, et al. Artificial intelligence to preoperatively predict proximal junction kyphosis following adult spinal deformity surgery: soft tissue imaging may be necessary for accurate models. *Spine* 2023;48:1688-95. DOI PubMed PMC
19. Kernbach JM, Staartjes VE. Foundations of machine learning-based clinical prediction modeling: part II - generalization and overfitting. In: Staartjes VE, Regli L, Serra C, editors. *Machine learning in clinical neuroscience*. Springer, Cham; 2022. pp. 15-21. DOI
20. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32-5. DOI
21. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis* 2020;128:336-59. DOI
22. Adogwa O, Buchowski JM, Lenke LG, et al. Comparison of rod fracture rates in long spinal deformity constructs after transforaminal versus anterior lumbar interbody fusions: a single-institution analysis. *J Neurosurg Spine* 2020;32:42-9. DOI PubMed
23. Passias PG, Bortz C, Alas H, et al. Alcoholism as a predictor for pseudarthrosis in primary spine fusion: an analysis of risk factors and 30-day outcomes for 52,402 patients from 2005 to 2013. *J Orthop* 2019;16:36-40. DOI PubMed PMC
24. Marques MF, Fiere V, Obeid I, et al; Société Française de Chirurgie Rachidienne, SFCR. Pseudarthrosis in adult spine deformity surgery: risk factors and treatment options. *Eur Spine J* 2021;30:3225-32. DOI PubMed
25. Kawabata A, Yoshii T, Sakai K, et al. Identification of predictive factors for mechanical complications after adult spinal deformity surgery: a multi-institutional retrospective study. *Spine* 2020;45:1185-92. DOI PubMed
26. Kuo CC, Soliman MAR, Aguirre AO, et al. Vertebral bone quality score independently predicts proximal junctional kyphosis and/or

- failure after adult spinal deformity surgery. *Neurosurgery* 2023;92:945-54. DOI PubMed
27. Akbik OS, Al-Adli N, Pernik MN, et al. A comparative analysis of frailty, disability, and sarcopenia with patient characteristics and outcomes in adult spinal deformity surgery. *Global Spine J* 2023;13:2345-56. DOI PubMed PMC

Meta-Analysis

Open Access



Deep learning for automated spinopelvic parameter measurement from radiographs: a meta-analysis

Dylan Glaser¹, Ahmad K. AIMekawi², James P. Caruso³, Candace Y. Chung⁴, Eshal Z. Khan¹, Hicham M. Daadaa⁵, Salah G. Aoun³, Carlos A. Bagley² 

¹Department of Neurosurgery, The university of Missouri-Kansas City, School of Medicine, Kansas City, MO 64108, USA.

²Department of Neurosurgery, Saint Luke's Hospital, Kansas City, MO 64111, USA.

³Department of Neurosurgery, The University of Texas Southwestern, Dallas, TX 75235, USA.

⁴Department of Neurosurgery, College of Osteopathic Medicine, Kansas City University, Kansas City, MO 64106, USA.

⁵Department of Hematology, King's College Hospital NHS Foundation Trust, London SE5 9RS, UK.

Correspondence to: Dr. Carlos A. Bagley, Saint Luke's Marion Bloch Neuroscience Institute, Department of Neurosurgery, Saint Luke's Hospital, 4401 Wornall Rd., Kansas City, MO 64111, USA. E-mail: cabagley@saint-lukes.org

How to cite this article: Glaser D, AIMekawi AK, Caruso JP, Chung CY, Khan EZ, Daadaa HM, Aoun SG, Bagley CA. Deep learning for automated spinopelvic parameter measurement from radiographs: a meta-analysis. *Art Int Surg.* 2025;5:1-15. <https://dx.doi.org/10.20517/ais.2024.36>

Received: 31 May 2024 **First Decision:** 14 Oct 2024 **Revised:** 26 Nov 2024 **Accepted:** 6 Dec 2024 **Published:** 4 Jan 2025

Academic Editors: Eyad Elyan, Andrew Gumbs **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Aim: Quantitative measurement of spinopelvic parameters from radiographs is important for assessing spinal disorders but is limited by the subjectivity and inefficiency of manual techniques. Deep learning may enable automated measurement with accuracy rivaling human readers.

Methods: PubMed, Embase, Scopus, and Cochrane databases were searched for relevant studies. Eligible studies were published in English, used deep learning for automated spinopelvic measurement from radiographs, and reported performance against human raters. Mean absolute errors and correlation coefficients were pooled in a meta-analysis.

Results: Fifteen studies analyzing over 10,000 radiographs met the inclusion criteria, employing convolutional neural networks (CNNs) and other deep learning architectures. Pooled mean absolute errors were 4.3° [95% confidence interval (CI) 3.2-5.4] for Cobb angle, 3.9° (95%CI 2.7-5.1) for thoracic kyphosis, 3.6° (95%CI 2.8-4.4) for lumbar lordosis, 1.9° (95%CI 1.3-2.5) for pelvic tilt (PT), 4.1° (95%CI 2.7-5.5) for pelvic incidence (PI), and 1.3 cm (95%CI 0.9-1.7) for sagittal vertical axis (SVA). Intraclass correlation coefficients exceeded 0.81, indicating strong agreement between automated and manual measurements.



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



Conclusion: Deep learning demonstrates promising accuracy for automated spinopelvic measurement, potentially rivaling experienced human readers. However, further optimization and rigorous multicenter validation are required before clinical implementation. These technologies may eventually improve the efficiency and reliability of quantitative spine image analysis.

Keywords: Deep learning, spine parameters, pelvic parameters

INTRODUCTION

Quantitative evaluation of spine and pelvis anatomy has long interested clinicians and researchers in fields such as orthopedics, neurosurgery, and radiology. Assessing sagittal spinal balance - the geometric relationships between spinal curves and pelvic parameters - is considered essential for understanding normal posture and alignment^[1]. Sagittal balance encompasses important radiographic measures such as cervical and lumbar lordosis, thoracic kyphosis, pelvic tilt (PT), pelvic incidence (PI), and sacral slope (SS)^[2,3]. Abnormal spinopelvic alignment has been associated with pain, disability, and poor health outcomes^[4].

Traditionally, spinopelvic parameters were manually measured from plain radiographs using techniques like the Cobb method, with known limitations in accuracy and objectivity^[5]. Computer-assisted analysis tools later emerged to potentially improve measurement consistency, though substantial human input was still required^[6]. Deep learning has rapidly advanced in recent years but traces its origins back decades. The concepts of neural networks were initially developed in the 1950s and 60s. However, computational power limited applications. In the 1980s and 90s, techniques like convolutional neural networks (CNNs) were pioneered, laying the groundwork for modern deep learning. Major advancements in computing, along with the availability of large datasets, then enabled deep neural networks to surpass previous benchmarks across diverse tasks. Beginning in the 2010s, deep learning achieved remarkable performance in computer vision, natural language processing, and medical imaging analysis. The latest methods like CNNs now offer transformative opportunities to extract information from complex data. Over the past decade, advances in artificial intelligence and machine learning have enabled more automated approaches for quantitative radiology and medical imaging^[7,8].

Machine learning utilizes statistical models trained on known data to recognize patterns in new data^[9]. Deep learning is a subset of machine learning based on layered neural networks that can automatically learn optimal features directly from raw data, unlike traditional techniques requiring hand-crafted feature engineering^[10]. The latest deep learning methods have become integral for the automated analysis of medical images across specialties^[11,12], including quantitative characterization of spine disorders from radiographs and CT scans^[13,14].

Several recent studies have applied deep CNNs for automated measurement of key spinopelvic parameters from standard radiographs^[15]. Reported accuracy has been promising but varies widely across studies. However, a comprehensive synthesis of the latest achievements, methodological innovations, and measured performance has been lacking. This review aims to systematically summarize and critically appraise the existing literature on deep learning-based assessment of sagittal spinopelvic alignment on radiographs. It elucidates the current state of the field and future directions to potentially improve clinical adoption.

METHODS

This Meta-analysis was conducted according to the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines^[16] [Supplementary Table 1].

Search strategy

A comprehensive literature search was conducted using PubMed, Embase, Scopus, and Cochrane databases from inception to December 2023 to identify relevant studies. The search strategy included a combination of controlled vocabulary terms (e.g., MeSH) and keywords related to “artificial intelligence”, “deep learning”, “convolutional neural network”, “spine”, “spinopelvic parameters”, and related terms. Reference lists of included articles and relevant systematic reviews were hand-searched to identify any additional eligible studies.

Study selection

Studies were included if they met the following criteria: (1) published in English language peer-reviewed journals; (2) used deep learning models including CNNs to automatically estimate spinopelvic parameters from radiographs (X-ray); (3) reported model performance metrics compared to human rater measurements including mean absolute error and correlation coefficient. Conference abstracts, case reports, editorials, and non-peer reviewed articles were excluded.

Two reviewers (A.K.M and J.C) independently screened the titles, abstracts, and full texts of retrieved records against the eligibility criteria. Disagreements were resolved by consensus or consultation with a third reviewer if needed. The study selection process was documented using a PRISMA flow diagram [Figure 1].

Data extraction

A standardized data extraction form was created and pilot-tested on a subset of included studies. Two reviewers (A.K.M and J.C) then independently extracted data from the full set of included studies. Extracted information included: first author name, publication year, dataset details (number of images, resolution, pathology), imaging modality, model details, spinopelvic parameters analyzed (Accuracy Metrics), deep learning model details including architecture and training approach, mean absolute error, correlation coefficient, batch size, number of epochs, any additional reported performance metrics, computational efficiency, validation approach, and any key limitation. Any discrepancies in extracted data were resolved through discussion and mutual consensus. Additionally, studies focusing specifically on lumbosacral transitional vertebrae (LSTV) were excluded to maintain homogeneity in the analysis. While LSTV can significantly impact spinopelvic measurements, the unique challenges they present in parameter assessment warrant separate consideration from standard spinopelvic measurements. This exclusion allowed for a more consistent comparison of measurement accuracy across included studies.

Statistical analysis

A random-effects meta-analysis was performed to pool the mean absolute errors reported by the included studies for each spinopelvic parameter. The inverse variance method was used to calculate the weighted mean differences and 95% confidence intervals (CIs). Heterogeneity among the studies was assessed using the I^2 statistic, which represents the percentage of total variation across studies due to heterogeneity rather than chance. An I^2 value of 0% indicates no observed heterogeneity, while larger values indicate increasing heterogeneity. The pooled estimates and their 95% CIs were graphically presented using forest plots. All statistical analyses were conducted using R software (version 4.0.3) with the “meta” package (version 4.15-1).

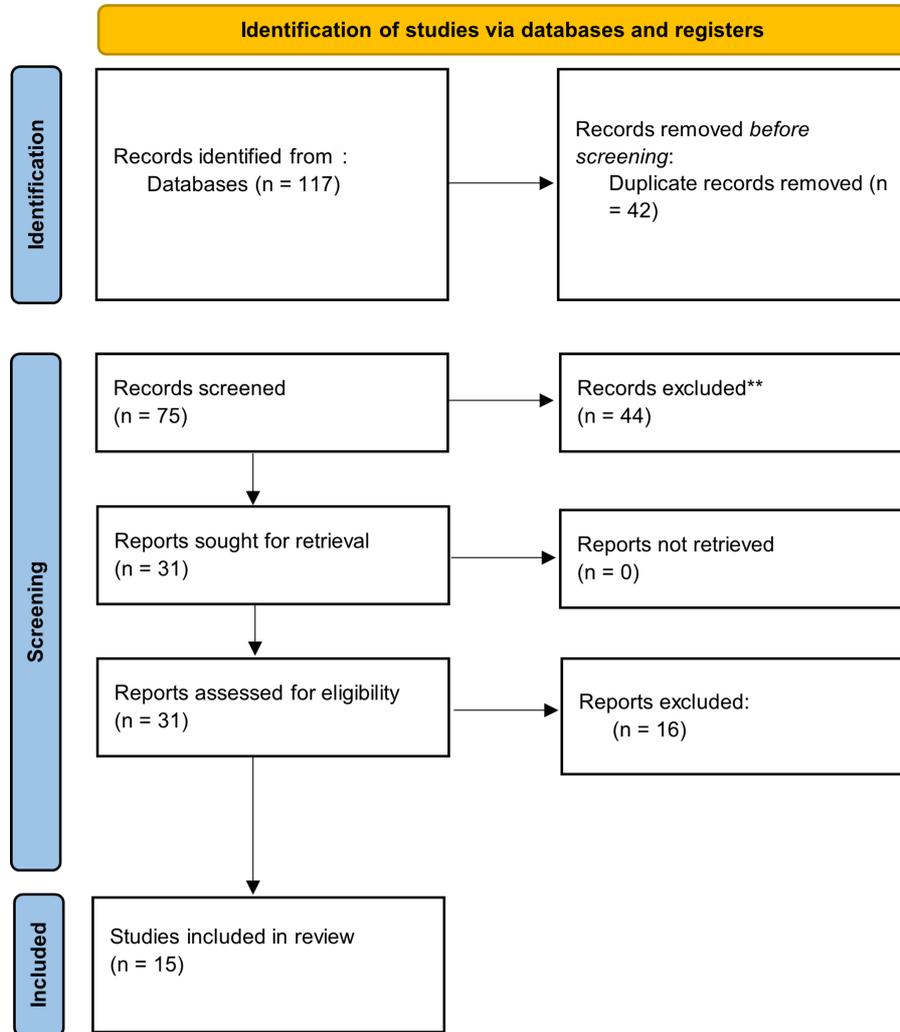


Figure 1. PRISMA flow chart. PRISMA: Preferred reporting items for systematic reviews and meta-analyses.

Quality assessment

The quality and risk of bias of included studies were assessed using the IJEMDI tool tailored specifically for diagnostic accuracy studies. Studies were evaluated across four domains: (1) clarity in the description of the research question, study objectives, and replicability of the study design; (2) availability of an open dataset or detailed instructions for data access; (3) comprehensive documentation of methods, including software details and statistical approaches, to facilitate replication; and (4) whether the results supported the conclusions, limitations were discussed, and conflicts of interest were disclosed. Each domain was rated as either present (2), absent (0), or unclear (1)^[17].

RESULTS

A total of 14 studies published between 2018-2023 were included in this systematic review, encompassing 10,727 subjects^[18-31]. The studies utilized various imaging modalities to develop and validate automated methods for measuring spinal alignment, including lateral X-rays^[18-31], biplanar radiographs^[30,31], and CT scans^[26]. Both preoperative and postoperative images were employed, with 6 studies incorporating cases with spinal implants^[19,20,23,24,28] to evaluate performance in surgically altered anatomy. The diversity of imaging captures numerous clinically relevant scenarios, although multicenter external validation was

lacking, with most datasets from single institutions [Table 1].

A range of deep learning models were applied for automated spinal measurement, including custom CNNs^[18-21,23-25,27-31], multi-view correlation networks^[19,20], and segmentation-based approaches^[23,24,28-30]. For Cobb angle measurement, mean absolute errors ranged from 1.2° to 7.81°^[18-21,23,27,28], with most studies achieving errors $\leq 5^\circ$. Similar trends were observed for other sagittal parameters, such as thoracic kyphosis, lumbar lordosis, and PI^[18,19,22-25,29,30]. Intraclass correlation coefficients between automated and manual measurements exceeded 0.75, indicating strong agreement^[22,25,26]. Computational efficiency was reported in several studies, with inference times ranging from 0.2 to 75 s per image^[22,23,27,28], demonstrating the potential for accelerated analysis compared to manual methods.

Cobb angle demonstrated a pooled mean error of 4.3° (95%CI: 3.2°-5.4°). Thoracic kyphosis and lumbar lordosis showed similar pooled errors of 3.9° (95%CI: 2.7°-5.1°) and 3.6° (95%CI: 2.8°-4.4°), respectively. PT had the lowest pooled error at 1.9° (95%CI: 1.3°-2.5°), while PI exhibited a slightly higher pooled error of 4.1° (95%CI: 2.7°-5.5°). Sagittal vertical axis (SVA) demonstrated a pooled mean error of 1.3 cm (95%CI: 0.9-1.7 cm). These results highlight the overall accuracy of deep learning models in automatically measuring key spinopelvic parameters from radiographic images [Figure 2].

Manual measurement of spinopelvic parameters has shown inter-observer variability ranging from 5° to 10° for Cobb angle measurements and similar ranges for other parameters in previous studies. The pooled AI measurement errors we found (4.3° for Cobb angle, 3.9° for thoracic kyphosis, and 3.6° for lumbar lordosis) demonstrate comparable or better accuracy than manual measurements while offering significantly improved efficiency.

Quality assessment

Utilizing the IJEMDI checklist, the papers address most checklist items sufficiently but have room for improvement around enabling replicability and providing more method/software specifics. Conflicts of interest and limitations also remain inconsistently addressed [Supplementary Table 2].

DISCUSSION

This systematic review and meta-analysis demonstrate the potential of deep learning for the automated measurement of spinopelvic parameters from radiographs. The comprehensive literature search identified 14 eligible studies between 2018-2023, analyzing over 10,000 radiographs with deep CNNs and other architectures^[18-31]. The studies utilized various imaging sources to develop and validate automated methods for measuring spinal alignment, including lateral X-rays^[19,20,23-25,27,32-34], biplanar radiographs^[31,35], and CT scans^[36]. Both preoperative^[19,23,25,27,31-33,35,36] and postoperative^[19,20,24] images were employed, with 6 studies incorporating cases with spinal implants^[19,20,23,24,33,36] to evaluate performance in surgically altered anatomy. The diversity of imaging captures numerous clinically relevant scenarios. However, multicenter external validation was lacking, with most datasets from single institutions. Aspects like vendor variability could impact segmentation. Model development must be capable of analyzing all imaging protocols for translation.

A range of model types were applied for automated spinal measurement, from conventional machine learning^[21,25] and rule-based systems^[31] to modern deep CNNs^[19,20,23-25,31,33-37]. Details for replication varied extensively - 4 studies provided no specific model details^[19,25,27,34], while 5 gave networks and parameters^[19,20,25,31,35]. Public code/data availability remains limited. Custom architectures were common for direct spinal measurement^[19,33-35,37], rather than off-the-shelf models. Multi-task^[25,33,37], multi-view^[19,33], and

Table 1. Main table describing study characteristics

Paper	Dataset details	Imaging	Model details	Mean absolute error	Correlation coefficient	Accuracy metrics	Comparison with other methods	Key limitations	Validation approach	Neural network architecture	Batch size	No. of epochs
Chae et al. 2020 ^[18]	Training - 400; resolution - 3,240 × 1,080 pixels; variety - 57% normal spine, 20% lumbar lordosis, 24% thoracic kyphosis	X-ray	Decentralized CNN; multiple orders	1.45°-3.52°	NA	Mean absolute error: 1.45°-3.52° for parameters	Compared to manual measurement by experienced surgeons, as well as regression CNN model	Requires multiple ordered datasets, training time; limited diversity	40 test radiographs; comparison to manual measurements by experienced surgeons	Custom decentralized CNN	NA	Initial: 0.001, SGDM momentum 0.95
Wu et al. 2018 ^[19]	526 (154 patients); resolution: 128 × 256 pixels	X-ray	Custom MVC-Net	Landmark: 0.0398-0.0459; Cobb: 4.04°-4.07°	NA	Mean absolute error (landmark): 0.0398 (AP) - 0.0459 (LAT); circular mean absolute error (Cobb angle): 4.04° (AP) - 4.07° (LAT)	Compared to manual measurement and other deep learning methods	Single clinic dataset; no metal artifact images	10-fold patient-wise cross-validation; comparison to manual "gold standard"	Custom MVC-Net	100	Starting: 0.01, halved every 10 epochs
Wang et al. 2019 ^[20]	526; resolution: 0.26 mm/pixel	X-ray	Custom MVE-Net	Cobb: 6.26°-7.81°	NA	Circular mean absolute error (Cobb angle): 7.81° (AP) - 6.26° (LAT); SMAPE (Cobb angle): 24.94% (AP) - 11.90% (LAT)	Compared to manual measurement and other deep learning methods	Single clinic dataset	Used same dataset as previous study; compared to other deep learning methods	Custom MVE-Net	NA	Starting: 0.01
Zhang et al. 2022 ^[21]	2,738 pairs (AP & LAT X-rays); from local hospital	X-ray	Custom MPF-Net	Landmark: 0.0046-0.0050; Cobb: 3.52°-4.05°	NA	Scaled mean absolute error (landmark): 0.0046 (AP) - 0.0050 (LAT); circular mean absolute error (Cobb angle): 3.52° (AP) - 4.05° (LAT); SMAPE (Cobb angle): 13.71% (AP) - 12.60% (LAT)	Compared to manual measurement and other deep learning methods	Single clinic dataset	10-fold cross-validation; comparison to manual "gold standard" measurements	Custom MPF-Net	120	Initial: 0.001, decayed by 0.2 every 30 epochs
Zerouali et al. 2023 ^[22]	100 patients with coronal & sagittal whole spine radiographs	X-ray	SmartXpert (Milvue)	≤ 2.9° or ≤ 2.7 mm	≥ 0.85 except thoracic kyphosis = 0.58	Mean absolute error: ≤ 2.9° or ≤ 2.7 mm for parameters; intraclass correlation coefficient: ≥ 0.85 except thoracic kyphosis = 0.58	Compared to measurements by senior musculoskeletal radiologist (ground truth)	Mainly pediatric population, exclusions restricted analysis to preoperative patients	Comparison to "gold standard" manual measurements; visual assessment of reliability by radiologists	NA	NA	NA
Korez et al. 2020 ^[23]	145 images to train model, 97 test images with variety of conditions	X-ray	RetinaNet + U-Net CNNs	1.2°-5.0°	NA	Mean absolute difference vs. manual measurements: 1.2°-5.0° for parameters	Compared to manual measurements by spine surgeon	Single center data; did not evaluate intra-observer	Statistical analysis (mean absolute difference, correlation)	RetinaNet + U-Net	NA	NA

Author (Year)	Study Description	Modality	Method	Accuracy	Precision	Comparison	Limitations	Analysis	Model	Batch Size	Epochs
Kim <i>et al.</i> 2023 ^[24]	1,807 lateral radiographs; variety of spinal conditions	X-ray	Mask R-CNN for vertebral segmentation	0.4°-3.0°	NA	Mean absolute error vs. manual measurements: 0.4°-3.0° for parameters; dice similarity coefficient: 92.6% for segmentation	(reference standard)	variability analysis, etc.) against manual measurements	Mask R-CNN (ResNet 101 backbone)	18 images per batch	36 epochs
Yeh <i>et al.</i> 2021 ^[25]	2,210 whole spine lateral radiographs; variety of spinal conditions	X-ray	Cascaded pyramid network + differentiable spatial to numerical transform layer	Landmark: 1.75-3.39 mm; parameter: 0.1°-6.6°	NA	Median error: 1.75-3.39 mm for landmarks; parameter errors: mean 0.1°-6.6°, median 0.03-5.3°	Compared to measurements by 3 doctors (ground truth)	Single center data; did not include images with vertebral anomalies	Cascaded pyramid net	NA	120 epochs (early stopping applied)
Orosz <i>et al.</i> 2022 ^[26]	600 lateral spine radiographs for training; 200 lumbar spine radiographs (100 pre-op, 100 post-op) for testing	X-ray	CNN for segmentation + U-Net for landmark detection	Not reported	0.75-0.92	Intraclass correlation coefficient between AI and human raters: 0.85-0.92 pre-op, 0.75-0.91 post-op	Compared to measurements by expert human raters	Single-center data for validation; did not assess intra-rater reliability	Convolutional NN + U-Net	NA	NA
Gami <i>et al.</i> 2022 ^[27]	100 images to train model, 130 images to test model	X-ray	YOLO version 3 CNN	Cobb: 1.726°	NA	Average absolute difference - Cobb angle: 1.726°, plumb line: 0.415 cm	Compared to radiographic measurements in cadaver model	Testing only on single cadaver model and artificial templates	YOLOv3 CNN	NA	NA
Schwartz <i>et al.</i> 2021 ^[28]	816 lateral lumbar radiographs including some with instrumentation/hip prostheses	X-ray	MultiResUNet CNN + computer vision pipeline	≤ 4.6°	NA	Mean absolute difference vs. surgeons: ≤ 4.6° for parameters; success rate: 90%-100%	Compared to measurements by 3 orthopedic spine surgeons	10% failure rate for Cobb angle; potential for measurement skew	MultiResUNet	NA	NA
Aubert <i>et al.</i> 2019 ^[29]	68 biplanar radiographs with variety of spinal conditions	X-ray	CNN for anatomical landmark detection to fit statistical spine model	Landmark: 1.6-2.3 mm; parameter: 2.8°-4.7°	NA	Mean error: 1.6-2.3 mm for landmarks; 2.8°-4.7° for spinal parameters; 1°-2.1° for pelvic parameters	Compared to expert supervised reconstructions (ground truth)	Small dataset from single center	CNNs	NA	NA
Nguyen <i>et al.</i> 2022 ^[30]	500 whole spine lateral radiographs with variety of conditions	X-ray	Decentralized CNN	1.156°-6.318°	≥ 0.8 for 10 of 12 parameters	Correlation coefficient: ≥ 0.8 for 10 of 12 parameters; mean absolute error:	Compared to manual measurements by experienced	Difficulty with parameters related to T1 vertebrae;	VGG-net based CNN architecture	Batch size: 32	50 epochs

						1.156°-6.318°	doctors (standard reference)	requires separate datasets for each model order	measurements			
Galbusera <i>et al.</i> 2019 ^[31]	493 biplanar radiographs; variety of spinal disorders and deformities	X-ray	Fully CNN + differentiable spatial to numerical transform layer	Not explicitly reported	NA	Standard error between DL predictions & ground truth: 2.7°-11.5° for parameters	Compared to parameters extracted from sterEOS 3D reconstructions (ground truth)	Limited training dataset size ($n = 443$ image pairs); polynomial interpolation introduced error	50 test cases; statistical analysis (linear regression, Bland-Altman analysis) against ground truth	Fully convolutional network	NA	100 epochs

CNN: Convolutional neural network; NA: not applicable; SGDM: stochastic gradient descent; AP: anteroposterior; LAT: lateral; MVC-Net: multi-view correlation network; MVE-Net: multi-view extrapolation net; MPF-Net: multi-task, proposal correlation, feature fusion network; MAE: mean absolute error; ICC: intraclass correlation coefficient; AI: artificial intelligence; YOLO: You Only Look Once.

vertebral correlation^[25] learning schemes showed benefits for parameter accuracy through inter-relationship modeling, overcoming imaging challenges like occlusion.

Studies assessed accuracy via comparison to expert manual measurement, using metrics such as mean absolute differences (all studies) and voxel overlap measures where segmentation was evaluated^[19,23,24,31,35,36]. For Cobb angle measurement, mean errors ranged from 1.7° to 8.1°, but most CNN methods achieved $\leq 5^\circ$ mean difference^[23-25,31,32,34,35], adequate for clinical usage^[38]. Similar trends were held for other sagittal measurements^[19,20,23,24,31,35]. Notably, Wang *et al.* employed extrapolation methods atop initial estimates to give the best overall accuracies of 6.2°/7.8° Cobb angle errors in lateral/AP views vs. 4.0°/4.1° for MCV-Net^[20,37]. Intraclass coefficients of 0.86-0.99^[19,23-25] confirmed automated/manual measurement agreement.

Comparisons were made to traditional manual measurement^[19,20,23-25,31,35], manual tools^[19,25,27], early machine learning applications^[25], and different iterations of automated algorithms^[19,33]. Automated methods met or exceeded both classic and contemporary techniques. Particular benefits arose in reproducibility, efficiency, and standardization vs. manual approaches prone to subjectivity and variability^[19,23,24]. Deep learning methods showed headroom over alternate automated implementations in accuracy, overcoming limitations such as occlusion. Wang *et al.* achieved better Cobb measurement than MCV-Net^[19] (7.8° lateral error vs. 4.1°), through vertebral correlation and extrapolation augmentations^[20].

Studies cited small datasets^[31], external validity^[19,24,31,35,36], surgical cases^[19,20,23,24,33], implant handling^[33,36], need for inter-rater evaluations^[33], pelvic measurement gaps^[27], follow-up studies^[24], and real-world clinical workflow integration^[24,27] as main limitations. Anonymization, reproducibility, negative societal impacts, and public data availability were generally not addressed. Small samples particularly restricted subgroup analysis - only Gami *et al.* reported metrics by spinal pathology^[27]. Building large heterogeneous benchmark datasets could facilitate model development and address generalizability. Standardized reporting guidelines for spine AI could also benefit the field.

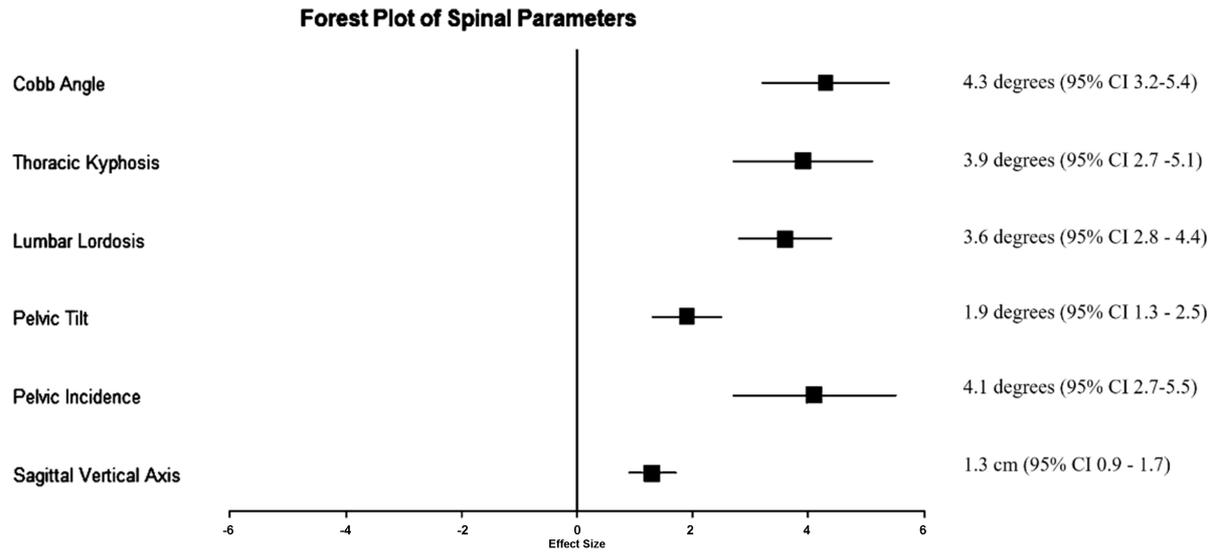


Figure 2. Forest plot showing weight distribution of the different spino-pelvic parameters.

End-to-end runtimes ranged from 2 to 75 s for automated measurement pipelines^[23,24], up to 17× faster than manual analysis; most systems took under 20 s^[19,23,35], adequate for surgical usage. Inference-only times were often sub-second^[23,27]. Accelerated measurement enables more intraoperative images for improved surgical decisions. However, detailed computational profiling was generally lacking, impeding comparisons. Cloud-based implementations could broadly enable these techniques.

Studies used statistical comparisons between automated and manual measurements for validation, incorporating Bland-Altman analysis^[19,23,25,27,31,35], paired significance tests^[19,23,27,35], linear regression^[19,23,25,27,31,35], Pearson correlation coefficients^[19,23,25,27,31,35], and intra-class coefficients^[19,23-25]. Manual measurement reliability was sometimes quantified^[27]. Both preoperative^[19,23-25,27,31,32,35] and postoperative subjects^[19,24,37] were included, although only Kim *et al.* performed validation in distinct pre- and postoperative cohorts^[24]. Most evaluations used held-out testing data from the same institution as model development; multicenter validation was absent. Generalizability beyond the typically homogeneous training populations requires further scrutiny.

CNN backbones ranged from VGG^[19] and U-Nets^[31,36] to ResNets^[24,25,33]. Both feedforward^[19,25] and fully convolutional layouts were used. Custom network engineering was common^[19,23-25,27,31,32,35], given insufficient anatomical representational power in generic classification architectures. Pretraining on natural images via Mask R-CNN^[36] and DetectNet^[34] helped offset smaller target dataset sizes. Segmentation-based approaches employed secondary algorithms on CNN outputs to estimate spinal parameters^[24,25,31,35,36], adding measurement variability. End-to-end sagittal measurement could minimize error propagation within integrated networks.

Reported batch sizes during neural network training spanned 16-256. However, 10 studies did not specify this optimization detail at all^[18-31,34,36]. Small batches can enhance generalization and reduce overfitting, but at a computational cost. Larger batches offer efficiency yet may miss anomalous cases. Standardization would benefit reproducibility. The median batch size was 64^[24,31,33,36], aligning with typical practices.

The number of training epochs ranged from 30 to 6,000 for deep neural networks. But again, most studies omitted specifics^[18-31,34,36]. Two reports described adaptive epoch counts based on validation improvements^[24,36], rather than fixed values. Typical regimes were 30-50 epochs^[31,33]. Standardized detail would aid reproducibility. Generalizability with shorter training requires scrutiny where transfer learning was not employed.

The IJMEDI checklist for medical imaging AI highlighted several shortcomings (see tabulated results in requests), particularly around enabling reproducibility. Areas such as software details, computational resource usage, model accessibility, and evaluation set specificity suffered poor reporting. However, studies did well in conveying overall aims, statistical and evaluation methodology, and limitations. Recent initiatives for standardizing ML reporting^[39,40], plus reproducibility checklists^[38], may benefit new spine AI imaging research.

Despite promising accuracy, certain limitations remain. Most studies used single-institutional data lacking sufficient diversity^[19-21,23,25,28-30]. Reference standards from manual radiograph measurements intrinsically incorporate subjectivity from inter-observer variations^[41]. CT imaging remains unevaluated. Studies for some parameters are still few. Real-world clinical validation is lacking^[42]. Our subgroup analyses found that studies using CNN architectures demonstrated higher accuracy for parameters like lumbar lordosis compared to other models. This highlights the importance of selecting appropriate architectures tailored to the specific radiographic quantification task. As deep learning continues advancing, further research is still needed to optimize model design and determine the most effective architectures for automated spinopelvic measurement. Larger comparative studies evaluating different network architectures on common datasets would help elucidate the relative merits and guide selection.

Moving forward, larger multicenter studies should validate these models before clinical implementation^[40,43]. Continued research on handling label noise and measurement uncertainty is required^[13,41]. Standardized reporting guidelines could enhance reproducibility^[40]. Models should be optimized across diverse settings and pathologies^[42,43]. Clinically meaningful accuracy metrics deserve focus beyond errors^[41].

The application of deep learning models and their potential role in spine surgery has already begun to be explored. Of value to spine surgeons, models have demonstrated success in diagnosing various musculoskeletal and spinal disorders, including sarcopenia, scoliosis, and low back pain^[44-47]. In regard to prognosis, deep learning models have been successful in predicting postoperative complications such as surgical site infections and 30-day readmission rates after lumbar fusion procedures^[48,49]. While these initial findings are promising, further research validating the use of these models in other realms of patient care, particularly surgical planning, is needed.

Spinopelvic parameters are of great importance to the surgeon for planning, and methods of measurement have evolved significantly. Early assessments began with the Cobb angle and focused on spinal curvatures but overlooked the pelvis. In the 1980s and 1990s, the introduction of parameters such as PI, PT, and SS revolutionized the understanding of sagittal balance. These measurements linked pelvic alignment to spinal posture. By the 2000s, global spinal alignment gained attention, with the SVA and newer measures like the T1 pelvic angle (TPA) becoming essential for surgical planning in adult spinal deformity (ASD).

Up until the early 2000s, measurement of spinopelvic parameters was mostly done manually and, on average, took 3-15 min. The manual measurement process is tedious and time-consuming while also being

prone to rater-dependent error^[50]. Advancements in imaging techniques, including full-body electron optic system (EOS) radiographs, CT scans, and MRI, have enabled more accurate measurements of spinopelvic parameters. The development of more sophisticated software has led to accelerated measurement times via semi-automated computer-aided tools, such as SurgiMap^[50]. Software tools such as SurgiMap have demonstrated a mean time efficiency of 75 ± 25 s to perform a full spinopelvic analysis, significantly reducing the burden associated with manual measurements^[50]. Our review of the existing literature on deep learning models for spinopelvic parameter measurement revealed processing times ranging from 0.2 to 1 s per image. A set of radiographs for spinopelvic parameter measurement typically involves 2-3 images on average: a lateral X-ray, an anterior and posterior X-ray, and possibly a full-body EOS image in more complicated cases. Regarding time saved, deep learning models would require an estimated 0.6-3 s to analyze a full set of images compared to the 75-second mean from the studies mentioned previously. Deep learning models are, therefore, roughly 25× more efficient. Additionally, there were studies included in our analysis that involved pathological images, whereas the study using SurgiMap involved images with no pathology, further demonstrating the capability and efficiency of deep learning technology. To contextualize these efficiency gains with accuracy: Manual measurements typically show inter-observer variability of 5° - 10° for the Cobb angle and similar ranges for other parameters. Semi-automated tools reduced this variability to 3° - 7° . Our meta-analysis found AI measurement errors of 4.3° for Cobb angle, 3.9° for thoracic kyphosis, and 3.6° for lumbar lordosis - comparable to or better than both manual and semi-automated methods. This suggests AI can dramatically improve measurement efficiency without compromising accuracy, potentially offering both time savings and measurement reliability improvements in clinical practice.

No one model stood out as superior to the others. Each study and the model they used had advantages and disadvantages that are open to interpretation. For example, the model used by Zerouali *et al.* was mainly tested in a pediatric population; therefore, this model would likely only be of interest to a surgeon who operates on this population^[22]. Many studies only involved a single clinical dataset, which is a key reason why we argue for multicenter validation to demonstrate reproducibility. Additionally, some studies did not train their models on patients who had implants. Therefore, these models would require further validation to be useable in scenarios such as postoperative evaluation and planning for revision surgery. What was consistent across all models was that they all were more efficient than current methods without compromising accuracy.

Despite the demonstrated accuracy and efficiency of these models, there remains a gap in understanding their practical utility for surgeons across various clinical contexts, including preoperative and intraoperative stages. Theoretically, the enhancement in efficiency should offer surgeons more time to review images and make surgical plans. Pending multicenter validation, future research should explore whether or not the integration of deep learning truly enhances efficiency throughout the entire perioperative continuum. For example, a surgeon may use deep learning as an adjunct for formulating a preoperative plan. Within surgery, intraoperative X-ray image evaluation may allow synchronous measurement of spinopelvic parameters to assess the efficacy of hardware placement. Lastly, in the postoperative phase, the technology can be used to predict postoperative complications and 30-day readmission rates as stated earlier, with the potential for much more. No one model stood out as superior to the others. Each study and the model it used had advantages and disadvantages that are open to interpretation.

A notable limitation in measuring PI deserves specific attention. Our meta-analysis found PI measurements had a relatively higher pooled error of 4.1° compared to other pelvic parameters such as PT (1.9°). This larger error can be attributed to several specific challenges: First, the presence of double-dome endplates can

make it difficult to precisely identify the sacral endplate angle. Second, femoral head overlapping, particularly in patients with high BMI or osteoarthritis, can obscure the precise center of the femoral head. Third, the quality of lateral radiographs, especially in patients with wide pelvises, can result in poor visualization of anatomical landmarks due to increased soft tissue density. Fourth, metallic implants such as total hip replacements can create artifacts that interfere with landmark identification. These factors compound measurement uncertainty and likely contribute to the higher error rates observed for PI across studies. Future deep learning models should specifically address these challenges, perhaps through specialized preprocessing steps or architectural modifications designed to better handle landmark obscurity and anatomical variations.

As this technology continues to evolve, it is highly unlikely that it will not play a role in patient healthcare. It is of great importance for future research to ensure adequate ethical standards, as new concepts and technologies are often met with some resistance. Issues with accountability, transparency, and permissions could come into question by involving deep learning in the decision-making process. Therefore, the integration of deep learning technology should come as a complementary tool in the surgical decision-making processes, where surgeons can potentially optimize patient care pathways and improve overall clinical outcomes.

Limitations

This review has certain limitations. The literature search was restricted to studies published in English, potentially excluding some relevant non-English studies. Searches were limited to four databases, although additional sources were hand-searched. Study screening and data extraction were performed by only two reviewers. The meta-analysis combined studies using different deep learning architectures and imaging modalities, which may have introduced heterogeneity. Only mean absolute errors and correlation coefficients were synthesized, although various other accuracy metrics were reported in the studies.

An additional limitation that should be taken into consideration is that the included studies did not account for anatomic variations such as LSTV. The prevalence of LSTV varies widely within the literature, ranging anywhere from 3.3% to 35.6%. A recent study by Khalifé *et al.* demonstrated that patients with low-grade LSTV, defined as Castelv I and II, have similar alignments as PI-matched no-LSTV and, therefore, should have their measurements taken from S1. Patients with high-grade LSTV, defined as Castelv III and IV, have more kyphotic L5-S1 segments with more cranial lumbar apex and thoracolumbar inflection point and, therefore, should have their measurements taken from L5. Future studies involving machine learning models for measuring spinopelvic parameters may have to pre-identify patients with LSTV and manually input the starting point to account for these anatomic variations^[51].

Conclusion

In conclusion, the breadth of imaging, network architecture details, spine pathologies, and statistical validation encompassed within these studies support automated measurement of spinal curvature as viable for clinical integration pending minor reporting enhancements. Multicenter datasets and model access could additionally reinforce external validity and enable incremental developments in this space.

Overall, this review supports deep learning as a potentially transformative technique for automated spinopelvic measurement from radiographs pending rigorous multicenter validation. These AI technologies may eventually improve efficiency, accuracy, and reliability for quantitative spine image analysis.

DECLARATIONS

Authors' contributions

Manuscript writing and revision: Glaser D

Data collection, analysis, and manuscript revision: AlMekkawi AK, Caruso JP

Data contribution and manuscript revision: Chung CY, Khan EZ, Daadaa HM

Conceptualization, progress monitoring, and final manuscript revision: Aoun SG, Bagley CA

Availability of data and materials

The data are available from the corresponding author upon reasonable request.

Financial support and sponsorship

None.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Le Huec JC, Thompson W, Mohsinaly Y, Barrey C, Faundez A. Sagittal balance of the spine. *Eur Spine J.* 2019;28:1889-905. DOI PubMed
2. Vrtovec T, Janssen MM, Likar B, Castelein RM, Viergever MA, Pernuš F. A review of methods for evaluating the quantitative parameters of sagittal pelvic alignment. *Spine J.* 2012;12:433-46. DOI PubMed
3. Legaye J, Duval-Beaupère G, Hecquet J, Marty C. Pelvic incidence: a fundamental pelvic parameter for three-dimensional regulation of spinal sagittal curves. *Eur Spine J.* 1998;7:99-103. DOI PubMed PMC
4. Glassman SD, Bridwell K, Dimar JR, Horton W, Berven S, Schwab F. The impact of positive sagittal balance in adult spinal deformity. *Spine.* 2005;30:2024-9. DOI PubMed
5. Maillot C, Ferrero E, Fort D, Heyberger C, Le Huec JC. Reproducibility and repeatability of a new computerized software for sagittal spinopelvic and scoliosis curvature radiologic measurements: Keops®. *Eur Spine J.* 2015;24:1574-81. DOI PubMed
6. Lafage R, Ferrero E, Henry JK, et al. Validation of a new computer-assisted tool to measure spino-pelvic parameters. *Spine J.* 2015;15:2493-502. DOI PubMed
7. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60-88. DOI
8. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng.* 2017;19:221-48. DOI PubMed PMC
9. Galbusera F, Casaroli G, Bassani T. Artificial intelligence and machine learning in spine research. *JOR Spine.* 2019;2:e1044. DOI PubMed PMC
10. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521:436-44. DOI PubMed
11. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys.* 2019;29:102-27. DOI PubMed
12. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health.* 2019;1:e271-97. DOI
13. Chea P, Mandell JC. Current applications and future directions of deep learning in musculoskeletal radiology. *Skeletal Radiol.* 2020;49:183-97. DOI PubMed
14. Jamaludin A, Lootus M, Kadir T, et al; Genodisc Consortium. ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. *Eur Spine J.* 2017;26:1374-83. DOI
15. Lopez CD, Boddapati V, Lombardi JM, et al. Artificial learning and machine learning applications in spine surgery: a systematic

- review. *Global Spine J*. 2022;12:1561-72. [DOI](#) [PubMed](#) [PMC](#)
16. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int J Surg*. 2021;88:105906. [DOI](#) [PubMed](#)
 17. Cabitza F, Campagner A. The IJMEDI checklist for assessment of medical AI. *Int J Med Inform*. 2021;153. [DOI](#)
 18. Chae DS, Nguyen TP, Park SJ, Kang KY, Won C, Yoon J. Decentralized convolutional neural network for evaluating spinal deformity with spinopelvic parameters. *Comput Methods Programs Biomed*. 2020;197:105699. [DOI](#) [PubMed](#)
 19. Wu H, Bailey C, Rasoulinejad P, Li S. Automated comprehensive adolescent idiopathic scoliosis assessment using MVC-Net. *Med Image Anal*. 2018;48:1-11. [DOI](#) [PubMed](#)
 20. Wang L, Xu Q, Leung S, Chung J, Chen B, Li S. Accurate automated Cobb angles estimation using multi-view extrapolation net. *Med Image Anal*. 2019;58:101542. [DOI](#)
 21. Zhang K, Xu N, Guo C, Wu J. MPF-net: an effective framework for automated cobb angle estimation. *Med Image Anal*. 2022;75:102277. [DOI](#)
 22. Zerouali M, Parpaleix A, Benbakoura M, Rigault C, Champsaur P, Guenoun D. Automatic deep learning-based assessment of spinopelvic coronal and sagittal alignment. *Diagn Interv Imaging*. 2023;104:343-50. [DOI](#) [PubMed](#)
 23. Korez R, Putzier M, Vrtovec T. A deep learning tool for fully automated measurements of sagittal spinopelvic balance from X-ray images: performance evaluation. *Eur Spine J*. 2020;29:2295-305. [DOI](#) [PubMed](#)
 24. Kim YT, Jeong TS, Kim YJ, Kim WS, Kim KG, Yee GT. Automatic spine segmentation and parameter measurement for radiological analysis of whole-spine lateral radiographs using deep learning and computer vision. *J Digit Imaging*. 2023;36:1447-59. [DOI](#) [PubMed](#) [PMC](#)
 25. Yeh YC, Weng CH, Huang YJ, Fu CJ, Tsai TT, Yeh CY. Deep learning approach for automatic landmark detection and alignment analysis in whole-spine lateral radiographs. *Sci Rep*. 2021;11:7618. [DOI](#) [PubMed](#) [PMC](#)
 26. Orosz LD, Bhatt FR, Jazini E, et al. Novel artificial intelligence algorithm: an accurate and independent measure of spinopelvic parameters. *J Neurosurg Spine*. 2022;37:893-901. [DOI](#)
 27. Gami P, Qiu K, Kannappan S, et al. Semiautomated intraoperative measurement of Cobb angle and coronal C7 plumb line using deep learning and computer vision for scoliosis correction: a feasibility study. *J Neurosurg Spine*. 2022;37:713-21. [DOI](#)
 28. Schwartz JT, Cho BH, Tang P, et al. Deep learning automates measurement of spinopelvic parameters on lateral lumbar radiographs. *Spine*. 2021;46:E671-8. [DOI](#)
 29. Aubert B, Vazquez C, Cresson T, Parent S, de Guise JA. Toward automated 3D spine reconstruction from biplanar radiographs using CNN for statistical spine model fitting. *IEEE Trans Med Imaging*. 2019;38:2796-806. [DOI](#) [PubMed](#)
 30. Nguyen TP, Jung JW, Yoo YJ, Choi SH, Yoon J. Intelligent evaluation of global spinal alignment by a decentralized convolutional neural network. *J Digit Imaging*. 2022;35:213-25. [DOI](#) [PubMed](#) [PMC](#)
 31. Galbusera F, Niemeier F, Wilke HJ, et al. Fully automated radiological analysis of spinal disorders and deformities: a deep learning approach. *Eur Spine J*. 2019;28:951-60. [DOI](#)
 32. Zhang T, Zhu C, Lu Q, Liu J, Diwan A, Cheung JPY. A novel tool to provide predictable alignment data irrespective of source and image quality acquired on mobile phones: what engineers can offer clinicians. *Eur Spine J*. 2020;29:387-95. [DOI](#) [PubMed](#)
 33. Horng MH, Kuok CP, Fu MJ, Lin CJ, Sun YN. Cobb angle measurement of spine from X-ray images using convolutional neural network. *Comput Math Methods Med*. 2019;2019:6357171. [DOI](#) [PubMed](#) [PMC](#)
 34. Sun H, Zhen X, Bailey C, Rasoulinejad P, Yin Y, Li S. Direct estimation of spinal Cobb angles by structured multi-output regression. In: Niethammer M, Styner M, Aylward S, Zhu H, Oguz I, Yap P, Shen D, editors. Information processing in medical imaging. Cham: Springer International Publishing; 2017. pp. 529-40. [DOI](#)
 35. Weng CH, Wang CL, Huang YJ, et al. Artificial intelligence for automatic measurement of sagittal vertical axis using ResUNet framework. *J Clin Med*. 2019;8:1826. [DOI](#) [PubMed](#) [PMC](#)
 36. H A, Prabhu GK. Automatic quantification of spinal curvature in scoliotic radiograph using image processing. *J Med Syst*. 2012;36:1943-51. [DOI](#) [PubMed](#)
 37. Zhang J, Lou E, Le LH, Hill DL, Raso JV, Wang Y. Automatic Cobb measurement of scoliosis based on fuzzy Hough Transform with vertebral shape prior. *J Digit Imaging*. 2009;22:463-72. [DOI](#) [PubMed](#) [PMC](#)
 38. 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. [DOI](#) [PubMed](#) [PMC](#)
 39. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health*. 2020;2:e537-48. [DOI](#) [PubMed](#) [PMC](#)
 40. Mongan J, Moy L, Kahn CE Jr. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell*. 2020;2:e200029. [DOI](#) [PubMed](#) [PMC](#)
 41. Willemink MJ, Koszek WA, Hardell C, et al. Preparing medical imaging data for machine learning. *Radiology*. 2020;295:4-15. [DOI](#) [PubMed](#) [PMC](#)
 42. Ghaednia H, Lans A, Sauder N, et al. Deep learning in spine surgery. *Semin Spine Surg*. 2021;33:100876. [DOI](#)
 43. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17:195. [DOI](#) [PubMed](#) [PMC](#)
 44. Cho BH, Kaji D, Cheung ZB, et al. Automated measurement of lumbar lordosis on radiographs using machine learning and computer

- vision. *Global Spine J.* 2020;10:611-8. [DOI](#) [PubMed](#) [PMC](#)
45. Burns JE, Yao J, Chalhoub D, Chen JJ, Summers RM. A machine learning algorithm to estimate sarcopenia on abdominal CT. *Acad Radiol.* 2020;27:311-20. [DOI](#) [PubMed](#)
 46. Yang J, Zhang K, Fan H, et al. Development and validation of deep learning algorithms for scoliosis screening using back images. *Commun Biol.* 2019;2:390. [DOI](#) [PubMed](#) [PMC](#)
 47. Hu B, Kim C, Ning X, Xu X. Using a deep learning network to recognise low back pain in static standing. *Ergonomics.* 2018;61:1374-81. [DOI](#) [PubMed](#)
 48. Staartjes VE, de Wispelaere MP, Vandertop WP, Schröder ML. Deep learning-based preoperative predictive analytics for patient-reported outcomes following lumbar discectomy: feasibility of center-specific modeling. *Spine J.* 2019;19:853-61. [DOI](#) [PubMed](#)
 49. Hines AL, Barrett ML, Jiang HJ, Steiner CA. Conditions with the largest number of adult hospital readmissions by payer, 2011. In: Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. [PubMed](#)
 50. Valenzuela JG, Cirillo Totera JI, Turkieltaub DH, Echaurren CV, Álvarez Lemos FL, Arriagada Ramos FI. Spino-pelvic radiological parameters: comparison of measurements obtained by radiologists using the traditional method versus spine surgeons using a semi-automated software (Surgimap). *Acta Radiol Open.* 2023;12:20584601231177404. [DOI](#) [PubMed](#) [PMC](#)
 51. Khalifé M, Lafage R, Daniels AH, et al; International Spine Study Group. Assessing abnormal proximal junctional angles in adult spinal deformity: a normative data approach to define proximal junctional kyphosis. *Spine.* 2025;50:103-9. [DOI](#) [PubMed](#)

Commentary

Open Access



Spino-plastic surgery, back to the future

Casey Martinez¹, Camryn Payne¹, Jonathan L. Jeger², Nicole Van Spronsen², Sebastian Winocou³, Maziyar A. Kalani⁴, Michael Bohl⁵, Alexander E. Ropper⁶, Edward M. Reece²

¹Alix School of Medicine, Mayo Clinic, Scottsdale, AZ 85259, USA.

²Division of Plastic Surgery, Mayo Clinic, Phoenix, AZ 85054, USA.

³Department of Plastic Surgery, Baylor College of Medicine, Houston, TX 77030, USA.

⁴Department of Neurosurgery, Mayo Clinic, Phoenix, AZ 85054, USA.

⁵Carolina Neurosurgery & Spine Associates, Charlotte, NC 28204, USA.

⁶Department of Neurosurgery, Baylor College of Medicine, Houston, TX 28204, USA.

Correspondence to: Dr. Edward M. Reece, Division of Plastic Surgery, Mayo Clinic, 5777 E Mayo Blvd, Phoenix, AZ 85054, USA.
E-mail: reece.edward@mayo.edu

How to cite this article: Martinez C, Payne C, Jeger JL, Van Spronsen N, Winocou S, Kalani MA, Bohl M, Ropper AE, Reece EM. Spino-plastic surgery, back to the future. *Art Int Surg.* 2025;5:16-23. <https://dx.doi.org/10.20517/ais.2024.73>

Received: 3 Sep 2024 **First Decision:** 31 Oct 2024 **Revised:** 30 Nov 2024 **Accepted:** 2 Dec 2024 **Published:** 4 Jan 2025

Academic Editor: Andrew Gumbs **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Artificial intelligence (AI) is a powerful computational tool that is being utilized more frequently in healthcare. AI holds promise within surgical practice, including application in the care of challenging patient populations. Complex spine reconstruction requires thorough multi-variable preoperative analysis and then the precise enactment of a surgical plan. Spino-plastics employs vascularized bone grafts (VBGs) to augment spinal fusion in these high-risk patients. In this article, we discuss the great breadth of AI and the tremendous potential for advancing the field of spino-plastics: surgical candidacy and patient selection, imaging and virtual surgical planning (VSP), intraoperative utilization, and future implementation.

Keywords: Artificial intelligence, spino-plastics, complex spinal reconstruction, machine learning, virtual surgical planning

INTRODUCTION

Artificial intelligence (AI) refers to computer systems that employ algorithms to analyze data, generate predictions, solve problems, and make decisions in a human-like fashion^[1-3]. A range of technologies fall



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



under the definition of AI, including machine learning (ML), deep learning (DL), natural language processing, and computer vision^[2]. Given the immeasurable quantity of patient health data and increasingly advanced technologies capable of processing it, there are ample applications for AI in the spino-plastics domain^[1].

Spino-plastic surgery is one surgical subspecialty that combines the talents of interdisciplinary surgical subspecialists within plastics, orthopedics, and neurosurgery to meet the needs of patients requiring complex spinal reconstruction. In brief, spino-plastics utilizes vascularized bone grafts (VBGs) from the iliac crest, spinous process, rib, scapula, clavicle, and occiput to augment the strength of spinal fusions necessitated by pathologies such as trauma, degeneration, or tumor^[4-9]. VBGs are pedicled on muscle and supplied by Sharpey's fibers, which physically connect the muscle to bone and allow small unnamed periosteal feeding vessels to vascularize Haversian canals^[10]. VBGs are increasingly indicated for the treatment of pseudoarthrosis, as they increase osteogenesis, osteoconductivity, and osteoinductivity compared to non-vascularized bone grafts (N-VBGs)^[10]. Rates of pseudoarthrosis following arthrodesis can reach 60% or higher, leading to reoperations and significant morbidity that negatively impact quality of life^[11,12]. VBGs have been incorporated into the existing reconstructive algorithm that is divided into six levels: allograft, bony substitution, autograft, N-VBG, VBG, pedicled vascularized bone flap, and free bone flap^[10]. As VBGs have been found to enhance the strength of spinal fusion^[13] and decrease rates of pseudoarthrosis, there is a need for an AI algorithm to identify those at risk for pseudoarthrosis who may benefit from VBG. Key areas of research interest within spino-plastics include the identification of optimal surgical candidates given the expanding indications for VBGs, as well as improving surgical techniques to enhance patient outcomes.

In the literature, there is already evidence of AI algorithms developed to screen for vulnerable patient populations and identify surgical candidates^[1,3,14]. Furthermore, there are many existing AI algorithms with similar functions of patient risk stratification. Within spine surgery, AI has already been applied to identify surgical candidates and treatment options for anterior decompression and fusion for cervical spondylotic myelopathy^[15,16], as well as to predict quality of life outcomes in adult spinal deformities^[17]. The future of spine surgery may be guided by bioinformaticians, data engineers, and computer scientists who process big data in a way that informs patient care and scientific discovery^[18]. In this article, we conducted a non-systematic narrative review of the literature to better understand AI's capability to transform the field of spino-plastics through assessment of surgical candidacy and patient selection, imaging and virtual surgical planning (VSP), and intraoperative instrument manipulation.

SURGICAL CANDIDACY AND PATIENT SELECTION

Disease classification systems are invaluable tools when applied appropriately within medical practice. While a classification score does not solely drive available treatment options, it is a standardized entry point and a piece to the overall puzzle in the care of patients with complex pathology. Unsupervised AI data analysis can create new hierarchical clustering that accounts for patient frailty scores, functional status, radiographic characteristics, and many demographic factors^[19]. Sophisticated pattern analysis incorporates more data than could have been previously imagined, making surgeon education easier with elaborate risk-benefit grids for various treatment pathways^[19].

Predictive algorithms are an excellent way to identify high-risk patients more effectively, such as those who are at a greater than average risk of pseudarthrosis, wound breakdown, or morbidity/mortality associated with spinal fusion. In general, earlier identification of high-risk patients allows for earlier intervention with proactive employment of strategies to mitigate the risks inherent to the patient or pathology itself. In spino-

plastics specifically, this concept may be harnessed to identify patients who might benefit from a risk-reducing VBG and mitigate the consequences of debilitating pseudoarthrosis and failed spinal fusion. Given the current novelty of spino-plastic surgery, this information is currently unvalidated. AI and ML move beyond the traditional linear or logistic regression, incorporating greater dimensions of analysis to more accurately identify those who may benefit from VBGs and, therefore, bring novel prognostic indicators to light^[20-22]. In spinal tumor surgery, which is a common indication for VBG, AI already has a role in patient risk stratification^[23]. A Naïve Bayes ML algorithm developed by DiSilvestro *et al.* was better at predicting 30-day mortality following spinal tumor resection than the National Surgical Quality Initiative mortality probability calculator^[24]. This algorithm is based on Bayes' theorem and predicts mortality based on independent patient risk factors. For example, this study found smoking, cancer, and chronic obstructive pulmonary disease to all be independent risk factors for 30-day mortality in intraspinal neoplasm excision. AI's role in predicting outcomes is a powerful tool that could similarly be harnessed in spino-plastics decision making.

Informed, shared decision making between the patient and provider fosters an atmosphere conducive to the best outcomes for everyone. ML eliminates barriers to the availability of high-powered data by generating accurate model populations that are based on groups of detailed, real patient data^[25]. ML enhances understanding of the preoperative conditions and how this is likely to interact with desired surgical outcomes, effectively matching patients with the best available treatment options^[26]. Combining physical exam findings or patient presentation with patient-specific anatomy in advanced imaging studies has the possibility to address healthcare disparities, improving access to care and creating a higher standard in fine-tuning preoperative patient selection^[26]. In the context of spine surgery, one example of this concept in action is Wilson *et al.*'s AI model designed to predict when the degree of spinal stenosis by magnetic resonance imaging (MRI) requires specialist evaluation, streamlining subspecialty referrals for the benefit of earlier access to appropriate care and timely intervention^[27].

Taking this one step further, ML can intelligently engineer algorithms that can achieve a high negative predictive value in determining the need for surgical intervention, possibly, and alternatively, higher risk situations, thereby amplifying vigilance toward those patients and redirecting the limited resources of office visits^[28]. The quadruple aim of healthcare involves increasing patient and provider experience and improving population health while reducing overall healthcare costs^[29]. AI utilization aligns well with the quadruple aim of healthcare by making the surgical triage experience more beneficial to all parties involved, increasing the ratio of surgical bookings to total patients seen in the clinic and improving the quality of care^[28]. Overall, AI holds the potential to shift the paradigm of decision making in spine surgery.

IMAGING AND VSP

In addition to AI's contributions to surgical candidacy and patient risk stratification, its involvement in radiologic studies plays an integral role in several aspects of spine surgery. AI algorithms have already been developed to assist in the classification and localization of spinal tumors^[30,31]. Zhuo *et al.* developed a DL model to classify spinal tumors using T2-weighted MRIs^[30]. In a similar capacity, Liu *et al.* proposed a model utilizing a weighted fusion framework on MRI data to locate tumors and synthesize patient clinical information for more accurate tumor classification than doctors^[31].

The automation of qualitative and quantitative radiologic interpretation promises advancement in volumetric assessments of tumors, determination of tumor genotype from phenotypic characteristics, disease or treatment burden on tumor-adjacent tissues, and much more^[32]. As time progresses and technology improves, an increasing number of studies might attain results that reach clinical significance,

with hopes of significant alterations to radiologic evaluation. For example, the important work of Wang *et al.* proved the clinical utility of applying deep neural networks for the detection of spinal metastasis, reaching an accuracy of 90%^[33].

Specifically, within spino-plastics, there are many applications for AI in diagnostics and imaging. Spino-plastic reconstruction is concerned with (1) increased bony fusion, especially in the setting of previous failures; (2) decreased time to bony fusion; (3) optimizing the interface between soft tissues, hardware, and osseous structures in both form and skeletal function; (4) stable and long-term closure of wounds: these are data points that are amenable to AI application. As mentioned, there is a wide range of pathology warranting VBGs to augment spinal fusion, including prior failed fusion and extensive reconstruction after tumor extirpation. A closer and more comprehensive evaluation of radiologic studies might provide insight into patients that necessitate further intervention to offer a better chance at successful fusion. In addition, computer modeling based on multidimensional analysis of various imaging modalities might also propose the vertebral level incurring the greatest mechanical stress status post instrumentation and fusion^[34], further aiding the surgical team in deciding the final target for VBG fixation.

This brings us to the discussion of VSP. This technique employs patient imaging to construct a 3D surgical model that allows for surgical simulation, visualization of complex anatomy, and virtual mapping to assist with procedural planning^[35]. Over the past several decades, VSP has been widely adopted within orthognathic surgery, providing an alternative to traditional surgical planning techniques^[36]. VSP improves surgical accuracy, creating more symmetry than would have otherwise been possible without this technology^[36]. Therefore, VSP is trusted by orthognathic surgeons who operate in a field where aesthetic results are of paramount importance^[36]. While there is still much room for growth in this surgical tool, it has been suggested that AI will only increase the scope of VSP^[37]. In a recent 2023 study, Browd *et al.* describe how patient-specific quantitative metrics, such as bone density, sagittal balance, and Cobb angles, derived from imaging modalities can potentially be applied to AI and ML algorithms for better surgical planning^[26].

INTRAOPERATIVE UTILIZATION

Intraoperatively, AI can be very helpful in tumor resection and reconstruction of the spine. For instance, AI can assist surgeons in differentiating between normal tissue and glioblastoma multiforme^[38]. Alternatively, AI might be harnessed to improve existing technology and intraoperative decision making. Many devices and techniques have been described for improved intraoperative performance in spine surgery. Computer-assisted navigation systems such as stealth guidance assist in surgical planning and operational precision^[23]. Stealth guidance is a robotic technology that enhances intraoperative localization and accuracy through three-dimensional modeling^[39]. Stealth guidance systems such as Medtronic's StealthStation employ imaging data in the form of MRI and CT scans to create multidimensional anatomic models and real-time navigation that allows surgeons to know precisely where they are in space^[40,41]. The precision enhancement of robotic-assisted stealth guidance has been demonstrated to reduce operative times and decrease intra- and postoperative complications in neurosurgical and spinal procedures requiring a higher level of dexterity and accuracy^[42-44].

Another distinct piece of computer-assisted navigation systems that improve operative efficacy is augmented reality (AR). This technology assists with intraoperative navigation by overlaying graphics in the real world, enhancing the perception of surgical instruments in space^[45,46]. By incorporating an overlay of surgical plans or highlighting relevant anatomy, surgeons are provided with real-time information that enhances their visualization and proprioception without the need to divert their attention away from the patient toward a screen^[47]. AR has been integrated into fields like orthopedic surgery, trauma surgery, and

spinal surgery to help with preoperative planning and surgical training^[45,48]. The approach and positioning of pedicle screws, foraminotomies, percutaneous interventions, and biopsies can all be achieved more safely, with less margin of error, under the guidance of AR^[49]. AR also permits spine surgeons to view dissection planes and tumor volumes with microscopic virtual mapping for performing osteotomies^[50]. Ma *et al.* describe an ultrasound methodology to superimpose surgical planning *in situ* by incorporating CT images with 3D anatomic landmarks^[51].

Spine surgery can be challenging at baseline, as it is not uncommon to lack direct exposure or visualization of the intricate, densely organized vessels and nerves along the axial skeleton. By the very nature of the field, spino-plastics aims to treat an even more challenging subset of patients. The distortion of native anatomy in complex cases, whether caused by revision surgery or the mass effect of tumor bulk, presents additional obstacles to intraoperative identification of neurovasculature. AR can aid surgeons in this difficult task, employing visual information from MRI and CT scans to build surgical maps and chart paths around key anatomic structures^[45,46]. In spino-plastics cases, once the spinal instrumentation and fusion are complete, the surgeon may harvest the VBG utilizing the standard arthrodesis instruments that are already on the sterile field. If stealth guidance or AR is already being utilized for arthrodesis, it would be wise to consider keeping the system available to assist the surgeon in harvesting and ensuring adequate fixation of the VBG. Better spatial conceptualization of the instrumentation might reduce any risk of damaging nearby structures in the vertebral column or retroperitoneum.

LOOKING TO THE FUTURE

Notably, the ultimate boundaries of AI have yet to be uncovered. AI has already contributed to our understanding of driver mutations behind spinal cord tumors^[23]. This incredible technology will continue to improve basic science research and treatment modalities to address the needs of spino-plastic patients from many different perspectives. Despite the tremendous promise and exponential rise in these technological advancements, there is much work to be done before clinicians may be completely comfortable about incorporating this new technology into their workflow. Because ML is a powerful tool that is not fully understood, caution must be exercised regarding the input of information to avoid the perpetuation of misinformation and social biases. Overall, ML and AI currently lack transparency, which creates a “black box” that may be difficult for surgeons to trust when comparing results to well-published algorithms that have a more easily understood basis. However, there are methods currently being utilized to validate their efficacy in clinical practice. This includes the results of case studies and trials - where technologies such as imaging guidance can differentiate tumors from healthy tissue^[52] - comparative studies, and live integration with surgical teams^[3] that provide constant feedback to enhance the safety and predictive power of AI algorithms. Many metrics were used in these various studies to compare the performance of AI algorithms to traditional models, such as the area under the curve, accuracy, and the receiver operating characteristic curve. Furthermore, there is an upfront investment of time and resources essential for the development of novel algorithms bearing any clinical significance. In other words, there is a significant lag time between technological advancements and gaining necessary approvals for clinical application through the proper avenues, including national supervisory organizations and individual hospital systems^[53]. In this stage of conceptualization, there are limited existing data on AI in spino-plastic surgery and further long-term data collection is required.

Despite the harvest and fixation of VBGs not requiring any additional tools that are traditionally used in spinal fusion, the field of spino-plastics is in its nascent stages. Due to resource limitations or surgeon-specific comfort levels with working in the spine and retroperitoneum, not all institutions have access to plastic surgeons capable of performing this procedure. Developing strategies to implement novel AI

technologies beyond academic practice, particularly in rural communities, is essential to ensuring equity in an increasingly digital age. Research has already discussed several key strategies, including improving digital infrastructure, such as internet access, and networks of local health information that can be employed to train AI. Training local community healthcare workers to utilize novel technologies such as mobile health applications and engaging with community stakeholders to determine the most impactful implementation strategies are also crucial^[54].

CONCLUSION

Spino-plastics uses a long-standing well-accepted concept of VBGs and applies it to quite complex reconstructive problems. Innovation is at the heart of this field, and spine surgeons are no strangers to welcoming new technologies and techniques. AI holds great promise in advancing medicine overall, making data collection and processing easier than ever with seemingly unending applications for the delivery of patient care. Beyond its potential role in patient selection, the visual enhancement offered by AI technologies can assist in diagnostics, surgical planning, and intraoperative precision. Spinal tumor resection often results in complex spinal defects that are nonuniform and in close proximity to several critical structures. In the planning and intraoperative phases, AI can improve outcomes by enhancing the accuracy of instrument movements and assisting with surgical planning and decision making^[55,56]. When AI is used in conjunction with other advanced technologies such as AR or stealth guidance, three-dimensional visualization is further enhanced, reducing risks of intraoperative complications^[45]. Thus, AI may one day function as a spino-plastic surgeon's first assistant in future operating rooms. Spinal fusion calls for advancements and synergy in AI, robotics, and AR. There is great promise in the collaborative opportunities that telemedicine and telesurgery will bring, dismantling the geographic and socioeconomic barriers to centers of excellence in multidisciplinary care^[25]. In conclusion, the integration of AI into spino-plastic surgery not only has the power to further individualize and enhance VBG's precision and effectiveness, but also broaden their potential indications, ultimately transforming the landscape of complex spinal reconstruction and offering new possibilities for patient care.

DECLARATIONS

Authors' contributions

Drafted this original manuscript: Martinez C, Payne C, Jeger JL, Van Spronsen N

Conceptualized and edited this manuscript: Winocou S, Kalani MA, Bohl M, Ropper AE, Reece EM

Availability of data and materials

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Guni A, Varma P, Zhang J, Fehervari M, Ashrafian H. Artificial intelligence in surgery: the future is now. *Eur Surg Res*. 2024;65:22-39. DOI
2. Malhotra K, Wong BN, Lee S, et al. Role of artificial intelligence in global surgery: a review of opportunities and challenges. *Cureus*. 2023;15:e43192. DOI PubMed PMC
3. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg*. 2018;268:70-6. DOI PubMed PMC
4. Reece EM, Davis MJ, Wagner RD, et al. Vascularized bone grafts for spinal fusion-part 1: the iliac crest. *Oper Neurosurg*. 2021;20:493-6. DOI
5. Reece EM, Agrawal N, Wagner KM, et al. Vascularized bone grafts for spinal fusion-part 2: the rib. *Oper Neurosurg*. 2021;20:497-501. DOI
6. Bohl MA, Reece EM, Farrokhi F, Davis MJ, Abu-Ghname A, Ropper AE. Vascularized bone grafts for spinal fusion-part 3: the occiput. *Oper Neurosurg*. 2021;20:502-7. DOI PubMed
7. Reece EM, Davis MJ, Abu-Ghname A, et al. Vascularized bone grafts for spinal fusion-part 4: the scapula. *Oper Neurosurg*. 2021;20:508-12. DOI
8. Bohl MA, Almefty KK, Preul MC, et al. Vascularized spinous process graft rotated on a paraspinous muscle pedicle for lumbar fusion: technique description and early clinical experience. *World Neurosurg*. 2018;115:186-92. DOI
9. Abdulwadood I, Gomez DA, Martinez C, et al. Vascularized bone grafts in spinal reconstruction: an updated comprehensive review. *Orthop Surg*. 2024;17:7-14. DOI
10. Skochdopole AJ, Wagner RD, Davis MJ, et al. Vascularized bone grafts in spinal reconstruction: an overview of nomenclature and indications. *Semin Plast Surg*. 2021;35:50-3. DOI PubMed PMC
11. Pennington Z, Mehta VA, Lubelski D, et al. Quality of life and cost implications of pseudarthrosis after anterior cervical discectomy and fusion and its subsequent revision surgery. *World Neurosurg*. 2020;133:e592-9. DOI
12. McAnany SJ, Baird EO, Overley SC, Kim JS, Qureshi SA, Anderson PA. A meta-analysis of the clinical and fusion results following treatment of symptomatic cervical pseudarthrosis. *Global Spine J*. 2015;5:148-55. DOI PubMed PMC
13. Verla T, Xu DS, Davis MJ, et al. Failure in cervical spinal fusion and current management modalities. *Semin Plast Surg*. 2021;35:10-3. DOI PubMed PMC
14. Buchem MM, Boosman H, Bauer MP, Kant IMJ, Cammel SA, Steyerberg EW. The digital scribe in clinical practice: a scoping review and research agenda. *NPJ Digit Med*. 2021;4:57. DOI PubMed PMC
15. Park D, Cho JM, Yang JW, et al. Classification of expert-level therapeutic decisions for degenerative cervical myelopathy using ensemble machine learning algorithms. *Front Surg*. 2022;9:1010420. DOI PubMed PMC
16. Zhou C, Huang S, Liang T, et al. Machine learning-based clustering in cervical spondylotic myelopathy patients to identify heterogeneous clinical characteristics. *Front Surg*. 2022;9:935656. DOI PubMed PMC
17. Mekhael E, El Rachkidi R, Saliby RM, et al. Functional assessment using 3D movement analysis can better predict health-related quality of life outcomes in patients with adult spinal deformity: a machine learning approach. *Front Surg*. 2023;10:1166734. DOI PubMed PMC
18. Hornung AL, Hornung CM, Mallow GM, et al. Artificial intelligence in spine care: current applications and future utility. *Eur Spine J*. 2022;31:2057-81. DOI
19. Ames CP, Smith JS, Pellisé F, et al; European Spine Study Group, International Spine Study Group. Artificial intelligence based hierarchical clustering of patient types and intervention categories in adult spinal deformity surgery: towards a new classification scheme that predicts quality and value. *Spine*. 2019;44:915-26. DOI
20. Shah AA, Devana SK, Lee C, et al. Machine learning-driven identification of novel patient factors for prediction of major complications after posterior cervical spinal fusion. *Eur Spine J*. 2022;31:1952-9. DOI PubMed PMC
21. Hassan AM, Rajesh A, Asaad M, et al. Artificial intelligence and machine learning in prediction of surgical complications: current state, applications, and implications. *Am Surg*. 2023;89:25-30. DOI PubMed PMC
22. Martin BI, Bono CM. Artificial intelligence and spine: rise of the machines. *Spine J*. 2021;21:1604-5. DOI PubMed
23. Katsos K, Johnson SE, Ibrahim S, Bydon M. Current applications of machine learning for spinal cord tumors. *Life*. 2023;13:520. DOI PubMed PMC
24. DiSilvestro KJ, Veeramani A, McDonald CL, et al. Predicting postoperative mortality after metastatic intraspinal neoplasm excision: development of a machine-learning approach. *World Neurosurg*. 2021;146:e917-24. DOI
25. Benzakour A, Altsitzioglou P, Lemée JM, Ahmad A, Mavrogenis AF, Benzakour T. Artificial intelligence in spine surgery. *Int Orthop*. 2023;47:457-65. DOI PubMed
26. Brownd SR, Park C, Donoho DA. Potential applications of artificial intelligence and machine learning in spine surgery across the continuum of care. *Int J Spine Surg*. 2023;17:S26-33. DOI PubMed PMC
27. Wilson B, Gaonkar B, Yoo B, et al. Predicting spinal surgery candidacy from imaging data using machine learning. *Neurosurgery*.

- 2021;89:116-21. DOI PubMed PMC
28. Broida SE, Schrum ML, Yoon E, et al. Improving surgical triage in spine clinic: predicting likelihood of surgery using machine learning. *World Neurosurg.* 2022;163:e192-8. DOI
29. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med.* 2014;12:573-6. DOI PubMed PMC
30. Zhuo Z, Zhang J, Duan Y, et al. Automated classification of intramedullary spinal cord tumors and inflammatory demyelinating lesions using deep learning. *Radiol Artif Intell.* 2022;4:e210292. DOI
31. Liu H, Jiao M, Yuan Y, et al. Benign and malignant diagnosis of spinal tumors based on deep learning and weighted fusion framework on MRI. *Insights Imaging.* 2022;13:87. DOI PubMed PMC
32. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin.* 2019;69:127-57. DOI PubMed PMC
33. Wang J, Fang Z, Lang N, Yuan H, Su MY, Baldi P. A multi-resolution approach for spinal metastasis detection using deep Siamese neural networks. *Comput Biol Med.* 2017;84:137-46. DOI PubMed PMC
34. Cui Y, Zhu J, Duan Z, Liao Z, Wang S, Liu W. Artificial intelligence in spinal imaging: current status and future directions. *Int J Environ Res Public Health.* 2022;19:11708. DOI PubMed PMC
35. Singh GD, Singh M. Virtual surgical planning: modeling from the present to the future. *J Clin Med.* 2021;10:5655. DOI PubMed PMC
36. Chen Z, Mo S, Fan X, You Y, Ye G, Zhou N. A meta-analysis and systematic review comparing the effectiveness of traditional and virtual surgical planning for orthognathic surgery: based on randomized clinical trials. *J Oral Maxillofac Surg.* 2021;79:471.e1-19. DOI
37. Mangano FG, Admakin O, Lerner H, Mangano C. Artificial intelligence and augmented reality for guided implant surgery planning: a proof of concept. *J Dent.* 2023;133:104485. DOI PubMed
38. Marcus AP, Marcus HJ, Camp SJ, Nandi D, Kitchen N, Thorne L. Improved prediction of surgical resectability in patients with glioblastoma using an artificial neural network. *Sci Rep.* 2020;10:5143. DOI PubMed PMC
39. Atai NA, Mehta VA. Initial United States experience with Medtronic Stealth Autoguide cranial robotic guidance platform. *J Neurosurg.* 2024;141:1520-6. DOI PubMed
40. Medtronic. Available from: <https://www.medtronic.com/us-en/healthcare-professionals/products/neurological/surgical-navigation-systems/stealthstation.html>. [Last accessed on 30 Dec 2024].
41. Allina Health. Navigation systems. Available from: <https://www.allinahealth.org/united-hospital/services/john-nasseff-neuroscience-institute/technology/navigation-systems>. [Last accessed on 30 Dec 2024].
42. Silva AK, Preminger A, Slezak S, Phillips LG, Johnson DJ. Melting the plastic ceiling: overcoming obstacles to foster leadership in women plastic surgeons. *Plast Reconstr Surg.* 2016;138:721-9. DOI
43. Sharma JD, Seunarine KK, Tahir MZ, Tisdall MM. Accuracy of robot-assisted versus optical frameless navigated stereoelectroencephalography electrode placement in children. *J Neurosurg Pediatr.* 2019;23:297-302. DOI
44. Fan X, Mirza SK, Li C, Evans LT, Ji S, Paulsen KD. Accuracy of stereovision-updated versus preoperative CT-based image guidance in multilevel lumbar pedicle screw placement: a cadaveric swine study. *JB JS Open Access.* 2022;7:e21.00129. DOI PubMed PMC
45. Chen H. Application progress of artificial intelligence and augmented reality in orthopaedic arthroscopy surgery. *J Orthop Surg Res.* 2023;18:775. DOI PubMed PMC
46. Xiong J, Hsiang EL, He Z, Zhan T, Wu ST. Augmented reality and virtual reality displays: emerging technologies and future perspectives. *Light Sci Appl.* 2021;10:216. DOI PubMed PMC
47. Gao Y, Zhao Y, Xie L, Zheng G. A projector-based augmented reality navigation system for computer-assisted surgery. *Sensors.* 2021;21:2931. DOI PubMed PMC
48. Longo UG, De Salvatore S, Candela V, et al. Augmented reality, virtual reality and artificial intelligence in orthopedic surgery: a systematic review. *Appl Sci.* 2021;11:3253. DOI
49. Ghaednia H, Fourman MS, Lans A, et al. Augmented and virtual reality in spine surgery, current applications and future potentials. *Spine J.* 2021;21:1617-25. DOI
50. Kosterhon M, Gutenberg A, Kantelhardt SR, Archavlis E, Giese A. Navigation and image injection for control of bone removal and osteotomy planes in spine surgery. *Oper Neurosurg.* 2017;13:297-304. DOI PubMed
51. Ma L, Zhao Z, Chen F, Zhang B, Fu L, Liao H. Augmented reality surgical navigation with ultrasound-assisted registration for pedicle screw placement: a pilot study. *Int J Comput Assist Radiol Surg.* 2017;12:2205-15. DOI
52. Morris MX, Fiocco D, Caneva T, Yiapanis P, Orgill DP. Current and future applications of artificial intelligence in surgery: implications for clinical practice and research. *Front Surg.* 2024;11:1393898. DOI PubMed PMC
53. Morley J, Murphy L, Mishra A, Joshi I, Karpathakis K. Governing data and artificial intelligence for health care: developing an international understanding. *JMIR Form Res.* 2022;6:e31623. DOI PubMed PMC
54. Samant S. What role will AI play in resource-poor health care settings? 2019. Available from: <https://www.clinicallab.com/what-role-will-ai-play-in-resource-poor-health-care-settings-407>. [Last accessed on 30 Dec 2024].
55. Mithany RH, Aslam S, Abdallah S, et al. Advancements and challenges in the application of artificial intelligence in surgical arena: a literature review. *Cureus.* 2023;15:e47924. DOI PubMed PMC
56. Amin A, Cardoso SA, Suyambu J, et al. Future of artificial intelligence in surgery: a narrative review. *Cureus.* 2024;16:e51631. DOI PubMed PMC

Review

Open Access



Has artificial intelligence in spine surgery lived up to the hype? A narrative review of recent approaches, current challenges, and the path forward

Vardhaan S. Ambati, Satvir Saggi, Abraham Dada, Nima Alan

Department of Neurological Surgery, University of California San Francisco, San Francisco, CA 94143, USA.

Correspondence to: Dr. Nima Alan, Department of Neurological Surgery, University of California, 505 Parnassus Ave, M779, San Francisco, CA 94143, USA. E-mail: nima.alan@ucsf.edu

How to cite this article: Ambati VS, Saggi S, Dada A, Alan N. Has artificial intelligence in spine surgery lived up to the hype? A narrative review of recent approaches, current challenges, and the path forward. *Art Int Surg.* 2025;5:53-64. <https://dx.doi.org/10.20517/ais.2024.45>

Received: 18 Jun 2024 **First Decision:** 6 Nov 2024 **Revised:** 10 Dec 2024 **Accepted:** 16 Dec 2024 **Published:** 18 Dec 2025

Academic Editor: Andrew Gumbs **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Healthcare applications of artificial intelligence (AI) and machine learning (ML) are currently in a stage of exponential growth; however, their adoption into clinical practice across clinical specialties remains uneven. In spine surgery, the presence of challenging clinical problems, advanced intraoperative technologies, and large multi-center datasets positions the field well for the integration of these technologies into the clinic and operating room (OR). Here, we review recent advances in AI/ML applications in several key domains of spine surgery, identify methodological challenges shared by many approaches, and suggest solutions that may lead to these approaches becoming validated, commercialized tools that can reach clinical practice. Ultimately, we aim for this narrative review to help catalyze further progress in the development and commercialization of AI/ML to benefit future spine patients.

Keywords: Artificial intelligence, machine learning, spine surgery

INTRODUCTION

The past decades have seen spine surgery at the forefront of healthcare innovation, with countless advancements in surgical techniques, robotics, and medical devices positively impacting patient care^[1-3]. In



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



parallel, recent advances in artificial intelligence (AI) and machine learning (ML) have already touched nearly every facet of modern life, from industry and transportation to the arts and music. AI and ML rely on large datasets to recognize patterns in data and, when properly deployed, can perform specialized tasks more quickly and accurately than humans. Due to these promises, spine surgeons are looking to AI/ML to usher in the next generation of technical advancements for their patients.

We are currently at an inflection point for the impact of AI/ML in spine surgery. The number of studies on healthcare AI applications continues to grow exponentially^[4], a trend that is reflected in spine surgery as well^[5]. Furthermore, national funding agencies have established priorities in healthcare AI^[6], and the Food and Drug Administration has gained experience in regulating these tools^[7]. Simultaneously, private venture funding has grown substantially in healthcare AI^[8], but rates of progress have not been equal across applications^[4]. In particular, AI diagnostic tools in radiology and pathology have grown faster than other areas of medicine^[4], such as the surgical specialties. A better understanding of the landscape of AI/ML in spine surgery could help bridge the gap between research and commercialization of such tools.

Currently, disorders of the spine are among the major contributors to both healthcare costs and disability both in the United States and worldwide^[9,10]. Spine surgery is a major contributor to healthcare spending^[11,12], and while safety has improved significantly year over year^[13], when complications do occur, they cause a substantial impact on patient quality of life and healthcare spending^[14]. Due to the wide range of variables that influence patient selection, preoperative planning, intraoperative technique and decision making, opportunities for the potential impact of AI and the possible challenges are high. AI-based tools that could reliably make advances in efficiency, technical proficiency, or complication minimization would have immense clinical and economic impact.

This article provides a targeted primer on AI/ML algorithms and critically reviews select applications of AI/ML to spine surgery. We highlight those that aim to assist in pre-surgical planning, intraoperative assistance, and prediction of postoperative course. These tools span a spectrum of development and commercial deployment, employ a variety of data sources, and interface with clinicians and patients in a number of ways. Through this narrative review, we identify a set of shared challenges facing the field, namely the substantial heterogeneity in patients with spinal disorders, the uncertainty and subjectivity in outcome measures, and the quality and quantity of data available for algorithm development. Finally, we propose solutions to these challenges, which we hope can forecast a path toward incorporating robust AI/ML tools in spine clinics and operating rooms (ORs), thereby achieving the best outcomes for future patients.

AN AI/ML PRIMER FOR SPINE SURGEONS

AI is a group of computational approaches that aim to provide human-level expertise and decision making and predominantly rely on ML, a class of powerful statistical models that recapitulate in silico different facets of human sensory processing and cognition, ranging from vision and language to estimation and prediction tasks^[15,16]. While the technical aspects of AI and ML are beyond the scope of this review, we discuss several key concepts that all spine surgeons should familiarize themselves with, as these technologies continue to play an increasing role in our field.

Traditional ML algorithms adapt structured formulas to relate input and output variables and generate future predictions. Common types of these algorithms include logistic regression, decision trees and random forest, and support vector machines^[15]. They differ based on the types of input and output variables they can handle, as well as their ability to process noisy and non-linear data, which are prevalent in healthcare applications. The other major class of models is known as ‘deep learning’, which accounts for the

many recent groundbreaking applications of AI. These models leverage artificial neural networks, which can learn complex relationships between their complex input data and generate a wide variety of potential outputs, including text, images, and audio^[16]. Compared to traditional ML, deep learning is more powerful but requires massively more data and is specialized for data-intensive tasks such as vision and language processing. On the other hand, traditional ML has straightforward ways for users to understand what variables are most important in making predictions, which is particularly true for the more simple methods^[15]. Efforts to understand how deep learning makes predictions and how it weighs input data present an evolving challenge^[16].

Another key division among AI/ML algorithms is between supervised and unsupervised approaches. Supervised models require training between inputs and desired outputs using labeled data. The process of annotating and curating such datasets can be cumbersome and is particularly challenging in surgical specialties where patient numbers are low and factors such as patient privacy are essential. Unsupervised models such as k-means clustering instead can find patterns inherent in unstructured data; however, they cannot directly make predictions in the same way that supervised algorithms can^[15]. While AI algorithms are advancing at a staggering pace, developing a general framework that outlines both the capabilities and limitations of these models will be critical for spine surgeons in the coming decades.

RECENT AI/ML APPLICATIONS IN SPINE SURGERY

Preoperative planning

Spine surgeons face various clinical and radiographic factors in preoperative planning, which they must parse to make often difficult and subjective decisions about patient selection and surgical approach. AI could augment the ability of clinicians to understand patient disease states, weigh factors such as symptoms and disability, and assess anatomic and pathologic parameters from multimodal imaging.

A number of recent studies have used ML to assist in understanding clinical variability and phenotype in patients undergoing spine surgery. Unsupervised clustering of the clinical metadata of patients with degenerative spondylolisthesis revealed distinct phenotypes of disease severity, which had different levels of postoperative improvement, pain, and satisfaction despite sharing the preoperative severity on imaging^[17]. Clustering has also been shown to disentangle the interactions between patient characteristics and surgical procedures in adult spinal deformity. By inputting a variety of variables including clinical, disability, and spinopelvic parameters, the initial clustering algorithm grouped patients based on age and prior surgery, followed by a second clustering step based on the type of surgery performed, yielding distinct groups that vary in terms of the risk/benefit of surgery^[18]. In spinal deformity, preoperatively oriented algorithms can also forecast fine-grained aspects of postoperative responses to a standardized scoliosis questionnaire^[19]. Together, these studies reveal how AI can uncover patterns in clinical data that could guide preoperative patient counseling or patient selection and maximize quality and value.

Patient imaging is central to preoperative planning in spine surgery and its analysis is one of the most promising applications of AI [Figure 1]. Radiographic analysis includes image segmentation, which refers to the accurate identification and delineation of anatomical structures. Previous work has leveraged deep learning to segment spinal cord structures in a fully automated manner^[20,21], performing better than previous state-of-the-art techniques that did not leverage ML^[22]. Further studies incorporating data from patients with SCI improved on these initial advances. They can capture and identify lesions that correlate with motor scores at admission^[23] and predict thoracolumbar injury classification scores from CT alone, which typically require MR imaging to assess ligamentous integrity^[24]. Other algorithms can accurately segment other relevant anatomic structures, including vertebral bodies and discs^[25,26], as well as paraspinal

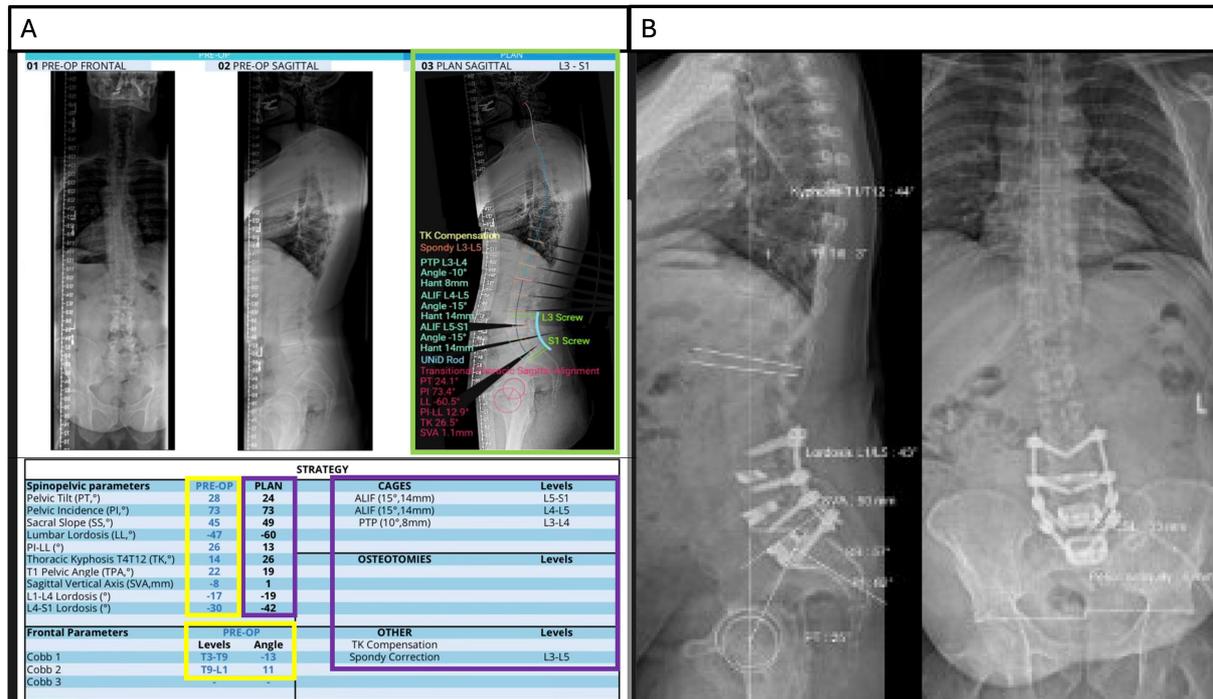


Figure 1. (A) Example of preoperative planning software. Yellow boxes highlight automated spinopelvic parameters and Cobb angle measurements performed by the software. Purple boxes indicate AI-recommended surgical plans and predicted postoperative spinopelvic parameters. The green box demonstrates the predicted postoperative sagittal standing X-ray with the recommended surgical plan; (B) Postoperative standing sagittal and coronal long-cassette radiographs. AI: Artificial intelligence.

musculature^[27]. The extent of osteoporosis^[28] and associated fractures^[29] can also be diagnosed by AI.

Building upon algorithms that segment spinal imaging, others can interpret degrees of neural element compression and estimate parameters of spinal deformity. In particular, these applications are promising as they are tedious, time-consuming, and subject to error and variability when performed by humans. For example, deep learning can estimate the degree of cervical central and foraminal stenosis^[30] and can detect lumbar spondylolisthesis^[31] and other important aspects of degeneration, such as the degree of disc degeneration and central canal stenosis with high accuracy^[32]. For deformity parameter calculation, AI has been applied to calculate coronal^[33,34], sagittal^[35,36], and combined coronal-sagittal parameters^[37]. By incorporating AI into deformity parameter calculation, clinicians can more accurately and efficiently perform both large deformity surgeries and use deformity principles in more limited surgery to ensure patients achieve the best anatomic and physiologic outcomes. These applications represent an ideal area for the strengths of AI to address current challenges in preoperative spine surgical evaluation, and indeed, these technologies have been among the first to reach clinical practice [Table 1].

Intraoperative tools

During surgery, a number of promising AI technologies may help clinicians optimize operative technique and efficiency. Compared to tools designed for pre- or postoperative settings, bringing AI into the OR requires algorithms that can deploy in real time and run on equipment that can interface with the patient, surgeon, and available intraoperative data streams.

Table 1. Summary of studies discussed in the preoperative planning subsection, highlighting key advancements in ML applications for preoperative radiographic and clinical tools

Area of investigation	Selected studies
Utilizing ML clustering methods to identify distinct phenotypes of spinal pathologies, presentation patterns, and radiographic parameters	Chan et al., 2021 ^[17] ; Ames et al., 2019 ^[18]
Preoperative counseling based on patient specific factors	Ames et al., 2019 ^[19]
Automated segmentation of anatomical structures from patient radiographs and films	Spinal cord: Gros et al., 2019 ^[20] ; Jamaludin et al., 2017 ^[21] Vertebral body and discs: Pang et al., 2022 ^[25] ; Pang et al., 2021 ^[26] Paraspinal musculature: Wesselink et al., 2022 ^[27]
Building upon segmentation algorithms to identify clinical correlates (e.g., neurologic exam, osteoporosis, disc degeneration, spinal stenosis)	Neurologic motor scores: McCoy et al., 2019 ^[23] Thoracolumbar injury classification: Doerr et al., 2022 ^[24] Osteoporosis and fractures: Zhang et al., 2020 ^[28] ; Yabu et al., 2021 ^[29] Spinal and foraminal stenosis: Jardon et al., 2023 ^[30] ; Grob et al., 2020 ^[32] Lumbar spondylolisthesis: Trinh et al., 2022 ^[31]
Automated spinopelvic parameter calculation	Berlin et al., 2023 ^[33] ; Wu et al., 2018 ^[34] ; Weng et al., 2019 ^[35] ; Korez et al., 2020 ^[36] ; Galbusera et al., 2019 ^[37]

ML: Machine learning.

Perhaps the main application of AI in the OR is to guide the next generation of surgical navigation, which currently relies on intraoperative radiography and the registration between pre- and intraoperative images. Today's approaches are limited by radiation exposure to the surgical team and patient, delays in operative time caused by acquiring such images, differences in patient anatomy between images acquired in prone and supine positions, and device failure causing navigational inaccuracy.

One promising technology dubbed the Paradigm™ system (Proprio, Seattle, WA) aims to lessen the need for intraoperative CT by using an optical imaging device and computer vision algorithms to align the intraoperative patient with their preoperative imaging, potentially unlocking radiation-free navigation and calculation of spinal anatomic parameters, which could improve safety and speed^[38]. Another competing technology, Flash 7D™ (SeaSpine, Carlsbad, CA), also aims to leverage optical imaging-based navigation powered by deep learning and computer vision. These technologies are being applied to instrumentation in lumbar degenerative disease^[39], pediatric deformity^[40,41], and trauma^[42,43], with potential safety benefits and reduced need for fluoroscopy.

Augmented or mixed reality, in which the surgeon wears goggles that permit them to view the operative field with graphic overlays, also leverages devices with the capability of AI-assisted computer vision [Figure 2]. These approaches are under development in spine surgery^[44] and also promise to help visualize underlying anatomy^[45] to guide pedicle screw placement^[46] and to help perform osteotomies. Early research suggests augmented reality (AR)-assisted pedicle screw placement may compare favorably to freehand techniques^[47] in spinal deformity cases and is also being studied for screw placement in workhorse approaches such as transforaminal lumbar interbody fusion^[48]. Our desire to minimize invasiveness while maximizing visualization of critical structures and accuracy of instrumentation necessitates intraoperative AI to continue making progress [Table 2].

Postoperative prognostication

One of the most common and accessible applications of AI/ML in spine surgery is predicting postoperative outcomes. National datasets from the NIH, American College of Surgeons, and NeuroPoint Alliance, which capture clinical and demographic data, metrics of surgical success, patient-reported outcomes, and complications, can allow clinicians to build models forecasting both perioperative and long-term outcomes

Table 2. Summary of studies discussed in the intraoperative tools subsection

Area of investigation	Selected studies
ML and computer vision tools for optical image-based intraoperative navigation (reduce radiation exposure)	General technique: Berven et al., 2023 ^[38] Lumbar degenerative disease: Abdelrahman et al., 2022 ^[39] Pediatric deformity: Comstock et al., 2023 ^[40] ; Lim et al., 2023 ^[41] Trauma: Yeretsian et al., 2022 ^[42] ; Malacon et al., 2022 ^[43]
Intraoperative AR, VR, and mixed reality tools	Eliahu et al., 2022 ^[44] ; Auloge et al., 2020 ^[45] ; Burström et al., 2019 ^[46] ; Elmi-Terander et al., 2020 ^[47] ; Charles et al., 2021 ^[48]

ML: Machine learning; AR: augmented reality; VR: virtual reality.



Figure 2. Percutaneous lumbar pedicle screw placement assisted by AR headset (yellow arrow). The green box highlights the AR overlay, which displays the screw trajectory, allowing the surgeon to maintain focus on the operative field without needing to look at monitors. AR: Augmented reality.

at the patient level. These methods may help identify risk factors that lead to poor outcomes, allowing surgeons to better select patients and tailor appropriate surgical interventions and postoperative care.

A broad spectrum of ML studies have aimed to predict perioperative patient outcomes and complications and identify variables that most strongly drive these events. ML tends to identify common factors that correlate with perioperative outcome, such as age, functional and nutritional status, BMI, Medicaid status,

intraoperative blood loss, smoking, and preoperative medical comorbidities, but the exact predictors vary by study^[49-53]. Some models tend to predict relatively common events such as postoperative delirium, hospital readmissions, and length of stay, whereas others aim to predict rarer and potentially more catastrophic events such as vascular injury during anterior lumbar surgery^[51]. Across tools that aim to quantify adverse events, the approaches that are trained on large databases, receive external validation and testing, and release their tools as open source or commercialized software are most likely to gain the most traction.

In addition to perioperative complications, ML is also well-suited to predict long-term outcomes. In cervical spondylotic myelopathy, previous studies have accurately predicted outcomes years after surgery from preoperative variables, with simple methods such as logistic regression performing well compared to more advanced methods^[54,55]. By examining feature importance methods in ML algorithms, drivers of long-term outcomes can be better understood. For example, in patients who underwent lumbar fusion, higher leg pain and back pain preoperatively were predictive of improvements in leg and back pain, respectively^[56]. In a separate study of both cervical and thoracolumbar fusion, preoperative axial pain and peripheral pain, nationality, the number of previous spine surgeries, age, type of intervention, preoperative quality-of-life, BMI, number of affected levels, and comorbidity were major predictors of outcome^[57]. Similarly, using preoperative MRIs, one study used neural network-based models to predict postoperative proximal junctional kyphosis (PJK). Analysis of the model found that soft tissue features were the strongest drivers of the accuracy of PJK prediction^[58]. A natural question is to ask: “How valuable are these models?” Indeed, they primarily identify obvious risk factors as drivers of short-term complications (age, sex, comorbidities), and those of long-term outcomes (how much patients stand to gain from their preoperative level of disability). We argue that the key to these models is to be able to quote and counsel patients about risks and outcomes in a patient-specific manner to improve informed consent and shared decision making [Table 3].

CHALLENGES AND OPPORTUNITIES

AI and ML tools throughout the spectrum of spine surgical care hold significant promise to improve patient outcomes; however, those at each point of care have sets of unique challenges. AI/ML focusing on preoperative planning may require prospective studies showing that it improves outcomes to gain traction from physicians and reimbursement from insurance companies. Intraoperative tools and robotics require significant hardware investment and may face regulatory challenges to reach clinical integration, and may encounter resistance from surgeons who fear inefficiencies and potential patient harm associated with early adoption of new technologies^[59,60]. Models that predict postoperative complications and long-term outcomes face difficulty in standardizing outcome metrics and in generalizing across centers^[61]. However, common to all AI/ML tools in spine surgery are several critical challenges, which we detail below, along with our proposed solutions.

Challenge 1: patient and surgical heterogeneity

Our varied clinical and research efforts in spine surgery reflect the immense heterogeneity in the patients we treat. Patients may undergo the same operation for a wide variety of indications, at a wide variety of initial states of health, and similarly, outcomes are driven by a wide variety of physiologic and psychosocial factors. In addition, the same patient with the same pathology may be offered differing surgical plans based on their surgeon’s training and preference. A central challenge in ML is the tradeoff between variables and observations (i.e., patients)^[15]. In spine surgery, where patient variability is high, this limitation means that for models to reach the expert level, they must incorporate both many variables and data from a large number of patients. However, as model complexity increases, the ability to understand such models decreases. To mitigate this tradeoff, it may be most expedient to focus AI development efforts on applications that are specifically tailored to quickly and accurately perform highly specific, otherwise time-

Table 3. Summary of AI/ML studies discussed in the postoperative prognostication subsection

Area of investigation	Selected studies
Perioperative complication prediction and risk stratification	General technique: Berven et al., 2023 ^[38] Lumbar degenerative disease: Abdelrahman et al., 2022 ^[39] Pediatric deformity: Comstock et al., 2023 ^[40] , Lim et al., 2023 ^[41] Trauma: Yeretsian et al., 2022 ^[42] ; Malacon et al., 2022 ^[43]
Long-term outcome prognostication	Eliahu et al., 2022 ^[44] ; Auloge et al., 2020 ^[45] ; Burström et al., 2019 ^[46] ; Elmi-Terander et al., 2020 ^[47] ; Charles et al., 2021 ^[48]

AI: Artificial intelligence; ML: machine learning.

consuming tasks (i.e., image segmentation of the spine or robotic navigation).

Challenge 2: subjective outcome measures

Despite the aforementioned challenges in developing a broad AI understanding of spine surgery arising from patient heterogeneity, one substantial barrier lies in challenges presented by current outcome measures. Many of the endpoints we follow are subjective or are influenced by a wide variety of factors that AI may not be able to accurately capture in an unbiased manner. For example, endpoints such as pain and functional status may be influenced by psychological factors. Endpoints such as the return to work may be influenced by socioeconomic status. Endpoints such as the need for revision surgery may be influenced by many factors, including preoperative comorbidities and postoperative access to care in addition to the surgery itself. Postoperative pain medication use is influenced by preoperative levels of tolerance and patterns of clinical prescription. It is critical that such models and their predictions do not lead clinicians to select patients or surgical approaches in a way that perpetuates present disparities. Solutions to this problem may be to focus on more immediate rather than long-term measures, on quantitative or radiographic endpoints that can be measured in a validated manner, and potentially to use AI and new technologies to develop novel outcome metrics that better capture the impact of spine surgery on patients' lives.

Challenge 3: tradeoffs in data quality and quantity

One of the central principles of ML is that capabilities and performance increase with ever-larger datasets^[15]. In particular, cutting-edge approaches such as deep learning and large language models (the types of models underlying self-driving cars and ChatGPT, respectively) rely on immense amounts of data to tune hundreds of billions of parameters, from which their intelligence emerges^[62]. In spine surgery, large registries such as the Quality Outcomes Database (QOD), British Spine Registry, and International Spine Study Group (ISSG) have aggregated patient data across numerous centers, and the largest ML studies may incorporate thousands of patients. However, these numbers are likely sufficient for certain tasks requiring only simple categorical and numerical variables as inputs rather than complex data such as cross-sectional images, text, and video that require immense amounts of data. Still, healthcare databases often encounter quality issues such as missing or incomplete data and variable practices across the sites where the data were collected. Furthermore, as the number of variables per patient in the database increases, the difficulty of expanding the dataset grows, limiting the number of patients incorporated and increasing the administrative burden on centers that participate.

Due to limitations in data quantity, many studies are validated using withheld patients or cross-validation from the same single-center datasets, which may result in model overfitting and limited clinical utility. Validation using independently collected external datasets will allow for improved assessments of model accuracy and generalizability. Even findings from multi-center studies may be affected by this problem, as the datasets are not completely independent of one another. In addition, some studies used large national datasets that may have limited granularity of clinically relevant variables, potentially limiting their models'

performances. We propose that the standards for multi-center validation and held-out validation and test sets for quantifying performance - standards commonly applied across AI applications - be rigorously applied in AI tools for spine surgery to ensure that published models have the best chance for successful clinical translation.

Future directions

While challenges remain in further integrating AI and ML technologies into spine surgery practice, these technologies have already made an impact on clinical care, operative planning, and procedures in the OR. For AI technologies to continue to develop, the field of spine surgery must make a concerted effort to collect high-quality data in the form of de-identified or HIPAA-compliant large multi-center databases and registries, as these data can be used to fine-tune existing and develop new AI technologies. Future surgical planning and prognostication models should leverage a wide variety of data sources for model training, ranging from demographic and clinical data to patient radiographs and free text from medical records. In addition, surgeons should work closely with industry and academic partners to create robotic and augmented/mixed reality tools. As with all new technology, these efforts will require careful oversight, fine-tuning, and comparison to existing best practices. Should spine surgery as a field successfully apply AI models and tools, our patients stand to benefit the most through patient-specific, data-driven surgical planning tools, increased surgical efficiency, and more accurate short and long-term prognostication.

CONCLUSION

This narrative review highlights a selection of current developments in AI for spine surgery. Despite the challenges discussed in the previous section, AI is already beginning to change how we practice spine surgery. By understanding the current landscape of AI/ML tools across stages of development and clinical scenarios ranging from pre- to intra- and postoperative contexts, we may target our efforts toward incorporating the methods most pertinent to the challenges in our practice. One can easily imagine a near future where AI assists in planning surgical approaches and counseling patients, integrates into intraoperative imaging and navigation systems to enhance anatomical recognition and guide instrumentation, and helps avoid and manage postoperative complications. By highlighting the path forward, we identify strategies that innovation-minded spine surgeons can adopt to expedite the development and clinical translation of these models for the benefit of our patients.

DECLARATIONS

Authors' contributions

Made substantial contributions to the drafting of the article: Ambati VS, Saggi S, Alan N

Made substantial contributions to background research and review of existing literature: Ambati VS, Saggi S, Dada A

Led the review article (supervising/senior author): Alan N

Availability of data and materials

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

Alan N serves as a consultant for Globus, Stryker, and Depuy Synthes. The other authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Cornwall GB, Davis A, Walsh WR, Mobbs RJ, Vaccaro A. Innovation and new technologies in spine surgery, circa 2020: a fifty-year review. *Front Surg.* 2020;7:575318. DOI PubMed PMC
2. Kazemi N, Crew LK, Tredway TL. The future of spine surgery: new horizons in the treatment of spinal disorders. *Surg Neurol Int.* 2013;4:S15-21. DOI PubMed PMC
3. Walker CT, Kakarla UK, Chang SW, Sonntag VKH. History and advances in spinal neurosurgery. *J Neurosurg Spine.* 2019;31:775-85. DOI PubMed
4. Zahlan A, Ranjan RP, Hayes D. Artificial intelligence innovation in healthcare: literature review, exploratory analysis, and future research. *Technol Soc.* 2023;74:102321. DOI
5. Zhou S, Zhou F, Sun Y, et al. The application of artificial intelligence in spine surgery. *Front Surg.* 2022;9:885599. DOI PubMed PMC
6. Suran M. New NIH Program for artificial intelligence in research. *JAMA.* 2022;328:1580. DOI PubMed
7. Clark P, Kim J, Aphinyanaphongs Y. Marketing and US Food and Drug Administration clearance of artificial intelligence and machine learning enabled software in and as medical devices: a systematic review. *JAMA Netw Open.* 2023;6:e2321792. DOI PubMed PMC
8. McNabb NK, Christensen EW, Rula EY, et al. Projected growth in FDA-approved artificial intelligence products given venture capital funding. *J Am Coll Radiol.* 2024;21:617-23. DOI
9. Hoy D, March L, Brooks P, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis.* 2014;73:968-74. DOI
10. Stewart Williams J, Ng N, Peltzer K, et al. risk factors and disability associated with low back pain in older adults in low- and middle-income countries. Results from the WHO Study on Global AGEing and Adult Health (SAGE). *PLoS One.* 2015;10:e0127880. DOI PubMed PMC
11. Missios S, Bekelis K. Hospitalization cost after spine surgery in the United States of America. *J Clin Neurosci.* 2015;22:1632-7. DOI PubMed
12. Parker SL, Chotai S, Devin CJ, et al. Bending the cost curve-establishing value in spine surgery. *Neurosurgery.* 2017;80:S61-9. DOI PubMed
13. Jiang F, Wilson JRF, Badhiwala JH, et al. Quality and safety improvement in spine surgery. *Global Spine J.* 2020;10:17S-28S. DOI PubMed PMC
14. Walid MS, Robinson JS Jr. Economic impact of comorbidities in spine surgery. *J Neurosurg Spine.* 2011;14:318-21. DOI PubMed
15. Dangeti P. Statistics for machine learning. Packt Publishing Ltd; 2017. Available from: <https://books.google.com/books?id=C-DDDwAAQBAJ&printsec=frontcover&hl=zh-CN#v=onepage&q&f=false>. [Last accessed on 28 Dec 2024].
16. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016. Available from: <https://www.deeplearningbook.org/>. [Last accessed on 28 Dec 2024].
17. Chan AK, Wozny TA, Bisson EF, et al. Classifying patients operated for spondylolisthesis: a K-means clustering analysis of clinical presentation phenotypes. *Neurosurgery.* 2021;89:1033-41. DOI
18. Ames CP, Smith JS, Pellisé F, et al; European Spine Study Group, International Spine Study Group. Artificial intelligence based hierarchical clustering of patient types and intervention categories in adult spinal deformity surgery: towards a new classification scheme that predicts quality and value. *Spine* 2019;44:915-26. DOI
19. Ames CP, Smith JS, Pellisé F, et al; European Spine Study Group, International Spine Study Group. Development of predictive models for all individual questions of SRS-22R after adult spinal deformity surgery: a step toward individualized medicine. *Eur Spine J* 2019;28:1998-2011. DOI
20. Gros C, De Leener B, Badji A, et al. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. *Neuroimage.* 2019;184:901-15. DOI PubMed PMC
21. Jamaludin A, Loots M, Kadir T, et al; Genodisc Consortium. ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. *Eur Spine J* 2017;26:1374-83. DOI
22. De Leener B, Cohen-Adad J, Kadoury S. Automatic segmentation of the spinal cord and spinal canal coupled with vertebral labeling. *IEEE Trans Med Imaging.* 2015;34:1705-18. DOI PubMed
23. McCoy DB, Dupont SM, Gros C, et al; TRACK-SCI Investigators. Convolutional neural network-based automated segmentation of

- the spinal cord and contusion injury: deep learning biomarker correlates of motor impairment in acute spinal cord injury. *AJNR Am J Neuroradiol* 2019;40:737-44. DOI PubMed PMC
24. Doerr SA, Weber-Levine C, Hersh AM, et al. Automated prediction of the thoracolumbar injury classification and severity score from CT using a novel deep learning algorithm. *Neurosurg Focus*. 2022;52:E5. DOI
 25. Pang S, Pang C, Su Z, et al. DGMSNet: spine segmentation for MR image by a detection-guided mixed-supervised segmentation network. *Med Image Anal*. 2022;75:102261. DOI
 26. Pang S, Pang C, Zhao L, et al. SpineParseNet: spine parsing for volumetric MR image by a two-stage segmentation framework with semantic image representation. *IEEE Trans Med Imaging*. 2021;40:262-73. DOI
 27. Wesselink EO, Elliott JM, Coppieters MW, et al. Convolutional neural networks for the automatic segmentation of lumbar paraspinous muscles in people with low back pain. *Sci Rep*. 2022;12:13485. DOI PubMed PMC
 28. Zhang B, Yu K, Ning Z, et al. Deep learning of lumbar spine X-ray for osteopenia and osteoporosis screening: a multicenter retrospective cohort study. *Bone*. 2020;140:115561. DOI
 29. Yabu A, Hoshino M, Tabuchi H, et al. Using artificial intelligence to diagnose fresh osteoporotic vertebral fractures on magnetic resonance images. *Spine J*. 2021;21:1652-8. DOI
 30. Jardon M, Tan ET, Chazen JL, et al. Deep-learning-reconstructed high-resolution 3D cervical spine MRI for foraminal stenosis evaluation. *Skeletal Radiol*. 2023;52:725-32. DOI
 31. Trinh GM, Shao HC, Hsieh KL, et al. Detection of lumbar spondylolisthesis from X-ray images using deep learning network. *J Clin Med*. 2022;11:5450. DOI PubMed PMC
 32. Grob A, Loibl M, Jamaludin A, et al. External validation of the deep learning system “SpineNet” for grading radiological features of degeneration on MRIs of the lumbar spine. *Eur Spine J*. 2022;31:2137-48. DOI PubMed
 33. Berlin C, Adomeit S, Grover P, et al. Novel AI-based algorithm for the automated computation of coronal parameters in adolescent idiopathic scoliosis patients: a validation study on 100 preoperative full spine X-rays. *Global Spine J*. 2024;14:1728-37. DOI PubMed PMC
 34. Wu H, Bailey C, Rasoulinejad P, Li S. Automated comprehensive adolescent idiopathic scoliosis assessment using MVC-Net. *Med Image Anal*. 2018;48:1-11. DOI PubMed
 35. Weng CH, Wang CL, Huang YJ, et al. Artificial intelligence for automatic measurement of sagittal vertical axis using ResUNet framework. *J Clin Med*. 2019;8:1826. DOI PubMed PMC
 36. Korez R, Putzier M, Vrtovec T. A deep learning tool for fully automated measurements of sagittal spinopelvic balance from X-ray images: performance evaluation. *Eur Spine J*. 2020;29:2295-305. DOI PubMed
 37. Galbusera F, Niemeyer F, Wilke HJ, et al. Fully automated radiological analysis of spinal disorders and deformities: a deep learning approach. *Eur Spine J*. 2019;28:951-60. DOI
 38. Berven S, Bisson E, Glassman S, et al. Optimizing surgical alignment: intraoperative assessment of alignment using zero radiation, volumetric intelligence. 2023. Available from: https://assets-global.website-files.com/6296c202ecfe835c15f4e757/649d0b69ba75cbd927e0717b_PROPRIO_WhitePaperLayout_R2.pdf. [Last accessed on 28 Dec 2024].
 39. Abdelrahman A, Bangash OK, Bala A. Percutaneous posterior lumbar interbody fusion using optical topographic navigation: operative technique. *Interdiscip Neurosur*. 2022;29:101561. DOI
 40. Comstock CP, Wait E. Novel machine vision image guidance system significantly reduces procedural time and radiation exposure compared with 2-dimensional fluoroscopy-based guidance in pediatric deformity surgery. *J Pediatr Orthop*. 2023;43:e331-6. DOI PubMed PMC
 41. Lim KBL, Yeo ISX, Ng SWL, Pan WJ, Lee NKL. The machine-vision image guided surgery system reduces fluoroscopy time, ionizing radiation and intraoperative blood loss in posterior spinal fusion for scoliosis. *Eur Spine J*. 2023;32:3987-95. DOI PubMed
 42. Yeretsian T, Lai C, Guha D, Ramjst J, Yang VXD. Machine-vision image guided C4-C5 unilateral cervical pedicle screw insertion: case report and review of literature. *AME Case Rep*. 2022;6:9. DOI PubMed PMC
 43. Malacon K, Fatemi P, Zygourakis CC. First reported use of machine vision image guided system for unstable thoracolumbar fusion: technical case report. *Interdiscip Neurosur*. 2022;30:101641. DOI
 44. Eliahu K, Liounakos J, Wang MY. Applications for augmented and virtual reality in robot-assisted spine surgery. *Curr Robot Rep*. 2022;3:33-7. DOI
 45. Auloge P, Cazzato RL, Ramamurthy N, et al. Augmented reality and artificial intelligence-based navigation during percutaneous vertebroplasty: a pilot randomised clinical trial. *Eur Spine J*. 2020;29:1580-9. DOI
 46. Burström G, Buerger C, Hoppenbrouwers J, et al. Machine learning for automated 3-dimensional segmentation of the spine and suggested placement of pedicle screws based on intraoperative cone-beam computer tomography. *J Neurosurg Spine*. 2019;31:147-54. DOI
 47. Elmi-Terander A, Burström G, Nachabé R, et al. Augmented reality navigation with intraoperative 3D imaging vs fluoroscopy-assisted free-hand surgery for spine fixation surgery: a matched-control study comparing accuracy. *Sci Rep*. 2020;10:707. DOI PubMed PMC
 48. Charles YP, Cazzato RL, Nachabe R, Chatterjea A, Steib JP, Gangi A. Minimally invasive transforaminal lumbar interbody fusion using augmented reality surgical navigation for percutaneous pedicle screw placement. *Clin Spine Surg*. 2021;34:E415-24. DOI PubMed
 49. 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. DOI PubMed PMC

50. Khazanchi R, Bajaj A, Shah RM, et al. Using machine learning and deep learning algorithms to predict postoperative outcomes following anterior cervical discectomy and fusion. *Clin Spine Surg*. 2023;36:143-9. DOI
51. Karhade AV, Bongers MER, Groot OQ, et al. Development of machine learning and natural language processing algorithms for preoperative prediction and automated identification of intraoperative vascular injury in anterior lumbar spine surgery. *Spine J*. 2021;21:1635-42. DOI
52. Karabacak M, Margetis K. A machine learning-based online prediction tool for predicting short-term postoperative outcomes following spinal tumor resections. *Cancers*. 2023;15:812. DOI PubMed PMC
53. Han SS, Azad TD, Suarez PA, Ratliff JK. A machine learning approach for predictive models of adverse events following spine surgery. *Spine J*. 2019;19:1772-81. DOI PubMed
54. Park C, Mummaneni PV, Gottfried ON, et al. Which supervised machine learning algorithm can best predict achievement of minimum clinically important difference in neck pain after surgery in patients with cervical myelopathy? A QOD study. *Neurosurg Focus*. 2023;54:E5. DOI
55. Khan O, Badhiwala JH, Witiw CD, Wilson JR, Fehlings MG. Machine learning algorithms for prediction of health-related quality-of-life after surgery for mild degenerative cervical myelopathy. *Spine J*. 2021;21:1659-69. DOI PubMed
56. Rushton AB, Jadhakhan F, Verra ML, et al. Predictors of poor outcome following lumbar spinal fusion surgery: a prospective observational study to derive two clinical prediction rules using British Spine Registry data. *Eur Spine J*. 2023;32:2303-18. DOI
57. Müller D, Haschtmann D, Fekete TF, et al. Development of a machine-learning based model for predicting multidimensional outcome after surgery for degenerative disorders of the spine. *Eur Spine J*. 2022;31:2125-36. DOI
58. Johnson GW, Chanbour H, Ali MA, et al. Artificial intelligence to preoperatively predict proximal junction kyphosis following adult spinal deformity surgery: soft tissue imaging may be necessary for accurate models. *Spine*. 2023;48:1688-95. DOI PubMed PMC
59. Fiani B, De Stefano F, Kondilis A, Covarrubias C, Reier L, Sarhadi K. Virtual reality in neurosurgery: “Can you see it?” *World Neurosurg*. 2020;141:291-8. DOI PubMed
60. Goldberg JL, Härtl R, Elowitz E. Challenges hindering widespread adoption of minimally invasive spinal surgery. *World Neurosurg*. 2022;163:228-32. DOI PubMed
61. Ghanem M, Ghaith AK, El-Hajj VG, et al. Limitations in evaluating machine learning models for imbalanced binary outcome classification in spine surgery: a systematic review. *Brain Sci*. 2023;13:1723. DOI PubMed PMC
62. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. *arXiv*. 2020;arXiv:2005.14165. Available from: <https://doi.org/10.48550/arXiv.2005.14165>. [accessed 28 Dec 2024]

AUTHOR INSTRUCTIONS

1. Submission Overview

Before you decide to publish with us, please read the following items carefully and make sure that you are well aware of Editorial Policies and the following requirements.

1.1 Topic Suitability

The topic of the manuscript must fit the scope of the journal. Please refer to Aims and Scope for more information.

1.2 Open Access and Copyright

The journal adopts Gold Open Access publishing model and distributes content under the Creative Commons Attribution 4.0 International License. Copyright is retained by authors. Please make sure that you are well aware of these policies.

1.3 Publication Fees

Artificial Intelligence Surgery is an open access journal. When a paper is accepted for publication, authors are required to pay Article Processing Charges (APCs) to cover its editorial and production costs. The APC for each submission is \$2600. There are no additional charges based on color, length, figures, or other elements. For more details, please refer to OAE Publication Fees.

1.4 Language Editing

All submissions are required to be presented clearly and cohesively in good English. Authors whose first language is not English are advised to have their manuscripts checked or edited by a native English speaker before submission to ensure the high quality of expression. A well-organized manuscript in good English would make the peer review even the whole editorial handling more smoothly and efficiently.

If needed, authors are recommended to consider the language editing services provided by OAE to ensure that the manuscript is written in correct scientific English before submission. An extra charge is required to enjoy this service. Please visit https://www.oaepublish.com/index/author_services or contact English-Editing@oaepublish.com for more details.

1.5 Work Funded by the National Institutes of Health

If an accepted manuscript was funded by National Institutes of Health (NIH), the author may inform editors of the NIH funding number. The editors are able to deposit the paper to the NIH Manuscript Submission System on behalf of the author.

1.6 Financial Disclosure and Conflict of Interest

Please provide financial disclosure and any conflict of interest on the title page of manuscript.

2. Submission Preparation

2.1 Cover Letter

A cover letter is required to be submitted accompanying each manuscript. It should be concise and explain why the study is significant, why it fits the scope of the journal, and why it would be attractive to readers, *etc.*

Here is a guideline of a cover letter for authors' consideration:

In the first paragraph: include the title and type (e.g., Original Article, Review, *etc.*) of the manuscript, a brief description on the background of the study, the question the author sought out to answer and why;

In the second paragraph: concisely explain what was done, the main findings and why they are significant;

In the third paragraph: indicate why the manuscript fits the Aims and Scope of the journal, and why it would be attractive to readers;

In the fourth paragraph: confirm that the manuscript has not been published elsewhere and is not under consideration of any other journal. All authors have approved the manuscript and agreed on its submission to the journal. Journal's specific requirements have been met if any.

If the manuscript is contributed to a Special Issue, please also mention it in the cover letter.

If the manuscript was presented partly or entirely in a conference, the author should clearly state the background information of the event, including the conference name, time and place in the cover letter.

2.2 Types of Manuscripts

There is no restriction on the number of figures, tables and references, provided that the manuscript is concise and comprehensive. The journal publishes Original Article, Review, Meta-Analysis, Commentary, *etc.* For more details about paper type, please refer to the following table.

Manuscript Type	Definition	Word Limit	Abstract	Keywords	Main Text Structure
Original Article	An Original Article describes detailed results from novel research. All findings are extensively discussed.	5000 max	Structured abstract including Aim, Methods, Results and Conclusion. No more than 250 words.	3-8 keywords	The main content should include four sections: Introduction, Methods, Results and Discussion.
Review	A review article should provide readers with an in-depth understanding of a field by summarizing existing literature, and highlight key gaps and challenges to address future research. A Scoping Review can be submitted as Review article with a similar structure. Authors should adhere to the PRISMA-ScR checklist (available at PRISMA Scoping) and include the completed checklist as supplementary material. A statement about PRISMA compliance and registration details (if applicable) should be included in the Methods section.	7000 max	Unstructured abstract. No more than 250 words..	3-8 keywords	The main text may consist of several sections with unfixed section titles. We suggest that the author include an "Introduction" section at the beginning, several sections with unfixed titles in the middle part, and a "Conclusion" section in the end.
Report/ Guideline	A document developed by an independent panel of experts, usually multidisciplinary, convened to review the research literature for the purpose of advancing the understanding of an issue, procedure, or method.	7000 max	Unstructured abstract. No more than 250 words.	3-8 keywords	The main text may consist of several sections with unfixed section titles. We suggest that the author include an "Introduction" section at the beginning, several sections with unfixed titles in the middle part, and a "Conclusion" section in the end.
Meta-Analysis	A Meta-Analysis is a statistical analysis combining the results of multiple scientific studies. It is often an overview of clinical trials.	5000 max	Structured abstract including Aim, Methods, Results and Conclusion. No more than 250 words.	3-8 keywords	The main content should include four sections: Introduction, Methods, Results and Discussion.
Systematic Review	A Systematic Review collects and critically analyzes multiple research studies, using methods selected before one or more research questions are formulated, and then finding and analyzing related studies and answering those questions in a structured methodology.	3000 max	Structured abstract including Aim, Methods, Results and Conclusion. No more than 250 words.	3-8 keywords	The main content should include four sections: Introduction, Methods, Results and Discussion.
Editorial	An Editorial is a short article describing news about the journal or opinions of senior Editors or the publisher.	1000 max	None required.	None required	/

Letter to Editor	A Letter to Editor is usually an open post-publication review of a paper from its readers, often critical of some aspect of a published paper. Controversial papers often attract numerous Letters to Editor.	1000 max	Unstructured abstract (optional). No more than 250 words.	3-8 keywords (optional)	/
Commentary	A Commentary is to provide comments on a newly published article or an alternative viewpoint on a certain topic.	2500 max	Unstructured abstract. No more than 250 words.	3-8 keywords	/
Perspective	A Perspective provides personal points of view on the state-of-the-art of a specific area of knowledge and its future prospects. Links to areas of intense current research focus can also be made. The emphasis should be on a personal assessment rather than a comprehensive, critical review. However, comments should be put into the context of existing literature. Perspectives are usually invited by the Editors.	2000 max	Unstructured abstract. No more than 150 words.	3-8 keywords	/

2.3 Manuscript Structure

2.3.1 Front Matter

2.3.1.1 Title

The title of the manuscript should be concise, specific and relevant, with no more than 16 words if possible. When gene or protein names are included, the abbreviated name rather than full name should be used.

2.3.1.2 Authors and Affiliations

Authors' full names should be listed. The initials of middle names can be provided. Institutional addresses and email addresses for all authors should be listed. At least one author should be designated as corresponding author. In addition, corresponding authors are suggested to provide their Open Researcher and Contributor ID upon submission. Please note that any change to authorship is not allowed after manuscript acceptance.

2.3.1.3 Abstract

Original research, systematic reviews, and meta-analyses require structured abstracts. The abstract should provide the context or background for the study and should state the study's purpose, basic procedures (selection of study participants, settings, measurements, analytical methods), main findings (giving specific effect sizes and their statistical and clinical significance, if possible), and principal conclusions. It should emphasize new and important aspects of the study or observations, note important limitations, and not overinterpret findings. Clinical trial abstracts should include items that the CONSORT group has identified as essential. It is not allowed to contain results which are not presented and substantiated in the manuscript, or exaggerate the main conclusions. Citations should not be included in the abstract.

2.3.1.4 Graphical Abstract

The graphical abstract is essential as this can catch first view of your publication by readers. We request the authors submit an eye-catching figure during the revision stage. It should summarize the content of the article in a concise graphical form. It is recommended to use it because this can make online articles get more attention. The graphic abstract should be submitted as a separate document in the online submission system along with the revised version. Please provide an image with a minimum of 730 × 1,228 pixels (h × w) or proportionally more. The image should be readable at a size of 7 × 12 cm using a regular screen resolution of 96 dpi. Preferred file types: TIFF, PSD, AI, JPG, JPEG, EPS, PNG, ZIP and PDF files.

2.3.1.5 Keywords

Three to eight keywords should be provided, which are specific to the article, yet reasonably common within the subject discipline.

2.3.2 Main Text

Manuscripts of different types are structured with different sections of content. Please refer to Types of Manuscripts to make sure which sections should be included in the manuscripts.

2.3.2.1 Introduction

Provide a context or background for the study (that is, the nature of the problem and its significance). State the specific purpose or research objective of, or hypothesis tested by, the study or observation. Cite only directly pertinent references, and do not include data or conclusions from the work being reported.

2.3.2.2 Methods

The guiding principle of the Methods section should be clarity about how and why a study was done in a particular way. The Methods section should aim to be sufficiently detailed such that others with access to the data would be able to reproduce the results. In general, the section should include only information that was available at the time the plan or protocol for the study was being written; all information obtained during the study belongs in the Results section. If an organization was paid or otherwise contracted to help conduct the research (examples include data collection and management), then this should be detailed in the methods. The Methods section should include a statement indicating that the research was approved by an independent local, regional or national review body (e.g., ethics committee, institutional review board). If doubt exists whether the research was conducted in accordance with the Helsinki Declaration, the authors must explain the rationale for their approach and demonstrate that the local, regional or national review body explicitly approved the doubtful aspects of the study.

2.3.2.2.1 Selection and Description of Participants

Clearly describe the selection of observational or experimental participants (healthy individuals or patients, including controls), including eligibility and exclusion criteria and a description of the source population. Because the relevance of such variables as age, sex, or ethnicity is not always known at the time of study design, researchers should aim for inclusion of representative populations into all study types and at a minimum provide descriptive data for these and other relevant demographic variables. Comment on how representative the study sample is of the larger population of interest.

Ensure correct use of the terms sex (when reporting biological factors) and gender (identity, psychosocial or cultural factors), and, unless inappropriate, report the sex and/or gender of study participants, the sex of animals or cells, and describe the methods used to determine sex and gender. If the study was done involving an exclusive population, for example in only one sex, authors should justify why. Authors should define how they determined race or ethnicity and justify their relevance. In the case where race or ethnicity was not collected, explain why it was not collected. Race and ethnicity are social and not biological constructs; authors should interpret results associated with race and ethnicity in that context. Authors should use neutral, precise, and respectful language to describe study participants and avoid the use of terminology that might stigmatize participants.

2.3.2.2.2 Technical Information

Specify the study's main and secondary objectives—usually identified as primary and secondary outcomes. Identify methods, equipment (give the manufacturer's name and address in parentheses), and procedures in sufficient detail to allow others to reproduce the results. Give references to established methods, including statistical methods (see below); provide references and brief descriptions for methods that have been published but are not well-known; describe new or substantially modified methods, give the reasons for using them, and evaluate their limitations. Identify precisely all drugs and chemicals used, including generic name(s), dose(s), and route(s) of administration. Identify appropriate scientific names and gene names.

2.3.2.2.3 Statistics

Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to judge its appropriateness for the study and to verify the reported results. When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Avoid relying solely on statistical hypothesis testing, such as P values, which fail to convey important information about effect size and precision of estimates. References for the design of the study and statistical methods should be to standard works when possible (with pages stated). Define statistical terms, abbreviations, and most symbols. Specify the statistical software package(s) and versions used. Distinguish prespecified from exploratory analyses, including subgroup analyses.

2.3.2.3 Results

Present your results in logical sequence in the text, tables, and figures, giving the main or most important findings first. Do not repeat all the data in the tables or figures in the text; emphasize or summarize only the most important observations. Provide data on all primary and secondary outcomes identified in the Methods Section. Extra or supplementary materials and technical details can be placed in an appendix where they will be accessible but will not interrupt the flow of the text, or they can be published solely in the electronic version of the journal.

Give numeric results not only as derivatives (for example, percentages) but also as the absolute numbers from which the derivatives were calculated. Restrict tables and figures to those needed to explain the argument of the paper and to assess supporting data. Use graphs as an alternative to tables with many entries; do not duplicate data in graphs and tables. Avoid nontechnical uses of technical terms in statistics, such as “random” (which implies a randomizing device), “normal,” “significant,” “correlations,” and “sample.”

Separate reporting of data by demographic variables, such as age and sex, facilitate pooling of data for subgroups across studies and should be routine, unless there are compelling reasons not to stratify reporting, which should be explained.

2.3.2.4 Discussion

It is useful to begin the discussion by briefly summarizing the main findings, and explore possible mechanisms or explanations for these findings. Emphasize the new and important aspects of your study and put your findings in the context of the totality of the relevant evidence. State any limitations of your study, and explore the implications of your findings for future research and for clinical practice or policy. Discuss the influence or association of variables, such as sex and/or gender, on your findings, where appropriate, and the limitations of the data. Do not repeat in detail data or other information given in other parts of the manuscript, such as in the Introduction or the Results section. Link the conclusions with the goals of the study but avoid unqualified statements and conclusions not adequately supported by the data. In particular, distinguish between clinical and statistical significance, and avoid making statements on economic benefits and costs unless the manuscript includes the appropriate economic data and analyses. Avoid claiming priority or alluding to work that has not been completed. State new hypotheses when warranted, but label them clearly.

2.3.2.5 Conclusion

It should state clearly the main conclusions and include the explanation of their relevance or importance to the field.

2.3.3 Back Matter

2.3.3.1 Acknowledgments

Anyone who contributed towards the article but does not meet the criteria for authorship, including those who provided professional writing services or materials, should be acknowledged. Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgments section. This section is not added if the author does not have anyone to acknowledge.

2.3.3.2 Authors' Contributions

Each author is expected to have made substantial contributions to the conception or design of the work, or the acquisition, analysis, or interpretation of data, or the creation of new software used in the work, or have drafted the work or substantively revised it.

Please use Surname and Initial of Forename to refer to an author's contribution. For example: made substantial contributions to conception and design of the study and performed data analysis and interpretation: Salas H, Castaneda WV; performed data acquisition, as well as provided administrative, technical, and material support: Castillo N, Young V.

If an article is single-authored, please include "The author contributed solely to the article." in this section.

2.3.3.3 Availability of Data and Materials

In order to maintain the integrity, transparency and reproducibility of research records, authors should include this section in their manuscripts, detailing where the data supporting their findings can be found. Data can be deposited into data repositories or published as supplementary information in the journal. Authors who cannot share their data should state that the data will not be shared and explain it. If a manuscript does not involve such issue, please state "Not applicable." in this section.

2.3.3.4 Financial Support and Sponsorship

All sources of funding for the study reported should be declared. The role of the funding body in the experiment design, collection, analysis and interpretation of data, and writing of the manuscript should be declared. Any relevant grant numbers and the link of funder's website should be provided if any. If the study is not involved with this issue, state "None." in this section.

2.3.3.5 Conflicts of Interest

Authors must declare any potential conflicts of interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. If there are no conflicts of interest, please state "All authors declared that there are no conflicts of interest." in this section. Some authors may be bound by confidentiality agreements. In such cases, in place of itemized disclosures, we will require authors to state "All authors declare that they are bound by confidentiality agreements that prevent them from disclosing their conflicts of interest in this work." If authors are unsure whether conflicts of interest exist, please refer to the "Conflicts of Interest" of OAE Editorial Policies for a full explanation.

2.3.3.6 Ethical Approval and Consent to Participate

Research involving human subjects, human material or human data must be performed in accordance with the Declaration of Helsinki and approved by an appropriate ethics committee. An informed consent to participate in the study should also be obtained from participants, or their parents or legal guardians for children under 16. A statement detailing the name of the ethics committee (including the reference number where appropriate) and the informed consent obtained must appear in the manuscripts reporting such research.

Studies involving animals and cell lines must include a statement on ethical approval. More information is available at Editorial Policies.

If the manuscript does not involve such issue, please state “Not applicable.” in this section.

2.3.3.7 Consent for Publication

Manuscripts containing individual details, images or videos, must obtain consent for publication from that person, or in the case of children, their parents or legal guardians. If the person has died, consent for publication must be obtained from the next of kin of the participant. Manuscripts must include a statement that a written informed consent for publication was obtained. Authors do not have to submit such content accompanying the manuscript. However, these documents must be available if requested. If the manuscript does not involve this issue, state “Not applicable.” in this section.

2.3.3.8 Copyright

Authors retain copyright of their works through a Creative Commons Attribution 4.0 International License that clearly states how readers can copy, distribute, and use their attributed research, free of charge. A declaration “© The Author(s) 2025.” will be added to each article. Authors are required to sign License to Publish before formal publication.

2.3.3.9 References

References should be numbered in order of appearance at the end of manuscripts. In the text, reference numbers should be placed in square brackets and the corresponding references are cited thereafter. If the number of authors is less than or equal to six, we require to list all authors’ names. If the number of authors is more than six, only the first three authors’ names are required to be listed in the references, other authors’ names should be omitted and replaced with “et al.”. Abbreviations of the journals should be provided on the basis of Index Medicus. Information from manuscripts accepted but not published should be cited in the text as “Unpublished material” with written permission from the source. The names of journals should be italicized.

References should be described as follows, depending on the types of works:

Types	Examples
Journal articles by individual authors	Weaver DL, Ashikaga T, Krag DN, et al. Effect of occult metastases on survival in node-negative breast cancer. <i>N Engl J Med</i> . 2011;364:412-21. [PMID: 21247310 DOI: 10.1056/NEJMoa1008108]
Organization as author	Diabetes Prevention Program Research Group. Hypertension, insulin, and proinsulin in participants with impaired glucose tolerance. <i>Hypertension</i> 2002;40:679-86. [PMID: 12411462]
Both personal authors and organization as author	Vallancien G, Emberton M, Harving N, van Moorselaar RJ; Alf-One Study Group. Sexual dysfunction in 1,274 European men suffering from lower urinary tract symptoms. <i>J Urol</i> . 2003;169:2257-61. [PMID: 12771764 DOI: 10.1097/01.ju.0000067940.76090.73]
Journal articles not in English	Zhang X, Xiong H, Ji TY, Zhang YH, Wang Y. Case report of anti-N-methyl-D-aspartate receptor encephalitis in child. <i>J Appl Clin Pediatr</i> . 2012;27:1903-7. (in Chinese)
Journal articles ahead of print	Odibo AO. Falling stillbirth and neonatal mortality rates in twin gestation: not a reason for complacency. <i>BJOG</i> . 2018; Epub ahead of print [PMID: 30461178 DOI: 10.1111/1471-0528.15541]
Books	Sherlock S, Dooley J. Diseases of the liver and billiary system. 9th ed. Oxford: Blackwell Sci Pub; 1993. pp. 258-96.
Book chapters	Meltzer PS, Kallioniemi A, Trent JM. Chromosome alterations in human solid tumors. In: Vogelstein B, Kinzler KW, Editors. The genetic basis of human cancer. New York: McGraw-Hill; 2002. pp. 93-113.
Online resource	FDA News Release. FDA approval brings first gene therapy to the United States. Available from: https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm574058.htm . [Last accessed on 30 Oct 2017]
Conference proceedings	Harnden P, Joffe JK, Jones WG, Editors. Germ cell tumours V. Proceedings of the 5th Germ Cell Tumour Conference; 2001 Sep 13-15; Leeds, UK. New York: Springer; 2002.
Conference paper	Christensen S, Oppacher F. An analysis of Koza's computational effort statistic for genetic programming. In: Foster JA, Lutton E, Miller J, Ryan C, Tettamanzi AG, Editors. Genetic programming. EuroGP 2002: Proceedings of the 5th European Conference on Genetic Programming; 2002 Apr 3-5; Kinsdale, Ireland. Berlin: Springer; 2002. pp. 182-91.
Unpublished material	Tian D, Araki H, Stahl E, Bergelson J, Kreitman M. Signature of balancing selection in Arabidopsis. <i>Proc Natl Acad Sci U S A</i> . Forthcoming 2002.
Thesis or dissertation	Cable ML. Life in extreme environments: lanthanide-based detection of bacterial spores and other sensor design pursuits. Ph.D. Dissertation, California Institute of Technology, Pasadena, CA, 2010. Available from https://resolver.caltech.edu/CaltechTHESIS:05102010-145436548 [accessed 29 December 2023].
Thesis or dissertation in print	Enander RT. Lead particulate and methylene chloride risks in automotive refinishing. Ph.D. Thesis, Tufts University, Medford, MA, 2001.
Website	World Health Organization Home Page. Available from https://www.who.int/ [accessed 29 December 2023].

Newspaper	Beauge J. School district sued over burns girl suffered during chemistry class demonstration. PennLive (Harrisburg, PA), November 17, 2018, updated November 17, 2018. Available from https://www.pennlive.com/news/2018/11/school_district_sued_over_burn.html [accessed 29 December 2023].
Database	SpectraBase. Bio-Rad Laboratories. Available from https://spectrabase.com/ [accessed 29 December 2023].
Patent	Stern MK, Cheng BKM. Process for preparing N-(p-nitroaryl)amides via reaction of nitrobenzene with nitriles. US 5380946, 1995.
Preprint	Yamamoto TS, Inui R, Tada Y, Yokoyama S. Prospects of detection of subsolar mass primordial black hole and white dwarf binary mergers. arXiv 2004;arXiv:2401.00044. Available from https://arxiv.org/abs/2401.00044 [accessed 29 December 2023].

For other types of references, please refer to U.S. National Library of Medicine.

The journal also recommends that authors prepare references with a bibliography software package, such as EndNote to avoid typing mistakes and duplicated references.

2.3.3.10 Supplementary Materials

Additional data and information can be uploaded as Supplementary Materials to accompany the manuscripts. The supplementary materials will also be available to the referees as part of the peer-review process. Any file format is acceptable, such as data sheet (word, excel, csv, cdx, fasta, pdf or zip files), presentation (powerpoint, pdf or zip files), image (cdx, eps, jpeg, pdf, png or tiff), table (word, excel, csv or pdf), audio (mp3, wav or wma) or video (avi, divx, flv, mov, mp4, mpeg, mpg or wmv). All information should be clearly presented. Supplementary materials should be cited in the main text in numeric order (e.g., Supplementary Figure 1, Supplementary Figure 2, Supplementary Table 1, Supplementary Table 2, *etc.*). The style of supplementary figures or tables complies with the same requirements on figures or tables in main text. Videos and audios should be prepared in English, and limited to a size of 500 MB.

2.4 Manuscript Format

2.4.1 File Format

Manuscript files can be in DOC and DOCX formats and should not be locked or protected.

2.4.2 Length

The word limit is specified in the item “Types of Manuscripts”. There are no restrictions on number of figures or amount of supporting documents. Authors are encouraged to present and discuss their findings concisely.

2.4.3 Language

Manuscripts must be written in English.

2.4.4 Multimedia Files

The journal supports manuscripts with multimedia files. The requirements are listed as follows:

Video or audio files are only acceptable in English. The presentation and introduction should be easy to understand. The frames should be clear, and the speech speed should be moderate.

A brief overview of the video or audio files should be given in the manuscript text.

The video or audio files should be limited to a size of up to 500 MB.

Please use professional software to produce high-quality video files, to facilitate acceptance and publication along with the submitted article. Upload the videos in mp4, wmv, or rm format (preferably mp4) and audio files in mp3 or wav format.

2.4.5 Figures

Figures should be cited in numeric order (e.g., Figure 1, Figure 2) and placed after the paragraph where it is first cited;

Figures can be submitted in format of tiff, psd, AI or jpeg, with resolution of 300-600 dpi;

Figure caption is placed under the Figure;

Diagrams with describing words (including, flow chart, coordinate diagram, bar chart, line chart, and scatter diagram, *etc.*) should be editable in word, excel or powerpoint format. Non-English information should be avoided;

Labels, numbers, letters, arrows, and symbols in figure should be clear, of uniform size, and contrast with the background; Symbols, arrows, numbers, or letters used to identify parts of the illustrations must be identified and explained in the legend;

Internal scale (magnification) should be explained and the staining method in photomicrographs should be identified;

All non-standard abbreviations should be explained in the legend;

Permission for use of copyrighted materials from other sources, including re-published, adapted, modified, or partial figures and images from the internet, must be obtained. It is authors’ responsibility to acquire the licenses, to follow any citation instruction requested by third-party rights holders, and cover any supplementary charges.

2.4.6 Tables

Tables should be cited in numeric order and placed after the paragraph where it is first cited;
The table caption should be placed above the table and labeled sequentially (e.g., Table 1, Table 2);
Tables should be provided in editable form like DOC or DOCX format (picture is not allowed);
Abbreviations and symbols used in table should be explained in footnote;
Explanatory matter should also be placed in footnotes;
Permission for use of copyrighted materials from other sources, including re-published, adapted, modified, or partial tables from the internet, must be obtained. It is authors' responsibility to acquire the licenses, to follow any citation instruction requested by third-party rights holders, and cover any supplementary charges.

2.4.7 Abbreviations

Abbreviations should be defined upon first appearance in the abstract, main text, and in figure or table captions and used consistently thereafter. Non-standard abbreviations are not allowed unless they appear at least three times in the text. Commonly-used abbreviations, such as DNA, RNA, ATP, *etc.*, can be used directly without definition. Abbreviations in titles and keywords should be avoided, except for the ones which are widely used.

2.4.8 Italics

General italic words like *vs.*, *et al.*, *etc.*, *in vivo*, *in vitro*; *t* test, *F* test, *U* test; related coefficient as *r*, sample number as *n*, and probability as *P*; names of genes; names of bacteria and biology species in Latin.

2.4.9 Units

SI Units should be used. Imperial, US customary and other units should be converted to SI units whenever possible. There is a space between the number and the unit (i.e., 23 mL). Hour, minute, second should be written as h, min, s.

2.4.10 Numbers

Numbers appearing at the beginning of sentences should be expressed in English. When there are two or more numbers in a paragraph, they should be expressed as Arabic numerals; when there is only one number in a paragraph, number < 10 should be expressed in English and number > 10 should be expressed as Arabic numerals. 12345678 should be written as 12,345,678.

2.4.11 Equations

Equations should be editable and not appear in a picture format. Authors are advised to use either the Microsoft Equation Editor or the MathType for display and inline equations.

2.5 Submission Link

Submit an article via <https://www.oaecenter.com/login?JournalId=ais>.

3. Research and Publication Ethics

3.1 Research Involving Human Subjects

All studies involving human subjects must be in accordance with the Helsinki Declaration and seek approval to conduct the study from an independent local, regional, or national review body (e.g., ethics committee, institutional review board, *etc.*). Such approval, including the names of the ethics committee, institutional review board, *etc.*, must be listed in a declaration statement of Ethical Approval and Consent to Participate in the manuscript. If the study is judged exempt from ethics approval, related information (e.g., name of the ethics committee granting the exemption and the reason for the exemption) must be listed. Further documentation on ethics should also be prepared, as editors may request more detailed information. Manuscripts with suspected ethical problems will be investigated according to COPE Guidelines.

3.1.1 Consent to Participate

For all studies involving human subjects, informed consent to participate in the studies must be obtained from participants, or their parents or legal guardians for children under 16. Statements regarding consent to participate should be included in a declaration statement of Ethical Approval and Consent to Participate in the manuscript. If informed consent is not required, the name of the ethics committee granting the exemption and the reason for the exemption must be listed. If any ethical violation is found at any stage of publication, the issue will be investigated seriously based on COPE Guidelines.

3.1.2 Consent for Publication

All articles published by OAE are freely available on the Internet. All manuscripts that include individual participants' data in any form (i.e., details, images, videos, *etc.*) will not be published without Consent for Publication obtained from that person(s), or for children, their parents or legal guardians. If the person has died, Consent for Publication must be obtained from the next of kin. Authors must add a declaration statement of Consent for Publication in the manuscript, specifying written informed consent for publication has been obtained.

3.1.3 Ethical Approval and Informed Consent for Retrospective/Database Studies

Researchers must confirm they have obtained ethical approval from ethical review boards to perform the study, as well as permission from the dataset owner to use the information in databases for the purposes of the research they are performing. If permission to use information from a database is not required (e.g., it is publicly available and unrestricted re-use is permitted under an open license), a statement explaining this must be included in the manuscript. For studies which ethics approval has been waived, authors must state clearly in the manuscript and provide brief details of the waive policy. The statement should include details of the policies under which the waive was granted.

Authors must keep data anonymized. If participants' details are not to be anonymized, authors must ensure that written informed consent, including consent for publication, was obtained from each participant, and consent statement must be included in the manuscript.

3.1.4 Ethical Approval and Informed Consent for Survey Studies

Researchers must ensure the participant's right to confidentiality has been considered, and they must inform all participants about the aims of the research and if there are any possible risks, and how the collecting data is being stored. The voluntary consent to participate of participants should be recorded and any legal requirements on data protection should be adhered to. Same with all research studies, ethics approval from IRB/local ethics committee for survey studies must be obtained before performing study. If ethics approval for certain survey study is not required, authors must include a statement to explain this clearly in the manuscript.

3.1.5 Trial Registration

OAE requires all authors to register all relevant clinical trials that are reported in manuscripts submitted. OAE follows the World Health Organization (WHO)'s definition of clinical trials: "A clinical trial is any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes. Interventions include but are not restricted to drugs, cells, other biological products, surgical procedures, radiologic procedures, devices, behavioral treatments, process-of-care changes, preventive care, *etc.*"

In line with International Committee of Medical Journal Editors (ICMJE) recommendation, OAE requires the registration of clinical trials in a public trial registry at or before the time of first patient enrollment. OAE accepts publicly accessible registration in any registry that is a primary register of the WHO International Clinical Trials Registry Platform or in ClinicalTrials.gov. The trial registration number should be listed at the end of the Abstract section.

Secondary data analyses of primary (parent) clinical trials should not be registered as a new clinical trial, but rather reference the trial registration number of the primary trial.

Editors of OAE journals will consider carefully whether studies failed to register or had an incomplete trial registration. Because of the importance of prospective trial registration, if there is an exception to this policy, trials must be registered and the authors should indicate in the publication when registration was completed and why it was delayed. Editors will publish a statement indicating why an exception was allowed. Please note such exceptions should be rare, and authors failing to prospectively register a trial risk its inadmissibility to OAE journals.

Authors who are not sure whether they need trial registration may refer to ICMJE FAQs for further information.

3.2 Research Involving Animals

Experimental research on animals should be approved by an appropriate ethics committee and must comply with institutional, national, or international guidelines. OAE encourages authors to comply with the AALAS Guidelines, the ARRIVE Guidelines, and/or the ICLAS Guidelines, and obtain prior approval from the relevant ethics committee. Manuscripts must include a statement indicating that the study has been approved by the relevant ethical committee and the whole research process complies with ethical guidelines. If a study is granted an exemption from requiring ethics approval, the name of the ethics committee granting the exemption and the reason(s) for the exemption should be detailed. Editors will take account of animal welfare issues and reserve the right to reject a manuscript, especially if the research involves protocols that are inconsistent with commonly accepted norms of animal research.

3.3 Research Involving Cell Lines

Authors must describe what cell lines are used and their origin so that the research can be reproduced. For established cell lines, the provenance should be stated and references must also be given to either a published paper or to a commercial source. For *de novo* cell lines derived from human tissue, appropriate approval from an institutional review board or equivalent ethical committee, and consent from the donor or next of kin, should be obtained. Such statements should be listed on the Declaration section of Ethical Approval and Consent to Participate in the manuscript.

Further information is available from the International Cell Line Authentication Committee (ICLAC). OAE recommends that authors check the NCBI database for misidentification and contamination of human cell lines.

3.4 Publication Ethics Statement

The Editors of this journal enforce a rigorous peer-review process together with strict ethical policies and standards to guarantee to add high-quality scientific works to the field of scholarly publication. Unfortunately, cases of plagiarism,

data falsification, image manipulation, inappropriate authorship credit, and the like, do arise. The Editors of *Artificial Intelligence Surgery* take such publishing ethics issues very seriously and are trained to proceed in such cases with zero tolerance policy.

Authors wishing to publish their papers in *Artificial Intelligence Surgery* must abide to the following:

The author(s) must disclose any possibility of a conflict of interest in the paper prior to submission.

The authors should declare that there is no academic misconduct in their manuscript in the cover letter.

Authors should accurately present their research findings and include an objective discussion of the significance of their findings.

Data and methods used in the research need to be presented in sufficient detail in the manuscript so that other researchers can replicate the work.

Authors should provide raw data if referees and the Editors of the journal request.

Simultaneous submission of manuscripts to more than one journal is not tolerated.

Republishing content that is not novel is not tolerated (for example, an English translation of a paper that is already published in another language will not be accepted).

The manuscript should not contain any information that has already been published. If you include already published figures or images, please get the necessary permission from the copyright holder to publish under the CC-BY license.

Plagiarism, data fabrication and image manipulation are not tolerated.

Plagiarism is not acceptable in OAE journals.

Plagiarism involves the inclusion of large sections of unaltered or minimally altered text from an existing source without appropriate and unambiguous attribution, and/or an attempt to misattribute original authorship regarding ideas or results, and copying text, images, or data from another source, even from your own publications, without giving credit to the source.

As to reusing the text that is copied from another source, it must be between quotation marks and the source must be cited. If a study's design or the manuscript's structure or language has been inspired by previous studies, these studies must be cited explicitly.

If plagiarism is detected during the peer-review process, the manuscript may be rejected. If plagiarism is detected after publication, we may publish a Correction or retract the paper.

Falsification is manipulating research materials, equipment, or processes, or changing or omitting data or results so that the findings are not accurately represented in the research record.

Image files must not be manipulated or adjusted in any way that could lead to misinterpretation of the information provided by the original image.

Irregular manipulation includes: introduction, enhancement, moving, or removing features from the original image; grouping of images that should be presented separately, or modifying the contrast, brightness, or color balance to obscure, eliminate, or enhance some information.

If irregular image manipulation is identified and confirmed during the peer-review process, we may reject the manuscript.

If irregular image manipulation is identified and confirmed after publication, we may publish a Correction or retract the paper.

OAE reserves the right to contact the authors' institution(s) to investigate possible publication misconduct if the Editors find conclusive evidence of misconduct before or after publication. OAE has a partnership with iThenticate, which is the most trusted similarity checker. It is used to analyze received manuscripts to avoid plagiarism to the greatest extent possible. When plagiarism becomes evident after publication, we will retract the original publication or require modifications, depending on the degree of plagiarism, context within the published article, and its impact on the overall integrity of the published study. Journal Editors will act under the relevant COPE Guidelines.

4. Authorship

Authorship credit of OAE journals should be solely based on substantial contributions to a published study, as specified in the following four criteria:

1. Substantial contributions to the conception or design of the work, or the acquisition, analysis, or interpretation of data for the work;
2. Drafting the work or revising it critically for important intellectual content;
3. Final approval of the version to be published;
4. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

All those who meet these criteria should be identified as authors. Authors must specify their contributions in the section Authors' Contributions of their manuscripts. Contributors who do not meet all the four criteria (like only involved in a

acquisition of funding, general supervision of a research group, general administrative support, writing assistance, technical editing, language editing, proofreading, *etc.*) should be acknowledged in the section of Acknowledgement in the manuscript rather than being listed as authors.

If a large multiple-author group has conducted the work, the group ideally should decide who will be authors before the work starts and confirm authors before submission. All authors of the group named as authors must meet all the four criteria for authorship.

AI and AI-assisted technologies should not be listed as an author or co-author.

5. Reviewers Exclusions

You are welcome to exclude a limited number of researchers as potential Editors or reviewers of your manuscript. To ensure a fair and rigorous peer-review process, we ask that you keep your exclusions to a maximum of three people. If you wish to exclude additional referees, please explain or justify your concerns—this information will be helpful for Editors when deciding whether to honor your request.

6. Editors and Journal Staff as Authors

Editorial independence is extremely important and OAE does not interfere with editorial decisions.

Editorial staff or Editors shall not be involved in processing their own academic work. Submissions authored by editorial staff/Editors will be assigned to at least three independent outside reviewers. Decisions will be made by other Editorial Board members who do not have conflict of interests with the author. Journal staffs are not involved in the processing of their own work submitted to any OAE journals.

7. Policy of the Use of AI and AI-assisted Technologies in Scientific Writing

Generative AI and AI-assisted technologies (e.g., large language models) are expected to be increasingly used to create content. In the writing process of manuscripts, using AI and AI-assisted technologies to complete key researcher work, such as producing scientific insights, analyzing and interpreting data or drawing scientific conclusions, is not allowed, and they should only be used to improve the readability and language of manuscripts.

AI and AI-assisted technologies should be used under human control and supervision as they may generate incorrect or prejudiced output, and they should not be listed as an author or co-author, nor cited as an author.

The use of AI and AI-assisted technologies should be disclosed by authors in their manuscripts, and a statement will be required in the final publication.

OAE will keep monitoring the development and adjust the policy when necessary.

8. Conflict of Interests

OAE journals require authors to declare any possible financial and/or non-financial conflicts of interest at the end of their manuscript and in the cover letter, as well as confirm this point when submitting their manuscript in the submission system. If no conflicts of interest exist, authors need to state “The authors declare no conflicts of interest”. We also recognize that some authors may be bound by confidentiality agreements, in which cases authors need to state “The authors declare that they are bound by confidentiality agreements that prevent them from disclosing their competing interests in this work”.

9. Editorial Process

9.1 Initial check

9.1.1 Initial manuscript check

New submissions are initially checked by the Managing Editor from the perspectives of originality, suitability, structure and formatting, conflicts of interest, background of authors, etc. Poorly prepared manuscripts may be rejected at this stage. If your manuscript does not meet one or more of these requirements, we will return it for further revisions.

9.1.2 Publishing ethics

All manuscripts submitted to *Artificial Intelligence Surgery* are screened using iThenticate powered by CrossCheck to identify any plagiarized content. Your study must also meet all ethical requirements as outlined in our Editorial Policies. If the manuscript does not pass any of these checks, we may return it to you for further revisions or decline to consider your study for publication.

9.2 Editorial assessment

Once your manuscript has passed the initial check, it will be assigned to an Assistant Editor, and then the Editor-in-Chief, or an Associate Editor in the case of a conflict of interest, will be notified of the submission and invited to review.

Regarding Special Issue paper, after passing the initial check, the manuscript will be successively assigned to an Assistant Editor, Guest Editor, and then to the Editor-in-Chief, or an Associate Editor in the case of conflict of interest for the Editor-in-Chief to review. The Editor-in-Chief, or the Associate Editor may reject manuscripts that they deem highly unlikely to pass peer review without further consultation. Once your manuscript has passed the editorial assessment, the Assistant Editor will start to organize peer-review.

9.3 Process

Artificial Intelligence Surgery operates a single-blind review process. The technical quality of the research described in the manuscript is assessed by a minimum of three independent expert reviewers. The Editor-in-Chief is responsible for the final decision regarding acceptance or rejection of the manuscript.

9.4 Decisions

Your research will be judged on technical soundness only, not on its perceived impact as judged by editors or referees. There are three possible decisions: Accept (your study satisfies all publication criteria), Invitation to Revise (more work is required to satisfy all criteria), and Reject (your study fails to satisfy key criteria and it is highly unlikely that further work can address its shortcomings).

10. Contact Us

Managing Editor

Zoey Han

Email: editorialoffice@aisjournal.net

Locations

Los Angeles Office

245 E Main Street, ste122, Alhambra, CA 91801, USA

Tel: +1 323 9987086

Xi'an Office

Suite 1504, Tower A, Xi'an National Digital Publishing Base, No. 996 Tiangu 7th Road, Gaoxin District, Xi'an 710077, Shaanxi, China

Tel: +86 (0)29 8954 0089



*Artificial Intelligence Surgery
(AIS)*

Los Angeles Office
245 E Main Street ste122, Alhambra,
CA 91801, USA
Tel: +1 323 9987086
E-mail: editorialoffice@aisjournal.net
Website: www.oaepublish.com/ais

