Review

Check for updates

Applications and quality assurance of artificial intelligence in adult spinal deformity surgery

Hafthor Sigurdarson[1], Aditya Joshi[2], Aria Mohebi[2], Hamid Hassanzadeh[2]

[1]Department of Orthopaedic Surgery, Landspitali - The National University Hospital of Iceland, Reykjavík 108, Iceland.
[2]Department of Orthopaedic Surgery, Johns Hopkins Medicine, Bethesda, MD 20817, USA.

Correspondence to: Aditya Joshi, Department of Orthopaedic Surgery, Johns Hopkins Medicine, National Capital Region, 6420 Rockledge Dr, Suite 2200, Bethesda, MD 20817, USA. E-mail: adityajoshiprof@gmail.com

Abstract

Artificial intelligence (AI) is reshaping healthcare, particularly within the realm of spinal surgery, enhancing diagnostics, treatment, and patient management. AI is not only enhancing the technical aspects of spinal surgery but also revolutionizing patient care through personalized management, setting a new standard within the field. This computational renaissance has received increasing attention from providers and regulatory bodies to ensure novel technologies are being safely and effectively used. This review explores contemporary uses of AI in adult spinal deformity (ASD) surgery and the extent of their validation. Given the increasing complexity of ASD surgery and the expanding capabilities of AI, this review is essential to synthesize current applications, evaluate methodological strengths and limitations, and highlight future research opportunities in this evolving field.

Keywords: Artificial intelligence, spine surgery, machine learning, adult spinal deformity

INTRODUCTION

In recent years, the application of artificial intelligence (AI) in healthcare has initiated a monumental shift in the methodology used to diagnose patients[1,2], assess prognosis[3], and provide therapies[4-6]. AI has developed into an essential tool for evaluating large patient datasets[7-9], deriving insightful conclusions, and directing medical decisions through advanced computational techniques[10,11]. Its capacity to transform the

delivery of healthcare encompasses a wide range of medical specialties.

Spine surgery is a standout candidate for the applications of AI. Adult spinal deformity (ASD) is a branch of spinal surgery that encompasses a variety of conditions involving an abnormal curvature or alignment of the spine in adult patients. These deformities can arise from various pathologies, including degenerative diseases like arthritis, the progression of a pre-existing condition that was present but stable during childhood (like scoliosis or kyphosis), or due to the effects of trauma or previous spinal surgery. ASD particularly resonates with this technological integration because of the plethora of clinical presentations and their heavily debated management[12,13]. In prior literature, spinal surgeons have attempted to develop models to eliminate some subjective decision making, only to discover the nonlinearity of ASD cannot be tethered to a one-size-fits-all approach[14,15]. Given this complexity, integrating AI into spinal surgery is crucial to developing new ways to improve surgical interventions' accuracy, effectiveness, and results for patients with ASD[16-18]. AI is revolutionizing the management of spinal pathologies through predictive models, advancing surgical technologies, and enhanced therapeutic decision making.

A growing body of literature acknowledges AI is a turning point in the field of spinal surgery[19]. It brings the potential to improve clinical results, surgical technique, and patient care. Through the utilization of sophisticated computational techniques, predictive modeling, and surgical advancements, AI provides surgeons with tailored knowledge, accurate instruments, and research-backed approaches to effectively manage the complex terrain of adult spinal deformities and improve the quality of spine care in the modern healthcare environment. The purpose of this review is to outline the current applications of AI in spinal surgery and address their limitations. We aim to provide a commentary on the generalizability of these models and the validity of their performance. This narrative review was conducted by systematically searching PubMed and Embase using terms including "artificial intelligence", "machine learning", "deep learning", and "adult spinal deformity". Articles were selected based on relevance to clinical applications, with a focus on imaging, surgical planning, and predictive modeling. This is a narrative review intended to provide a broad synthesis of current trends and future directions. As such, it does not follow the PRISMA framework, although efforts were made to ensure transparency in the literature search and inclusion criteria. A brief synopsis of reviewed papers is included in Table 1.

IMAGING

The value of AI in spine imaging has increased tremendously over the last few years. Researchers believe that radiologists will use AI as a tool to aid the increasing demand for radiological inquiries from clinicians[20]. Cui *et al.* and Galbusera *et al.* emphasize how the application of AI could improve the quality, efficiency, and diagnostics of spine imaging[20,21]. These programs can improve patient satisfaction by giving faster and more reliable answers, while also helping doctors effectively cut down working time.

Numerous studies have looked at using model-based approaches for radiographs, CT, and MRI readings of spine images. Galbusera *et al.* showed promise for future usage of deep learning (DL) models with biplanar radiographs and a convolutional neural network (CNN) to obtain anatomical parameters, aiding in the interpretation of kyphosis, lordosis, Cobb angle, pelvic incidence, sacral slope, and pelvic tilt[21]. The study incorporated a CNN model with a C++ program to extract the 3D coordinates of the desired landmarks from the 2D radiographs. The model is a combination of a fully convolutional network with a differentiable spatial heatmap, allowing the transformation of graphical representation to be translated into quantitative data. The model training follows a 90:10 training-to-testing ratio, resulting in an overall success rate of landmark placement compared to standard measurements. Regression analysis showed all the spinopelvic parameters within 95% confidence intervals relative to ground truth data from sterEOS software. The

Table 1. AI and ML models used in spinal surgery

Model	Authors	Applications	Benefits	Risks
CNN	Galbusera *et al.*, 2019[21]; Löchel *et al.*, 2024[23]; Jamaludin *et al.*, 2017[26]; Schlemper *et al.*, 2018[28]; Souza *et al.*, 2010[29]; Yang *et al.*, 2018[38]; Chen *et al.*, 2017[39]; Xuan *et al.*, 2023[30]; Wu *et al.*, 2021[37]; Wang *et al.*, 2021[65]; Zhang *et al.*, 2023[66]; Zhao *et al.*, 2023[67]	Imaging, patient benefits, ASD progression	• Layer specialization • Efficient at processing large datasets • Noise reduction • Excellent generalization	• High computational cost • Potential overfitting
DLTG	Xuan *et al.*, 2023[30]	Imaging	• Temporal consistency • Excellent reconstruction quality	• Hyperparameter tuning • Scalability issues
DQN	Ghesu *et al.*, 2016[24]	Imaging	• Can handle high-dimensional input spaces • End-to-end learning	• Overestimation bias • Training instability
MSL	Kelm *et al.*, 2013[25]	Imaging	• More robust and accurate detection results than an exhaustive full space search	• Requires retraining for abnormal cases
GAN	Goodfellow *et al.*, 2020[27]; Yang *et al.*, 2018[38]	Imaging	• Improved image quality • Noise reduction • High-quality image generation with less data	• Generative models can be unpredictable • Requires careful tuning
RNN	Sri Lalitha *et al.*, 2023[31]; Nimal *et al.*, 2023[32]	Patient benefits	• Excellent handling of sequential data • Different architectures enable better performance	• Requires large datasets for training • Vanishing gradients
KNN	Sri Lalitha *et al.*, 2023[31]; Nimal *et al.*, 2023[32]	Patient benefits	• No assumption of data • Capture complex variables without defining separate model	• Does not provide relative importance of each predictor • Does not create a generalized separable model
ANN	Kim *et al.*, 2018[40]; Kuris *et al.*, 2021[41]; Hopkins *et al.*, 2019[42]; De la Garza Ramos *et al.*, 2022[63]	Risk calculator and decision-making tools	• Automatic feature extraction • Continuous learning	• Potential for overfitting • Requires large datasets and computational cost
Hierarchical clustering model	Ames *et al.*, 2019[43]	Risk calculator and decision-making tools	• Flexibility with different distance metrics	• Sensitivity to noise and outliers • Irreversible merge/split decisions
Random forest	Durand *et al.*, 2018[44]; Raman *et al.*, 2020[61]	Perioperative applications	• Minimal preprocessing • Parallel processing • Robust against overfitting	• Difficulty with imbalanced datasets • Interpretability can be challenging
SVM	Bissonnette *et al.*, 2019[55]	Surgical applications	• Kernel functions to handle non-linear data • Clear margins of separation	• Requires labeled training data • Can be sensitive to outliers
Geometric modeling	Klinder *et al.*, 2009[22]	Imaging	• Precise and accurate representation of shapes • Realistic rendering	• High resource requirements • Manual ground truth comparison required

AI: Artificial intelligence; ML: machine learning; CNN: convolutional neural network; ASD: adult spinal deformity; DLTG: dictionary learning with temporal gradient; DQN: deep Q-network; MSL: marginal space learning; GAN: generative adversarial network; RNN: recurrent neural network; KNN: K-nearest neighbors; ANN: artificial neural network; SVM: support vector machine.

authors were optimistic about the potential diagnostic capacity of this model; however, they also cautioned about the need for further training.

In 2013, Klinder *et al.* used a range of models [Geometric modeling, deformable models, curved planar reformation, generalized Hough transform (GHT) models, statistical models of shape, gradient, and appearance, 3-D deformable model approach, appearance model] to extract the spine curvature, along with vertebrae detection/identification/segmentation of CT scans[22]. The variety of scans tested included: cervical, thorax, lumbar, and whole spine images from multiple institutions. Pathological cases of scoliosis, kyphosis, compression fractures, postoperative pedicle screws, and arthritic spines were included in the dataset to diversify the training. A clinician manually created the ground truth comparison to ensure accuracy. The models were successful in 56/64 cases, achieving a mean point-to-surface error of 1.12 ± 1.04 mm. The identification rate for single vertebrae was 70%, increasing with the number of visible vertebrae and reaching 100% when 16 or more vertebrae were present in the image.

Parameters that have gained higher importance in the last decade are sagittal balance and spinopelvic angles. Until recently, no single model algorithm had been able to automatically analyze these sagittal balance parameters. However, in early 2024, Löchel *et al.* published a study analyzing 141 patients with ASD, both preoperatively and postoperatively, using a landmark detection algorithm with a high correlation coefficient ranging from 0.71 to 0.9[23]. The model employs a Mask Region-CNN to segment the images, with preprocessing steps that highlight bony structures and adjust image quality for analysis. Relevant anatomical landmarks are then identified, primarily on the sacrum and L1 body, and a regression line is fitted through the detected points. To validate the model's measurements, the processed images were compared with radiographs manually measured by two authors using SurgiMap Spine software as the ground truth. The model showed a detection rate of 91.5% for preoperative images and 84% for postoperative images.

Advancements in other non-orthopedic medical imaging demonstrate impressive progression in automation using machine learning (ML) and DL. Ghesu *et al.* employed a deep Q-network (DQN), a variant of CNN that utilizes a Markov decision process framework for decision making of anatomical landmark placement in cardiothoracic patients[24]. The system rewards actions proportionally to their proximity to the ground truth landmark of anatomical landmarks of the heart. The landmarks and their mean detection error (mm) are: left ventricle center (1.8 mm), right ventricle extremities (4.9 mm), right ventricle posterior (2.2 mm), and right ventricle anterior (3.7). The model performed well and predicted these landmarks with a convergence rate of 90%.

Kelm *et al.* proposed an automated analysis using marginal space learning (MSL) of disk selection and labeling, along with structure segmentation of MRI images, to provide a 3D model of vertebral bodies[25]. The spine's general location is found, and disk candidates are generated, including the position, orientation, and scale. A global probabilistic model is applied, encapsulating candidates based on appearance and pose, allowing the model to make educated guesses on the parameters of the spine. The training utilizes an iterative process and clustering, fine-tuning the selection by re-evaluating candidates constantly to make the outcome picture look as close as it can to the learned data. The study consisted of 42 MRI images of healthy volunteers, with a sensitivity of 98.64% and a PPV of 99.68%. This AI assistance can support segmental labeling and allows precise targeting of pathological disks when diagnosing ASD.

In 2017, Jamaludin *et al.* used a 2D and 3D CNN model named SpineNet to predict pathological features on MRI[26]. The study was on T2 sagittal spinal sequences from 2,009 patients to detect the following: foramen

grading, disc narrowing, upper/lower endplate defects and marrow changes, spondylolisthesis, and central canal stenosis. Using an 80:10:10 train:validation:test and a multi-task loss function, the model displayed a near-human performance compared to the intra-rater kappa value of the radiologist. Jamaludin also reports that the model improved when changed from single-task to multi-task CNN. As an example, the detection of lower endplate defects went from 79.5% to 86.4%. The intra-rater reliability score averaged 82.5%, ranging from 70.4% to 92.5%. The 2D and 3D models averaged 85.7% and 86.3% respectively. When comparing 2D *vs.* 3D models, the 3D model performed similar or better than the 2D model. The largest improvement between the models was in analyzing spondylolisthesis, where the transition to the 3D model jumped from 92.9% to 95.2%. Overall, the decision to use 2D *vs.* 3D models should be determined by the specific deformity at hand.

AI can increase the rate and accuracy of medical imaging, not only aiding radiologists in their diagnostic time but also reducing the scan time and reconstruction of each image. This will allow healthcare systems to minimize waiting time and give patients a faster diagnosis. Goodfellow *et al.* proposed a generative adversarial network (GAN), an implicit model that needs less data than other ML models[27]. The model is designed to generate entirely hypothetical sample images, like creating an image of a person that does not exist from images of existing celebrities. Although this technology is fascinating, it can be used in medicine to improve the image quality of scans from previously learned data.

Schlemper *et al.* and Souza *et al.* used CNN to accelerate data acquisition of MRI scans[28,29]. Schlemper's model used dynamic 2D sequences of cardiac images and outperformed state-of-the-art methods across all factors, taking only 23 ms to reconstruct each sequence and an average of 8.21 s on a GPU, much faster than dictionary learning with temporal gradient (DLTG), which took 6.6 h on average. To further aid radiologists and spine surgeons, Xuan *et al.* developed a DLTG with a PP-YOLOv2 object detection model to diagnose spinal diseases on MRI with 98% accuracy[30]. The models were trained on a generic data set, and labeled by experienced spine surgeons into: Normal, lumbar disc herniation, and spondylolisthesis. The model provided diagnostic results with an average time of 14.5 s, compared to the ten-minute average Xuan *et al.* reports a spine surgeon will spend on each case[30]. This can help maximize clinician efficiency and spot possible missed diagnoses.

Current limitations in AI imaging models include reduced performance in patients with atypical deformities, variability in image acquisition, and lack of standardization in annotation practices. Improving training data diversity and leveraging multimodal data inputs may enhance sensitivity and diagnostic accuracy in future applications.

PATIENT BENEFITS
AI has already turned heads around the world with chatbots and image creators such as open AI's ChatGPT, where any question can be asked, and the service will give an answer. The application of these chatbots can be useful for patients and clinicians, as demonstrated by Sri Lalitha *et al.* and Nimal *et al.* articles[31,32]. Sri Lalitha *et al.*'s system used a recurrent neural network (RNN) followed by a K-nearest neighbors (KNN) ML model for disease prediction and classification[31]. The training data, consisting of text and symptoms, was preprocessed into numerical format, and keywords and phrases of medical symptoms were extracted. The processed data are then fed into an input layer, followed by hidden layers of decision tree algorithms that classify the phrases and assign a score compared to the ground truth. The chatbot can then respond to the patient's questions using natural language processing (NLP), giving them information on symptoms and directing them to an appropriate healthcare site. The study showed promising results: the RNN had an accuracy of 96%, but the KNN had the lowest at 70% compared to ground truth. These

chatbots can shorten patients' waiting time from experiencing symptoms and can save clinicians time spent seeing patients that do not fit their respective fields.

AI-assisted real-time call centers are on the horizon, aiming to improve efficiency, boost productivity, and cut costs. Bian *et al.* reported alarming results with an AI-assisted follow-up conversational agent for postoperative orthopedic patients[33]. The system uses an automatic speech recognition (ASR) model that converts audio to text, which is then analyzed with natural language understanding (NLU) to determine an appropriate response. The response is formulated with natural language generation (NLG) and output as a human response to the patient. The results showed the AI system to be as effective as the manual method; it required no human intervention, spending close to 0 h per 100 patients, compared to the manual method, which required 9.3 h per 100 patients.

Ronckers *et al.*'s 2010 study on "Cancer mortality among women frequently exposed to radiographic examinations for spinal disorders" reported a worrisome 8% increase in cancer mortality, notably breast cancer, where they had a standardized mortality ratio (SMR) of 1.68 (68% increase) compared to control groups[34]. Maximizing patient safety should be the highest priority for any surgeon, and a reduction in radiation during spine surgery could add to the never-ending list of ways to provide a safer experience for patients and healthcare professionals. In 2006, Gebhard *et al.* illustrated that computer-assisted spine surgery (CAS) significantly reduced the duration and amount of radiation used in standard surgery[35]. CAS is often referred to now simply as "Navigation", using imaging before or during surgery, and registering perioperatively to track instrument use. This method, compared to the old way of using fluoroscopy of sagittal and transverse views, translated to a reduced radiation time from 177 to 40 s, and an 86% decrease in radiation value in mGy when using the Iso-C3D C-arm.

As important as Gebhard *et al.*'s findings were, they are now a thing of the past, as most operating rooms use navigation guidance during spinal surgeries[35]. To lower the dose even more, there has been the development of low-dose spine CT[36], which uses substantially less radiation than traditional CTs. AI has the capability, as mentioned above, to use models to create high-dose CT quality at a low-dose quality expense with image reconstruction models. Wu *et al.* proposed a DL-based algorithm to improve low-dose CT scans with a split unrolled grid-like alternative reconstruction (SUGAR) network[37]. The model leverages DL, physical modeling, and prior images to produce high-quality scans from minimal projection data. Yang *et al.* proposed a GAN network model with Wasserstein distance and perceptual loss, which showed an advantage in noise reduction while retaining critical image features compared to the mean squared error (MSE) network[38]. The Wasserstein distance measures how different two datasets are by seeing how much effort it takes to make one of the datasets look like the other. Perceptual loss measures how much information is lost or changed in the enhancement process of images and helps retain the overall feeling of the original image. Chen *et al.* introduced the residual encoder-decoder neural network (RED-CNN), designed to enhance images with a filtered back projection[39]. As the name suggests, it uses both a convolutional and deconvolutional layer to process scans within the image domain. The conventional layer compresses the input image into a smaller feature-rich representation, while the deconvolutional layer performs the opposite function to restore spatial details. This reconstruction process helps recover fine image details that may have been lost during encoding, thereby improving image quality and reducing noise. A backpropagation is implemented in the residual learning step, allowing the loss of deeper layers to be kept without diminution and, therefore, preserving image details. The RED-CNN uses small overlapping patches from full images, like observing each tile of a mosaic instead of the whole image, increasing training data and allowing the model to improve on finer details. The RED-CNN trains using an MSE loss function, quantitatively comparing output images to the ground truth using the peak signal-to-noise ratio (PSNR)

and structural similarity index (SSIM). Chen *et al.*'s results for the CNN-MSE model were the second best overall, with a PSNR score of 24.0637 and SSIM of 0.7966[39]. The only model to beat the CNN-MSE was Dictrecon, but the output images were blurry and had waxy artifacts.

RISK CALCULATOR AND DECISION-MAKING TOOLS

A collection of research has shown that machines, especially artificial neural networks (ANNs), have significant potential in predicting complications and readmission for spinal surgeries. In 2017, Kim *et al.* modeled a 70:30 training:testing ANN and logistic regression (LR) models, with data from 22,629 patients from the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQI), to identify risk factors for complications of posterior lumbar spine fusion[40]. The control for performance was the American Society of Anesthesiologists (ASA) class. ANN and LR performed better than ASA at predicting all four major types of complications: wound, cardiac, venous thromboembolism (VTE), and mortality. The ANN model had the best AUC prediction for cardiac (ANN: 0.710, ASA: 0.468), while the LR performed best for VTE (LR: 0.588, ASA: 0.435), wound (LR: 0.613, ASA: 0.491), and mortality (LR: 0.703, ASA: 0.369). Both the ANN models developed by Kuris *et al.* and Hopkins *et al.* successfully predicted readmissions using NSQIP data[41,42]. Kuris *et al.* reported being successful in predicting 30-day readmission for ALIF (94.6%), PLIF (94%), and PSF (92.6%)[41]. Hopkins *et al.*'s model achieved a mean and median PPV of 78.5% and 78.0% and an NPV of both mean and median 97%[42]. An impressive AUC was also found, averaging 0.812.

Accurately and efficiently predicting patient outcomes using AI involves: leveraging ML/DL trained on prior cases to generate background data, organizing new cases into subcategories, and then allowing AI to suggest what will be the best solution from what it has learned. Ames *et al.* applied a hierarchical clustering ML prediction model to analyze and categorize 570 confirmed ASD patients into three subcategories[43]: Young with coronal plane deformity, older with prior spine surgeries, and older without prior spine surgeries. The patients were split into either an objective group that included: age, sex, height, weight, and number of previous spine surgeries; and a subjective group that utilized patient-reported outcome measures that assessed the patient's health status at a given time. The selected surgical parameters included: the number of previous spine surgeries, approach, number of fused vertebral levels, pelvic fixation, operative time, estimated blood loss, and length of hospital stay. Additional covariates included the use of transforaminal lumbar interbody fusion (TLIF), anterior interbody fusion (ALIF), and osteotomy types. The model outcomes were then grouped to compare subcategories based on major complication rates and to identify which surgical parameters were associated with the highest success rates. This approach was further developed by Durand *et al.* to include the analysis of sagittal plane morphology[44]. Like Ames *et al.*, the retrospective study identified six different clusters (A-F) of preoperative lateral spine radiographs with unique spinal shapes and characteristics[43]. The outcome measures included: Oswestry disability index (ODI), proximal junctional kyphosis (PJK), proximal junctional failure (PJF), sagittal vertical axis (SVA), three column osteotomy (3-O), and upper instrumented vertebrae (UIV). The model showed certain clusters that were disproportionately higher in ODI, PJK, and PJF. Durand *et al.*'s findings conclude that the model is capable of predicting the factors mentioned, providing healthcare professionals with a reference to compare mean spine shapes for enhanced clinical decision making[44].

SURGICAL APPLICATIONS

Howe *et al.*'s paper published in 1999 foreshadowed what was to be a standard in modern healthcare[45]. Robots enable surgeons to perform minimally invasive surgeries, use image-guided assistance, and perform with higher accuracy. The main challenges reported in the 1999 article were: clinician acceptance, financial burdens of these robotics, performance reliability, and safety concerns. To be considered viable for clinical

use, robotic systems would have to improve patient safety, lower costs, or achieve both. The growth of surgical technology is exponential, demonstrated today with the Da Vinci telerobotic surgical system[46], allowing surgeons to operate from a computer screen with greater dexterity and accuracy. The navigation system used in spine and brain surgery mentioned previously is now standard, making techniques of the past seem primitive[47].

The advancement in AI-driven robotic spine surgery has shown promising results in improving accuracy, safety, and patient outcomes. Perioperative navigation and robotics, augmented reality (AR), and virtual reality (VR) training not only contribute to enhancing surgery precision and aiding patient recovery time but can expedite resident training on specific simulations. Rajasekaran *et al.*'s RCT study of 27 patients with thoracic spine deformity, involving a total of 478 thoracic pedicle screws, showed a staggering 23% pedicle breaches and 16% penetration into the anterior or lateral cortex in the non-navigation group, compared to 2% and 0.8% in the navigation group, respectively[48]. The navigation system also dropped screw insertion time by nearly half and minimized radiation, as discussed in the previous text. Kosmopoulos *et al.*'s meta-analysis of pedicle screw placement accuracy supported this finding when comparing screw violations[49].

Kim *et al.*'s prospective RCT on 78 patients with degenerative spinal disease undergoing PLIF compared minimally invasive robot-assisted to a conventional freehand (FH) technique using fluoroscopy[50]. The robot-assisted surgery used thin-slice CT scans to determine the optimal pedicle screw insertion path, select the appropriate implant size, and identify anatomical abnormalities. A Renaissance Surgical Guidance robot that can be precisely controlled was mounted on the operating table and secured on the spinous process of the desired vertebrae segment. A double verification via fluoroscopy is performed to ensure perfect alignment of the patient's anatomy. Results showed a non-significant intrapedicular accuracy ($P = 0.534$), but the robot-PLIF significantly reduced violations of the proximal facet joint ($P < 0.001$), and demonstrated a superior convergence orientation of the screws ($P < 0.001$) to ensure a safer distance from critical anatomical structures. This result of pedicle screw placement accuracy differs from that of studies such as Lonjon *et al.*, who reported an accuracy of 97.3% with robot-PLIF *vs.* 92.0% with the FH PLIF[51].

Kim *et al.* provided a comparative follow-up study of 1-year clinical and radiological outcomes[52]. The results stated that there were no significant differences between the two groups in ODI scores ($P = 0.688$), but a significant difference in disc height decrease in the robot-PLIF group ($P = 0.039$). D'Souza *et al.* performed a systematic review of robotic-assisted spine surgery, comparing several studies on accuracy, radiation exposure, and operative time[53]. They found that some studies showed that robot-assisted surgeries could improve accuracy and lower radiation exposure, but none of them resulted in faster operating times.

VR and AR are both perception-altering technologies that offer value in spinal surgery, but their experiences differ. VR allows surgeons and residents to train inside a simulation with the advantage of having a fail-safe for mistakes. Surgery resident training seeks to benefit from the innovative ways VR has evolved, as demonstrated in a pilot study performed by Ponce *et al.* on telementoring with a virtual interactive presence (VIP)[54]. The study assessed the effectiveness of the VIP in 15 surgeries, allowing attending surgeons to provide real-time assistance to resident surgeons performing surgery. The VIP uses a hybrid visual overlay and telestration, allowing attending surgeons to virtually "reach into" and even draw on the video feed to highlight anatomical structures and direct upcoming steps. Although a small sample size with only one attending surgeon participating, both attending surgeon and residents considered this favorable and easy to use, with possibilities to enhance training quality and quantity. Bissonnette *et al.* hypothesized whether AI could objectively distinguish between different levels of surgical training in a VR-

simulated hemilaminectomy[55]. Participating residents were classified into either senior or junior groups based on their training level, and twelve performance metrics were analyzed. A support vector machine (SVM) was trained to identify performance metrics with an accuracy of 97.6% to distinguish between the two groups, providing a standardized objective assessment. Immersive virtual reality (IVR) has been shown to enhance technical and non-technical surgical skills compared to traditional learning methods[56]. Lohre's blinded, multicenter RCT of 19 senior orthopedic residents and 7 consulting attending surgeons, aimed to compare IVR with traditional learning methods when performing glenoid exposure, using technical journal articles as a control. The IVR system included a head-mounted display (HMD) to immerse them in a virtual OR, a haptic controller with tactile feedback, and a feedback system on performance. No difference was in resident pre-surgical training, simulation familiarity, or previous VR training. The IVR group was significantly faster (14 ± 7 min) compared to the control group (21 ± 6 min) at completing the cadaveric dissection. Not only did the IVR show improved surgery and training module time, but resident instrument handling was also significantly better, and residents reported enjoying the learning activity (mean 4.8/5 IVR *vs.* 3.3 traditional). AR can apply computer-generated images to real-world scenarios, acting as a real-time guide. In 2013, Abe *et al.* introduced a novel AR system called virtual protractor with augmented reality (VIPAR), which aimed to improve the safety of percutaneous vertebroplasty of osteoporotic vertebral fractures[57]. The HMD was equipped with a tracking camera and ARToolKit AR software. Before surgery, patients undergo a 3D CT scan of the pathological spinal region, and a trajectory analysis is performed to optimize the patient's respective anatomy. The study is split into two parts: 40 computer simulations of spine phantoms that resemble human anatomy and then 5 patients to evaluate real-world practicality. The error of inserted angle (EIA) was assessed in two groups A and B, comparing groups using VIPAR and a non-VIPAR group. Both groups showed a statistically significant improvement using the VIPAR, with an average improvement from 4.34° to 0.96° in the axial plane, and 2.55° to 0.61° in the sagittal plane. Looking at the real-world application, a postoperative 3D CT scan of the bilateral needle insertion resulted in an EIA of 2.09° in the axial plane and 1.98° in the sagittal plane.

Great leaps have been made in the applications for minimally invasive spine surgery (MISS) in recent years, as it has the potential to shorten hospital stays, minimize blood loss, and cause fewer wound infections[58]. Burström *et al.*'s feasibility and accuracy study on the augmented reality surgical navigation (ARSN) system with instrument tracking on pig cadavers showed great promise[58]. The system found improved surgical precision, utilizing AR to track instruments and navigate by overlaying digital information into the surgeon's field of view, and VR to visualize 3D anatomical features in real time with enhanced spatial understanding. The study reported an advantage in accuracy, ranging from 97.4% to 100% depending on screw size, but also noted that the ARSN system does not require ionizing radiation during navigation. Elmi-Tarander *et al.*'s article was the first ARSN system researched on 20 patients undergoing pedicle screw placement[59], differing from Burströms *et al.* by using a C-arm with 2D/3D and only using AR to perform pedicle screw placement. The overall accuracy reported was 94.1%, with 5.9% of the screws being moderate breaches and none that were severely misplaced. The study also noted that most screws inserted were in the thoracic spine (64.4%), the most common range for lower accuracy insertion due to smaller pedicle size. Elmi-Tarander *et al.* followed up with a retrospective study comparing the ARSN group of 20 patients to a control group using traditional FH fluoroscopy[60]. A significantly higher screw placement accuracy was found in the ARSN, 93.9% compared to 89.6% FH, and a much higher rate of screws with no cortical breach (63.4% *vs.* 30.6%). These studies show how AR and VR can work together in the operating room to give surgeons more flexibility and insight for complex cases.

While still in the early stages, real-time AI-guided interventions are emerging in surgical planning and navigation. Challenges include integration with existing intraoperative systems, latency in image processing,

and ensuring surgeon interpretability and control. Further validation studies and close collaboration with engineers will be critical to safe implementation.

PERIOPERATIVE APPLICATIONS

Peri- and postoperative blood transfusions are a common practice, ranging from 27% up to 90% in some studies. An AI algorithm could minimize unnecessary RBC transfusions and highlight patients who are at risk for improved preoperative planning. Raman *et al.*'s study on conditional inference tree analysis of ASD surgeries was able to predict combinations of variables associated with intraoperative blood loss and perioperative RBC transfusions[61]. High-risk groups identified were: Fusion of more than 13 levels, ASA score greater than 1, history of hypertension, 3-column osteotomies, pelvic fixation, and surgery lasting longer than 8 h. Durand *et al.*'s random forest model supports these findings, reporting an 80:20 training:validation model that showed an AUC of 0.85[62]. A simple classification tree was compared, producing an AUC of 0.79, proving an advantage of the ML random forest model. Newer studies, such as that by De la Garza Ramos *et al.*, continue to back the use of AI in predicting patient groups that are likely to need blood transfusions[63]. The overall accuracy reported was 81% on the training data (70%) and 77% on the testing data (30%), with a sensitivity of 80% and an AUC of 0.84.

ASD PROGRESSION

An emergent application of AI in spine surgery has been the development of models to predict the progression of pathology in ASD, both in regard to nonoperative management and surgical intervention. In the early stages of predictive models for spinal progression, Nault *et al.* sought to develop a model that incorporated clinical characteristics, specifically skeletal maturity, and radiographic findings to predict the final Cobb angle of patients[64]. The authors developed a model with backward step regression that was then validated with a Bland-Altman method, which determined the goodness of fit to be 0.643. These findings demonstrated that while such primitive models had predictive merit, there was a significant need to refine and develop learning models with better performance.

In 2020, Wang *et al.* produced a DL model that implemented radiographs from an AIS patient's first visit to distinguish between progressive and non-progressive curves[65]. The use of a self-attentive capsule learning network was found to outperform traditional CNNs and clinical parameter-based models. This study suggests the potential for automated prediction of AIS curve progression using DL and radiomics, which could guide treatment decisions at the initial visit, such as early bracing for at-risk patients. Looking beyond the clinic, Zhang *et al.* sought to develop a model that could assess spinal deformity progression with the use of a patient's unclothed back appearance taken from a smartphone photograph[66]. This model, performing at an overall level of 91%, was found to have an AUC of 0.757 in distinguishing curve progression, suggesting a simple DL app has the promise for managing scoliosis in children outside of hospital settings, without radiation exposure. As researchers aim to expand the application horizons of AI, they also seek to strengthen the predictive value of these models. Zhao *et al.* produced a robust model known as SpineHRReformer that sought to characterize Cobb angles without the loss of efficiency from segmentation and endplate slope calculation[67]. This study produced a DL model that combined the transformer blocks and heatmap approaches to reliably produce spinal deformity assessments. One of the more worthy models described thus far, SpineHRReformer, performed at correlation coefficients exceeding 0.8 in predictive values.

DISCUSSION

As healthcare progressively digitalizes, a massive aggregate of patient data can materialize into new frontiers with AI and ML. These technologies are starting to revolutionize medical practices by providing

comprehensive, data-driven approaches for complex spinal pathologies in an efficient manner. They enhance decision making from patient selection to intraoperative strategies, fostering a cycle of self-improving computational analysis. Recent innovations have tapped into AI prowess in surgical settings as well as education. We sought to outline these current implementations, particularly emphasizing ones that have not previously been reviewed. The caveat with all emerging technologies is the limitations. The applications of AI/ML seem to have no bounds so the task at hand is the quality assurance of these applications.

When AI models are trained on datasets that are limited in diversity and size, they are prone to become too closely aligned with the dataset in question, capturing noise or specific patterns that do not generalize well to new datasets[68,69]. Consequently, these models run the risk of misdiagnosis or misinterpretation by consumers. In the various subsets of applications we outlined above, we attempted to capture the extent of validation of these models - how tried and true the results truly are. Frequently, we noted models that were internally validated with discrimination, decision, ROC and AUC analysis. The most observed splits are 70:30 or 80:20. To mitigate overfitting, many studies employed techniques such as L2 regularization, dropout layers in neural networks, and k-fold cross-validation. These methods enhance the robustness of models and ensure that performance is more likely to generalize across clinical settings. We rarely encountered instances where the models were externally validated against other datasets. This trend can partially be explained by the intersecting culture of academia and industry, where the goal is to be the front-runner of a concept and not necessarily the one to deliver the polished version[70,71]. External validation remains a key challenge in the development of AI models for ASD. The lack of multi-institutional datasets and standardized imaging protocols limits generalizability. Future efforts should focus on collaborative databases and benchmarking to enhance external validation opportunities. Thus, while we appreciate the endless possibilities presented by AI, we advise cautionary use of compromised products outputted purely to match this pace and not sufficiently validated. Regulatory bodies such as the US Food and Drug Administration have ongoing discussions about placing a nutrition label equivalent on these models to properly advise the scientific community and its consumers[72]. Patients and surgeons alike should aim to use AI in provision with their joint decision-making process.

However, we do not intend to simply provide a cautionary tale but also encourage the celebration of technology that will inherently spearhead the evolution of our field. Thus far, we have heavily commented on the potential AI demonstrates for spinal surgeons. There is also a plethora of future directions for patients. A growing body of literature substantiates the use of AI for health literacy among patients[73]. It has been well established that patient-reported outcomes increase with the level of transparency surrounding their care. AI can supplement the discussions patients have with their providers and inherently bolster satisfaction with their treatment plans[74]. AI can also provide a level of continuous health monitoring that can alert patients to health events not immediately apparent. Following spinal surgery, patients are frequently referred to physical therapy. AI can be utilized to record these sessions and evaluate the outcomes to gauge recovery progress. At the interface of operating surgeons and patients is a series of communication and care delivery that is often inefficiently done because of conflicting schedules. AI can automate and increase the availability of these services, maximizing the satisfaction patients have with ease of access while simultaneously reducing the financial strain on healthcare systems. Ultimately, the concomitant public health initiatives that surface from this technological trend will define the patient care continuum for generations to come.

CONCLUSION

AI has shown promising applications in imaging analysis, surgical planning, and outcome prediction for ASD. Despite these advancements, key challenges remain in data standardization, external validation, and real-time clinical integration. Future research should focus on multi-institutional collaborations, interpretability, and prospective clinical trials to fully realize the potential of AI in this domain.

DECLARATIONS

Authors' contributions
Conceptualization, data curation, writing - original draft preparation, writing - review and editing: Sigurdarson H
Conceptualization, writing - original draft preparation, writing - review and editing: Joshi A
Writing - original draft preparation, writing - review and editing: Mohebi A
Supervision, writing - review and editing: Hassanzadeh H

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. Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *J Ambient Intell Humaniz Comput.* 2023;14:8459-86. DOI PubMed PMC
2. Shen J, Zhang CJP, Jiang B, et al. Artificial intelligence versus clinicians in disease diagnosis: systematic review. *JMIR Med Inform.* 2019;7:e10010. DOI PubMed PMC
3. Chang MC, Kim JK, Park D, Kim JH, Kim CR, Choo YJ. The use of artificial intelligence to predict the prognosis of patients undergoing central nervous system rehabilitation: a narrative review. *Healthcare.* 2023;11:2687. DOI PubMed PMC
4. Poalelungi DG, Musat CL, Fulga A, et al. Advancing patient care: how artificial intelligence is transforming healthcare. *J Pers Med.* 2023;13:1214. DOI PubMed PMC
5. Reddy S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implement Sci.* 2024;19:27. DOI PubMed PMC
6. Ball HC. Improving healthcare cost, quality, and access through artificial intelligence and machine learning applications. *J Healthc Manag.* 2021;66:271-9. DOI PubMed
7. Rubinger L, Gazendam A, Ekhtiari S, Bhandari M. Machine learning and artificial intelligence in research and healthcare. *Injury.* 2023;54 Suppl 3:S69-73. DOI PubMed
8. Zhao R, Xie Z, Zhuang Y, Yu PLH. Automated quality evaluation of large-scale benchmark datasets for vision-language tasks. *Int J Neural Syst.* 2024;34:2450009. DOI
9. Rahmani AM, Azhir E, Ali S, et al. Artificial intelligence approaches and mechanisms for big data analytics: a systematic study. *PeerJ Comput Sci.* 2021;7:e488. DOI PubMed PMC

10. Nayarisseri A, Khandelwal R, Tanwar P, et al. Artificial intelligence, big data and machine learning approaches in precision medicine & drug discovery. *Curr Drug Targets.* 2021;22:631-55. DOI PubMed

11. Loftus TJ, Tighe PJ, Filiberto AC, et al. Artificial intelligence and surgical decision-making. *JAMA Surg.* 2020;155:148-58. DOI PubMed PMC

12. Kim HJ, Yang JH, Chang DG, et al. Adult spinal deformity: current concepts and decision-making strategies for management. *Asian Spine J.* 2020;14:886-97. DOI PubMed PMC

13. Kim HJ, Yang JH, Chang DG, et al. Adult spinal deformity: a comprehensive review of current advances and future directions. *Asian Spine J.* 2022;16:776-88. DOI PubMed PMC

14. Patel RV, Yearley AG, Isaac H, Chalif EJ, Chalif JI, Zaidi HA. Advances and evolving challenges in spinal deformity surgery. *J Clin Med.* 2023;12:6386. DOI PubMed PMC

15. Ailon T, Scheer JK, Lafage V, et al; International Spine Study Group. Adult spinal deformity surgeons are unable to accurately predict postoperative spinal alignment using clinical judgment alone. *Spine Deform.* 2016;4:323-9. DOI

16. Zhou S, Zhou F, Sun Y, et al. The application of artificial intelligence in spine surgery. *Front Surg.* 2022;9:885599. DOI PubMed PMC

17. 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

18. 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

19. Wirries A, Geiger F, Oberkircher L, Jabari S. An evolution gaining momentum - the growing role of artificial intelligence in the diagnosis and treatment of spinal diseases. *Diagnostics.* 2022;12:836. DOI PubMed PMC

20. 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

21. 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

22. Klinder T, Ostermann J, Ehm M, Franz A, Kneser R, Lorenz C. Automated model-based vertebra detection, identification, and segmentation in CT images. *Med Image Anal.* 2009;13:471-82. DOI

23. 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;33:4119-24. DOI

24. Ghesu FC, Georgescu B, Mansi T, Neumann D, Hornegger J, Comaniciu D. An artificial agent for anatomical landmark detection in medical images. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016. Springer, Cham; 2016. pp. 229-37. DOI

25. Michael Kelm B, Wels M, Kevin Zhou S, et al. Spine detection in CT and MR using iterated marginal space learning. *Med Image Anal.* 2013;17:1283-92. DOI

26. Jamaludin A, Kadir T, Zisserman A. SpineNet: automated classification and evidence visualization in spinal MRIs. *Med Image Anal.* 2017;41:63-73. DOI PubMed

27. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *Commun ACM.* 2020;63:139-44. DOI

28. Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging.* 2018;37:491-503. DOI PubMed

29. Souza R, Lebel R, Frayne R. A hybrid, dual domain, cascade of convolutional neural networks for magnetic resonance image reconstruction. *PMLR.* 2019;102:437-46. Available from: https://proceedings.mlr.press/v102/souza19a.html. [Last accessed on 6 Mar 2025]

30. Xuan J, Ke B, Ma W, Liang Y, Hu W. Spinal disease diagnosis assistant based on MRI images using deep transfer learning methods. *Front Public Health.* 2023;11:1044525. DOI PubMed PMC

31. Sri Lalitha Y, Ganapathi Raju NV, Vanimireddy RT, et al. Conversational AI Chatbot for HealthCare. *E3S Web Conf.* 2023;391:01114. DOI

32. Nimal KA, Nair VV, Jegdeep R, Nehru JA. Artificial intelligence based Chatbot for healthcare applications. *Adv Sci Technol.* 2023;124:370-7. DOI

33. Bian Y, Xiang Y, Tong B, Feng B, Weng X. Artificial intelligence-assisted system in postoperative follow-up of orthopedic patients: exploratory quantitative and qualitative study. *J Med Internet Res.* 2020;22:e16896. DOI PubMed PMC

34. Ronckers CM, Land CE, Miller JS, Stovall M, Lonstein JE, Doody MM. Cancer mortality among women frequently exposed to radiographic examinations for spinal disorders. *Radiat Res.* 2010;174:83-90. DOI PubMed PMC

35. Gebhard FT, Kraus MD, Schneider E, Liener UC, Kinzl L, Arand M. Does computer-assisted spine surgery reduce intraoperative radiation doses? *Spine.* 2006;31:2024-7; discussion 2028. DOI PubMed

36. Abul-Kasim K. Low-dose spine CT: optimisation and clinical implementation. *Radiat Prot Dosimetry.* 2010;139:169-72. DOI PubMed

37. Wu W, Niu C, Ebrahimian S, Yu H, Kalra MK, Wang G. AI-enabled ultra-low-dose CT reconstruction. *ArXiv* 2021, arXiv:2106.09834. Available from: https://doi.org/10.48550/arXiv.2106.09834. [Last accessed on 6 Mar 2025]

38. Yang Q, Yan P, Zhang Y, et al. Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Trans Med Imaging.* 2018;37:1348-57. DOI PubMed PMC

39. Chen H, Zhang Y, Kalra MK, et al. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans Med*

Imaging. 2017;36:2524-35. DOI PubMed PMC

40. 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

41. Kuris EO, Veeramani A, McDonald CL, et al. Predicting readmission after anterior, posterior, and posterior interbody lumbar spinal fusion: a neural network machine learning approach. *World Neurosurg.* 2021;151:e19-27. DOI

42. Hopkins BS, Yamaguchi JT, Garcia R, et al. Using machine learning to predict 30-day readmissions after posterior lumbar fusion: an NSQIP study involving 23,264 patients. *J Neurosurg Spine.* 2020;32:399-406. DOI

43. 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

44. 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

45. Howe RD, Matsuoka Y. Robotics for surgery. *Annu Rev Biomed Eng.* 1999;1:211-40. DOI PubMed

46. Ballantyne GH, Moll F. The da Vinci telerobotic surgical system: the virtual operative field and telepresence surgery. *Surg Clin North Am.* 2003;83:1293-304, vii. DOI PubMed

47. Mezger U, Jendrewski C, Bartels M. Navigation in surgery. *Langenbecks Arch Surg.* 2013;398:501-14. DOI PubMed PMC

48. Rajasekaran S, Vidyadhara S, Ramesh P, Shetty AP. Randomized clinical study to compare the accuracy of navigated and non-navigated thoracic pedicle screws in deformity correction surgeries. *Spine.* 2007;32:E56-64. DOI PubMed

49. Kosmopoulos V, Schizas C. Pedicle screw placement accuracy: a meta-analysis. *Spine.* 2007;32:E111-20. DOI PubMed

50. Kim HJ, Jung WI, Chang BS, Lee CK, Kang KT, Yeom JS. A prospective, randomized, controlled trial of robot-assisted vs freehand pedicle screw fixation in spine surgery. *Int J Med Robot.* 2017;13:e1779. DOI PubMed

51. Lonjon N, Chan-Seng E, Costalat V, Bonnafoux B, Vassal M, Boetto J. Robot-assisted spine surgery: feasibility study through a prospective case-matched analysis. *Eur Spine J.* 2016;25:947-55. DOI PubMed

52. Kim HJ, Kang KT, Chun HJ, et al. Comparative study of 1-year clinical and radiological outcomes using robot-assisted pedicle screw fixation and freehand technique in posterior lumbar interbody fusion: a prospective, randomized controlled trial. *Int J Med Robot.* 2018;14:e1917. DOI PubMed

53. 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

54. Ponce BA, Jennings JK, Clay TB, May MB, Huisingh C, Sheppard ED. Telementoring: use of augmented reality in orthopaedic education: AAOS exhibit selection. *J Bone Joint Surg Am.* 2014;96:e84. DOI PubMed

55. Bissonnette V, Mirchi N, Ledwos N, Alsidieri G, Winkler-Schwartz A, Del Maestro RF; Neurosurgical Simulation & Artificial Intelligence Learning Centre. Artificial intelligence distinguishes surgical training levels in a virtual reality spinal task. *J Bone Joint Surg Am.* 2019;101:e127. DOI PubMed PMC

56. Lohre R, Bois AJ, Athwal GS, Goel DP; Canadian Shoulder and Elbow Society (CSES). Improved complex skill acquisition by immersive virtual reality training: a randomized controlled trial. *J Bone Joint Surg Am.* 2020;102:e26. DOI PubMed

57. Abe Y, Sato S, Kato K, et al. A novel 3D guidance system using augmented reality for percutaneous vertebroplasty: technical note. *J Neurosurg Spine.* 2013;19:492-501. DOI

58. Burström G, Nachabe R, Persson O, Edström E, Elmi Terander A. Augmented and virtual reality instrument tracking for minimally invasive spine surgery: a feasibility and accuracy study. *Spine.* 2019;44:1097-104. DOI PubMed

59. Elmi-Terander A, Burström G, Nachabe R, et al. Pedicle screw placement using augmented reality surgical navigation with intraoperative 3D imaging: a first in-human prospective cohort study. *Spine.* 2019;44:517-25. DOI PubMed PMC

60. 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

61. 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

62. 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

63. 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

64. Nault ML, Beauséjour M, Roy-Beaudry M, et al. A predictive model of progression for adolescent idiopathic scoliosis based on 3D spine parameters at first visit. *Spine.* 2020;45:605-11. DOI

65. Wang H, Zhang T, Cheung KM, Shea GK. Application of deep learning upon spinal radiographs to predict progression in adolescent idiopathic scoliosis at first clinic visit. *EClinicalMedicine.* 2021;42:101220. DOI PubMed PMC

66. Zhang T, Zhu C, Zhao Y, et al. Deep learning model to classify and monitor idiopathic scoliosis in adolescents using a single smartphone photograph. *JAMA Netw Open.* 2023;6:e2330617. DOI PubMed PMC

67. Zhao M, Meng N, Cheung JPY, Yu C, Lu P, Zhang T. SpineHRformer: a transformer-based deep learning model for automatic spine deformity assessment with prospective validation. *Bioengineering.* 2023;10:1333. DOI PubMed PMC

68. Charilaou P, Battat R. Machine learning models and over-fitting considerations. *World J Gastroenterol.* 2022;28:605-7. DOI PubMed PMC

Sigurdarson *et al. Art Int Surg.* 2025;5:283-97 | https://dx.doi.org/10.20517/ais.2024.35   Page 297

69.   Subramanian J, Simon R. Overfitting in prediction models - is it a problem only in high dimensions? *Contemp Clin Trials.* 2013;36:636-41. DOI PubMed

70.   Bauer EA, Cohen DE. The changing roles of industry and academia. *J Invest Dermatol.* 2012;132:1033-6. DOI PubMed

71.   Yarborough M. Moving towards less biased research. *BMJ Open Sci.* 2021;5:e100116. DOI PubMed PMC

72.   U.S. Food & Drug Administration. Artificial intelligence & medical products: how CBER, CDER, CDRH, and OCP are working together. 2024. Available from: https://www.fda.gov/media/177030/download. [Last accessed on 6 May 2025].

73.   Liu T, Xiao X. A framework of AI-based approaches to improving eHealth literacy and combating infodemic. *Front Public Health.* 2021;9:755808. DOI PubMed PMC

74.   Yelne S, Chaudhary M, Dod K, Sayyad A, Sharma R. Harnessing the power of AI: a comprehensive review of its impact and challenges in nursing science and healthcare. *Cureus.* 2023;15:e49252. DOI PubMed PMC