BMC Musculoskeletal Disorders

SYSTEMATIC REVIEW

Open Access

Prevention and management of degenerative lumbar spine disorders through artificial intelligence-based decision support systems: a systematic review

Paolo Giaccone[1,2], Federico D'Antoni[1,2*], Fabrizio Russo[3,4*], Luca Ambrosio[3,4], Giuseppe Francesco Papalia[3,4], Onorato d'Angelis[5], Gianluca Vadalà[3,4], Albert Comelli[6], Luca Vollero[5], Mario Merone[2], Rocco Papalia[3,4] and Vincenzo Denaro[3,4]

Abstract

Background Low back pain is the leading cause of disability worldwide with a significant socioeconomic burden; artificial intelligence (AI) has proved to have a great potential in supporting clinical decisions at each stage of the healthcare process. In this article, we have systematically reviewed the available literature on the applications of AI-based Decision Support Systems (DSS) in the clinical prevention and management of Low Back Pain (LBP) due to lumbar degenerative spine disorders.

Methods A systematic review of Pubmed and Scopus databases was performed according to the PRISMA statement. Studies reporting the application of DSS to support the prevention and/or management of LBP due to lumbar degenerative diseases were included. The QUADAS-2 tool was utilized to assess the risk of bias in the included studies. The area under the curve (AUC) and accuracy were assessed for each study.

Results Twenty five articles met the inclusion criteria. Several different machine learning and deep learning algorithms were employed, and their predictive ability on clinical, demographic, psychosocial, and imaging data was assessed. The included studies mainly encompassed three tasks: clinical score definition, clinical assessment, and eligibility prediction and reached AUC scores of 0.93, 0.99 and 0.95, respectively.

Conclusions AI-based DSS applications showed a high degree of accuracy in performing a wide set of different tasks. These findings lay the foundation for further research to improve the current understanding and encourage wider adoption of AI in clinical decision-making.

Keywords Artificial intelligence, Machine learning, Deep learning, Low back pain, Spine, Prevention, Intervertebral disc, Decision support system

*Correspondence:
Federico D'Antoni
f.dantoni@unicampus.it
Fabrizio Russo
fabrizio.russo@policlinicocampus.it
Full list of author information is available at the end of the article

Introduction

Low back pain (LBP) is a debilitating multifactorial clinical condition, which affects millions of individuals worldwide [1], representing the main global cause of years lived with disability thus being associated with an enormous socioeconomic burden [2]. The prevalence of chronic LBP increases with advancing age, and it is commonly caused by degenerative lumbar spine disorders, such as intervertebral disc degeneration (IDD), lumbar spinal stenosis (LSS), degenerative spondylolisthesis (DLS), lumbar disc herniation (LDH), facet joint osteoarthritis, and adult spine deformities (ASD). In recent years, the evolution of artificial intelligence (AI) has allowed for significant developments in healthcare within different fields of medical research [3]. Due to technological advances resulting in an increased ability to recognize disease patterns in datasets, AI has been adopted to improve the knowledge, diagnosis, and treatment of several medical conditions [1]. With regard to the spine field, different applications of AI in LBP diagnosis and treatment have been investigated. To date, the potential of AI in spine surgery has been exploited for different tasks, including tasks such as the automatic segmentation of vertebral structures, the analysis of clinical and surgical notes, or the elaboration of patient-reported outcomes [4].

The use of AI in the spine field can be divided into three main categories, namely Computer Vision (CV), Computer-Aided Diagnosis (CAD), and Decision Support Systems (DSS). CV entails the ability of computers to achieve a comprehensive understanding of digital images and, as regards LBP, its application mainly pertains to feature extraction and segmentation from imaging data [5]. On the other hand, CAD comprises a group of techniques that assist medical practitioners in identifying a disease or quantifying its severity through classification and regression tasks, in which Machine Learning (ML) or deep learning (DL) models are used to assign a predefined label or generate a numeric output, respectively [6]. In this context, another promising technology is represented by natural language processing (NLP), which has been used to classify various conditions, label documents, and interpret heterogeneous data in different spinal disorders by transforming diverse source inputs in a natural, interpretable language [7]. In this scenario, DSS aim to enhance the decision-making process for medical practitioners and/or patients, with the ultimate goal of improving the outcomes for individuals affected by a specific disease. This encompasses activities ranging from defining and validating novel clinical indices to clinically evaluating the patient for diagnostic or prognostic support. Other applications include the suggestion of the most suitable treatment options and prediction of the potential improvements or complications a patient may experience after undergoing a specific therapy [8]. A schematic summary of the application fields of DSS in chronic LBP is reported in Fig. 1. It is worth noting that a clear distinction between the domains of prevention and outcome prediction is not always possible (e.g., eligibility for a certain treatment may be provided on the basis of multiple outcome prediction scores). In this study, we focused on studies that present prevention and management as the principal endpoint. To date, DSS systems have been recognized for their capacity

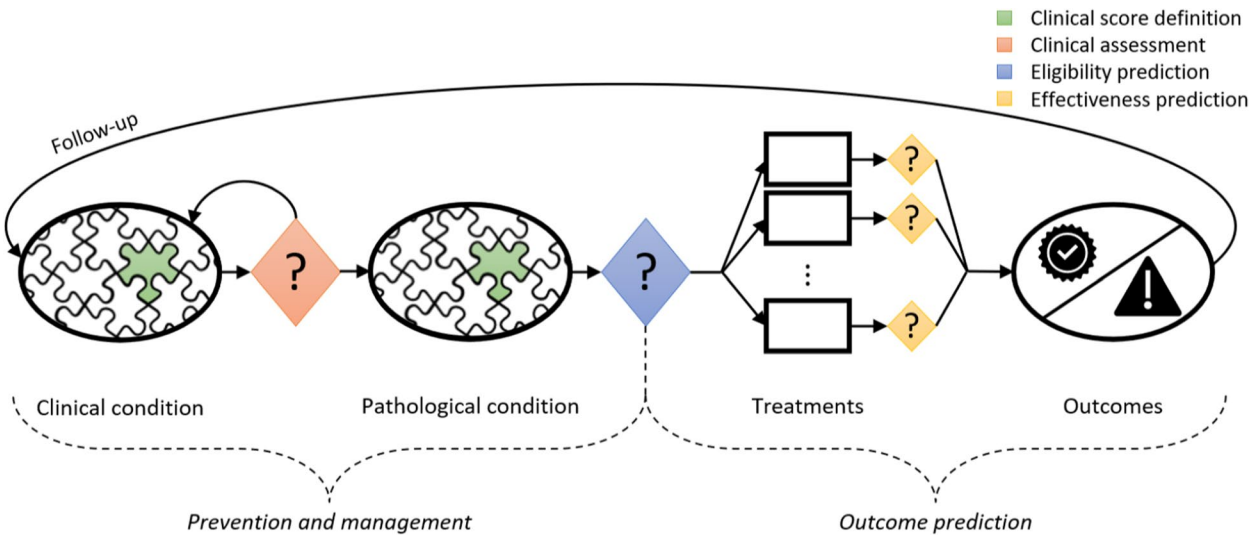

Fig. 1 AI DSS application fields. The colored boxes and diamonds represent the different possible goals of a DSS throughout the healthcare process, which macroscopically belong to either the prevention and management or the outcome prediction task

to increase compliance with guidelines and follow evidence-based recommendations in the daily clinical practice [9].

In the field of degenerative lumbar spine disorders, DSS may assist healthcare professionals in discriminating between different causes of LBP and providing personalized recommendations based on patients' history, clinical characteristics, and possible treatment options. The aim of this study was to systematically review the available evidence on the role and efficacy of AI-based DSS as supportive tools for making decisions on prevention and management in patients affected by LBP due to degenerative lumbar spine disorders.

Materials and methods

Electronic literature search

The protocol of this systematic review was registered in the Open Science Framework (OSF) database (https://osf.io/mksd3). A systematic search of PubMed and Scopus databases was performed on August 26th, 2024. The following search terms were used: "low back pain", "intervertebral disc degeneration", "intervertebral disc displacement", "spine surgery", "disc herniation", "artificial intelligence", "machine learning", "deep learning", "neural network", and "decision support system". The complete search strategy is reported as a Supplementary Material. General study characteristics extracted included: authors, year of publication, country, data type, sample size, clinical domain, clinical task, and performance-related metrics. The initial search of the articles was conducted by two reviewers (F.D.A. and L.A.). In case of disagreements, a third reviewer (F.R.) was involved to resolve conflicts. The following research order was used: first, papers were screened based on titles and abstract; then full texts of the remaining articles were assessed. The screening workflow is reported in a PRISMA flow diagram (Fig. 2). Briefly, we included studies describing the application of novel DSS approaches to support prevention and treatment strategies for the management of chronic LBP or lumbar degenerative conditions. We included articles exploring AI methodologies within the domains of CV, ML, and neural networks (NN), regardless of the specific type of employed data, such as imaging, text, or clinical data. Studies including < 10 patients, articles in languages other than English, reviews, meta-analyses, cadaveric studies, letters to the editor, case reports, technical notes, preclinical studies, gray literature, and commentaries were excluded from the analysis. Similarly, we discarded studies either proposing support

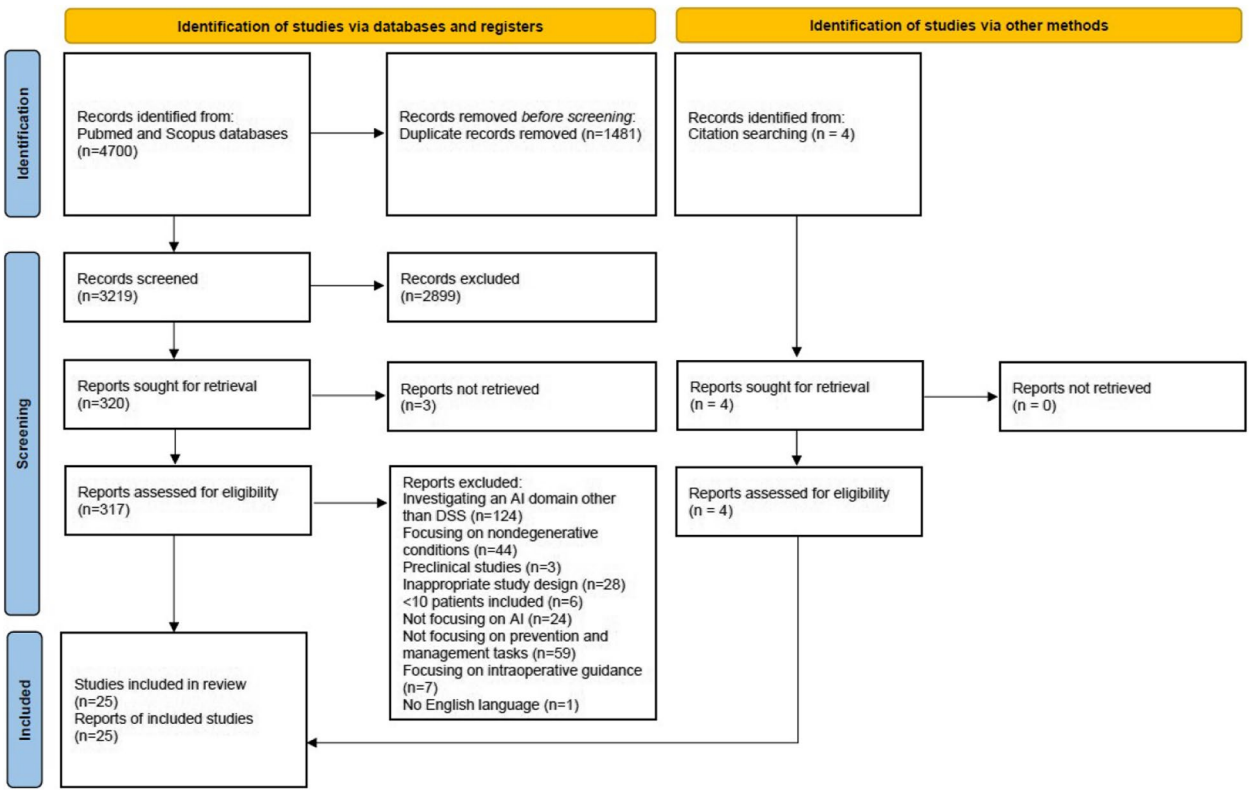

Fig. 2 Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flow diagram

systems without the employment of AI, or employing CV without providing any clinical decision support.

Evaluation metrics

Both classification and regression tasks were accomplished in the reported papers, thus different tasks adopted different metrics to evaluate their performance. However, depending on the specific task, different metrics have also been considered within the same paper.

As regards the classification task, we reported most of the results in terms of area under the curve (AUC) or accuracy (Acc). In brief, when considering a binary classification task (e.g., Positive label vs Negative label), given a test set composed of N samples, the true positives (TP) are defined as the number of positive samples correctly classified, while the true negatives (TN) as the number of negative samples correctly classified. Thus, the accuracy score is defined as:

$$Acc(\%) = \frac{TP + TN}{N} \times 100 \tag{1}$$

where greater values correspond to better performance. For each class, recall and precision can be computed as well. Defined the false positives (FP) and false negatives (FN) as the number of misclassified positive or negative samples, recall and precision are defined as:

$$Recall = \frac{TP}{TP + FN} \qquad Precision = \frac{TP}{TP + FP} \tag{2}$$

In binary problems, recall is also called TP rate and corresponds to sensitivity, whereas the TN rate is also called specificity. In the case of multi-class problems, accuracy is computed by considering the TP for each class, and recall and precision can be computed per each class.

Another widely used evaluation metric is the area under the curve (AUC), which corresponds to the area under the receiver operating characteristic (ROC) curve showing the performance of a classifier at all classification thresholds, which is plotted considering the TP rate against the FP rate. Its values range from 0 to 1 (the closer to 1, the better the performance).

With regard to the Regression task, let us consider a sequence of original values $x(t)$ and a sequence of predicted values $\tilde{x}(t)$. Moreover, defined $\hat{x}$ as the average value of the sequence, the residual sum of squares (RSS) and the total sum of squares (TSS) for a sequence of N timestamps are defined as:

$$RSS = \sum_{t=1}^{N} [x(t) - \tilde{x}(t)]^2, \qquad TSS = \sum_{t=1}^{N} [x(t) - \hat{x}(t)]^2 \tag{3}$$

These measures quantify the amount of variance of $x(t)$ that is not explained by the model and the overall variation in the original data, respectively. Besides, a unique comprehensive score, called coefficient of determination or r^2, can be defined as:

$$r^2 = 1 - \frac{RSS}{TSS} \tag{4}$$

thus, the closer to 1, the better the performance. In some cases, percentage error values are used to evaluate performance, the meaning of which varies based on the investigated task.

Risk of bias

The methodological quality of the included studies was independently evaluated by two reviewers (L.A. and G.F.P.), and any conflict was solved by the intervention of a third reviewer (G.V.). The risk of bias and applicability of included studies were assessed according to the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) [10]. This tool is based on 4 domains: patient selection, index test, reference standard, and flow and timing. Each domain is evaluated in terms of risk of bias, and the first 3 domains are also assessed in terms of concerns regarding applicability.

Results

Study selection

A total of 4700 articles were found. After duplicate removal, 3219 studies were screened at the title and abstract levels. 2899 articles were excluded following title and abstract screening and 3 reports were not retrieved. Then, 317 full-text articles were screened. Four additional studies were identified from hand searching bibliographies of included studies or identified systematic reviews. Out of these studies, 296 were excluded (articles focusing on CAD, n=87; articles focusing on CV, n=37, articles focusing on nondegenerative conditions, n=44; preclinical studies, n=3; no original research design, n=28; studies including < 10 patients, n=6; no use of AI, n=24; articles on intraoperative guidance, n=7; no English language, n=1; DSS-based studies not focusing on prevention and management, n=59). After this process, 25 articles were included (Fig. 2).

Study characteristics

Prevention and management systems may allow physicians or patients to prevent/mitigate the effects of degenerative lumbar spine conditions, monitor its progression, and identify the best treatment. The main characteristics of the included studies, such as underlying diagnosis, number of enrolled patients, clinical endpoint, data type, task, and AI model employed are summarized

Table 1 Summary characteristics of included studies related to definition of clinical score

Study	Country	Sample size	Diagnosis	Data type	Task	Results	AI model
Roller, 2021 [19]	USA	141	LSS treated with surgical decompression	MRI data	Predicting the lumbar microdecompression level	Acc = 65%	CNN
Jin, 2021 [26]	China	159	Lumbar IDD	Clinical, demographic and MRI data	Developing a new predictive classification system for LBP	Three LBP clusters were identified	K-means
Page, 2022 [28]	USA	89	Lumbar facet synovial cyst	Clinical and MRI data	Predict the risk of lumbar synovial cyst recurrence	AUC = 0.83	Supersparse Linear Integer Model
Yu, 2022 [29]	China	200	LDH	Clinical and MRI data	Predicting conservative vs. surgical treatment	AUC = 0.93	LASSO and LR
Campagner, 2020 [33]	Italy	72	Single-level discopathy undergone TLIF, XLIF or ALIF	Clinical data and biomarkers	Definition of a surgical invasiveness score for treatment planning	AUC = 0.87, 0.76	LR, kNN, NB, DT, SVM, RF
Xiong, 2023 [14]	China	59	LBP	Clinical and radiological data	Cage subsidence score definition and prediction for lumbar fusion surgical planning	AUC = 0.89	GB, NN

Abbreviations: *Acc* accuracy, *ALIF* anterior lumbar interbody fusion, *AUC* area under the curve, *CNN* convolutional neural network, *DT* decision tree, *GB* gradient boosting, *IDD* intervertebral disc degeneration, *kNN* k-nearest neighbors, *LASSO* least absolute shrinkage and selection operator, *LBP* low back pain, *LDH* lumbar disc herniation, *LR* logistic regression, *LSS* lumbar spinal stenosis, *MRI* magnetic resonance imaging, *NB* Naïve Bayes, *NN* neural network, *RF* random forest, *SVM* support vector machine, *TLIF* transforaminal lumbar interbody fusion, *XLIF* extreme lateral lumbar interbody fusion

in Tables 1, 2 and 3. Included patients were affected by LBP [11–16], DLS [17], ASD [18], LSS [19–23], unspecified lumbar IDD [24–27], lumbar facet joint synovial cyst [28], and LDH [29, 30]. Patients who underwent full endoscopic spine surgery [31] or laminectomy [32] for different lumbar degenerative conditions were included in two additional studies. Furthermore, in the study from Campagner et al. [33], patients with single-level lumbar discopathy treated with transforaminal lumbar interbody fusion (TLIF), anterior lumbar interbody fusion (ALIF) and extreme lateral lumbar interbody fusion (XLIF) were recruited. Included studies exploited different input sources, and more specifically: 14 studies utilized clinical data [13, 14, 16–20, 22, 27, 29–31, 33, 34], 11 exploited demographics [12, 13, 16–18, 20, 22, 23, 26, 28, 30], 3 used psychosocial data [12, 13, 16] and 1 utilized biomarkers [33]. In terms of radiological inputs, 10 studies employed MRI data [11, 19, 20, 24, 26–29, 32, 35], 1 study utilized CT data [11] and 3 studies used X-ray imaging [15, 18, 26]. Only 1 study evaluated plantar pressure data as an input source for DSS processing [25]. Regarding the task performed, 6 studies focused on the definition of specifically designed scores, i.e., for surgical invasiveness [33], cage subsidence in fusion surgery [14], risk of facet synovial cyst recurrence [28], prediction of the lumbar decompression level [19], anticipation of the most appropriate treatment for LDH [29], and a new predictive classification system for LBP [26]. 8 studies provided clinical assessment of patients by identifying risk factors for lumbar degenerative disease [12, 16, 22], predicting the progression of the disease [24, 35], or the presence of LBP [11], assessing recovery after surgery for lumbar IDD [25], and recommending referral to primary care for LBP [13]. Eventually, 11 studies performed eligibility prediction by anticipating surgical candidacy [17, 27, 32], assessing the suitability of surgery [15, 21, 23, 30, 34], recommending the most adequate endoscopic surgical corridor [31], and predicting allocation to conservative or surgical treatment [18, 20, 29].

As regards the AI models used, the vast majority of papers adopted supervised approaches employing manually labelled data in retrospective studies. Only 2 studies adopted unsupervised clustering techniques to distinguish between clinical and pathological conditions. They exploited data-driven algorithms to identify chronic LBP phenotypes, together with the most predictive profiling variables [16, 26].

The main difference in the supervised approaches lay in the use of machine learning (ML) instead of deep learning (DL) algorithms. The latter is a sub-field of the former that utilizes artificial neural networks (NN) inspired by the brain's structure and function. While DL algorithms often outperform traditional ML techniques in many applications, they require significantly larger datasets

Table 2 Summary characteristics of included studies related to clinical assessment

Study	Country	Sample size	Diagnosis	Data type	Task	Results	AI model
Aggarwal, 2021 [11]	USA	62	LBP	MRI and CT data	Predicting the presence of LBP	AUC = 0.836	LR, NB, NN, DT, RF, SVM, AdaBoost, Constant, kNN, SGD
Darvishi, 2017 [12]	Iran	160	LBP	Demographic and psychosocial data	Identifying LBP risk factors	Acc = 88%	NN
Nijeweme-d'Hollosy, 2018 [13]	The Netherlands	1326	LBP	Demographic, psychosocial and clinical data	Recommending self-referral to primary care for LBP	Acc = 71%	DT, boosted DT, RF
Cheung, 2021 [24]	China	1343	Lumbar IDD	MRI data	Predicting the progression of IDD	Acc = 90.2%, 90.4% and 89.9%	CNN
D. Liu, 2022 [25]	China	95	Lumbar IDD	Plantar pressure	Classifying subjects with lumbar IDD and assessing recovery after surgery	AUC = 0.998	SVM
Cheung, 2023 [35]	China	1152	Schmorl, HIZ Modic changes	MRI data	Predicting the progression of endplate defects	Weighted Acc: 89.5%, 91.8%, 87.5%	CNN
Abbas, 2023 [22]	Israel	345	LSS	Clinical, demographic and radiological data	Identifying the most predictive factors for degenerative LSS	The anteroposterior diameter of the bony canal at L5 and L4 levels is highly predictive	RF
Tagliaferri, 2023 [16]	Australia	42	LBP	Clinical, demographic, radiological and psychosocial	Identifying chronic LBP phenotypes and the most predictive profiling variables	Acc = 95.8%	Fuzzy C-means, SVM, kNN, NB, RF

Abbreviations: *Acc* accuracy, *AUC* area under the curve, *CNN* convolutional neural network, *CT* computed tomography, *DT* decision tree, *HIZ* high intensity zones, *IDD* intervertebral disc degeneration, *kNN* k-nearest neighbors, *LBP* low back pain, *LR* logistic regression, *LSS* lumbar spinal stenosis, *MRI* magnetic resonance imaging, *NB* Naïve Bayes, *NN* neural network, *RF* random forest, *SGD* Stochastic Gradient Descent, *SVM* support vector machine

Table 3 Summary characteristics of included studies related to eligibility prediction

Study	Country	Sample size	Diagnosis	Data type	Task	Results	AI model
Agarwal, 2022 [17]	USA	608	Grade 1 DLS	Demographic and clinical data	Predicting the surgical candidacy	Acc = 77.7%	RF
Durand, 2020 [18]	USA	1503	ASD	Clinical, demographic and X-ray data	Predicting nonsurgical vs. surgical treatment	AUC = 0.914, Acc = 86%	RF, LR, SVM, Elastic net regression
Tseng, 2020 [20]	Taiwan	103	LSS	Clinical, demographic and MRI data	Predicting conservative vs. surgical treatment	AUC = 0.952, Acc = 94.87%	LR, DT, NN, SVM
Xie, 2022 [27]	Australia	387	LBP, lower limb radiculopathy, and/or neurogenic claudication	Clinical, psychosocial and MRI data	Predicting the surgical candidacy	AUC = 0.9, Acc = 92.1%	MLP
Chen, 2022 [31]	Taiwan	899	Patients undergone endoscopic lumbar spine surgery	Clinical data	Recommending the most adequate endoscopic surgical corridor	AUC = 0.91, Acc = 85%	NN
Wilson, 2021 [32]	USA	240	Patients undergone lumbar laminectomy	MRI data	Predicting the surgical candidacy	AUC = 0.88	U-net
Bui, 2024 [15]	Taiwan	311	LBP	X-ray images	Predicting cage height and lumbar lordosis subtraction from the pelvic incidence after TLIF surgery	RMSE=1.01, 5.19; Acc = 81%	SVM, LASSO, DT, kNN, MLP
Chiu, 2023 [34]	China	181	LDDD treated with Nucleoplasty	Clinical and radiomics data	Predicting pain improvement after nucleoplasty for patients with LDDD	AUC = 0.77; Acc = 76%; F1 = 0.73	SVM, XGB, RF
Fan, 2023 [21]	China	1095	LSS	Radiomics data	Predicting decompression levels in patients with multilevel LSS	AUC = 0.86	MLP, GB, AdaBoost, LR, LDA, RF, SVM, DT, kNN, NB
Ren, 2024 [30]	China	1159	LDH	Clinical, demographic and radiological data	Predicting recurrent LDH following PELD	AUC = 0.93	NN, XGB, kNN, DT, RF, SVM
Pedersen, 2024 [23]	Denmark	6585	LSS	Demographic and patient-reported data	Predicting the improvement of different PROMs in patients with LSS one year after surgery	$0.67 \leq AUC \leq 0.88$	MARS

Abbreviations: *Acc* accuracy, *ASD* adult spine deformity, *AUC* area under the curve, *DLS* degenerative lumbar spondylolisthesis, *DT* decision tree, *GB* gradient boosting, *kNN* k-nearest neighbors, *LASSO* least absolute shrinkage and selection operator, *LBP* low back pain, *LDA* linear discriminant analysis, *LDDD* lumbar degenerative disc disease, *LDH* lumbar disc herniation, *LR* logistic regression, *LSS* lumbar spinal stenosis, *MARS* multivariate adaptive regression splines, *MLP* multilayer perceptron, *MRI* magnetic resonance imaging, *NB* Naïve Bayes, *NN* neural network, *PELD* percutaneous endoscopic lumbar discectomy, *PROM* patient reported outcome measure, *RF* random forest, *RMSE* root mean squared error, *SVM* support vector machine, *TLIF* transforaminal lumbar interbody fusion, *XGB* extreme gradient boosting

and are computationally expensive to train, especially without high-performance hardware like GPUs. These challenges likely explain why only a few of the reviewed papers extensively explored DL models. Figure 3 displays the usage percentages of the most commonly employed algorithms in the reviewed articles. As shown, Random Forest (RF), Support Vector Machine (SVM), and Fully Connected Networks (FCN) emerged as the most studied models, followed by Decision Tree (DT), Logistic Regression (LR) and k-Nearest Neighbour (kNN). For a comprehensive and detailed overview of the primary ML and DL algorithms, readers are referred to [36].

Seven studies resorted exclusively to DL algorithms [12, 19, 24, 27, 31, 32, 35], seventeen studies employed ML [17, 22, 23, 25, 28, 29], with some comparing the performances of different models [11, 13–16, 18, 20, 21, 30, 33, 34]. A detailed description of the adopted predictive strategies is reported in Results of individual studies section.

Risk of bias

Sixteen studies were rated on a 3-point scale, reflecting concerns about risk of bias and applicability as low, unclear, or high, as depicted in Fig. 4 (the detailed analysis is presented in Tables S1 and S2). With regard to risk of bias, almost 40% of studies displayed a high risk in the patient selection domain, mostly due to inaccurate reporting of eligibility criteria and inclusion of healthy controls.

Results of individual studies

Definition of clinical scores

Among the included studies, 6 were focused on the definition of clinical scores, intended as aggregated numerical indexes able to be highly predictive of clinical conditions, adverse events, or treatment appropriateness. In their study, Campagner et al. [33] cross-validated a number of ML techniques to assess and predict the invasiveness of TLIF, XLIF or ALIF in 72 patients affected by single-level discopathy between L3 and S1 levels. They found that, based on several pre- and post-surgical biomarkers as well as a ground truth score, random forest (RF) performed the best among different AI models, with an average AUC of 0.87 ± 0.05 on invasiveness classification and 0.76 ± 0.05 on invasiveness prediction. Intriguingly, the most important factors associated with invasiveness were the American Society of Anesthesiologists (ASA) grade, length of hospital stay, and serum concentration of interleukin (IL)−22, C reactive protein (CRP), and soluble cluster of differentiation 163 (sCD163). Jin et al. [26] used K-means and hierarchical agglomerative clustering on lumbar spine MRI images of 159 patients to define

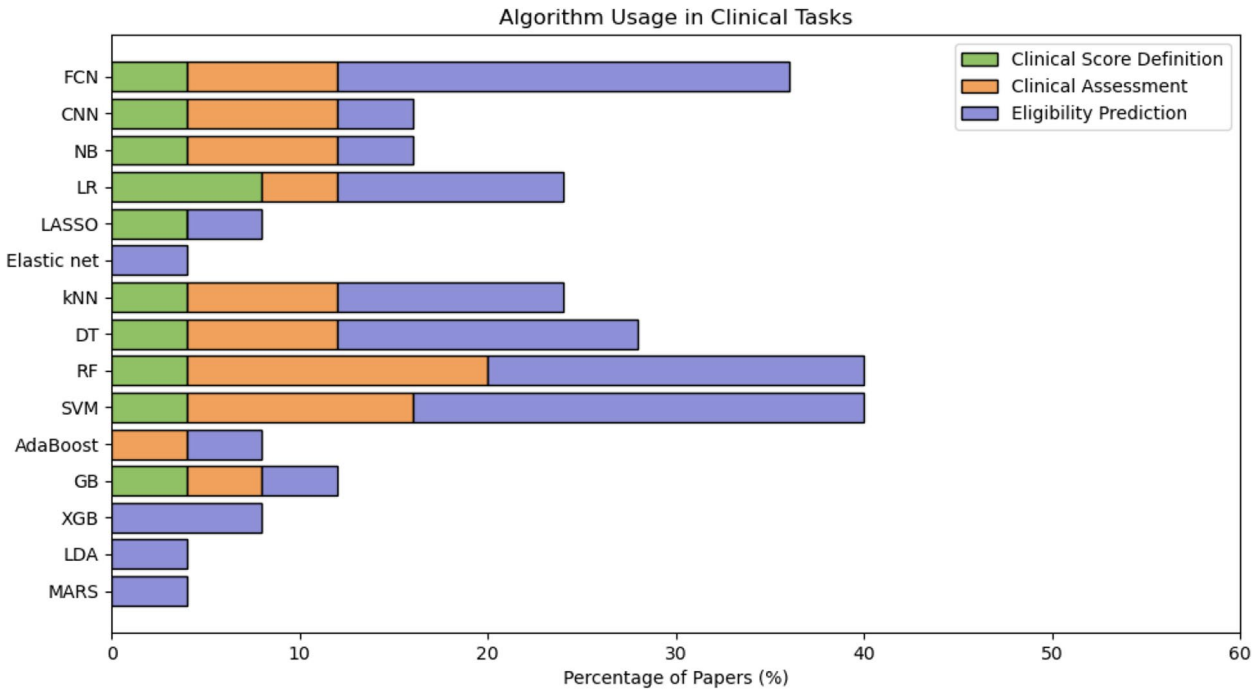

Fig. 3 Percentage values of supervised AI models' usage in the reviewed articles. CNN: Convolutional Neural Network; DT: Decision Tree; FCN: Fully Connected Network; GB: Gradient Boosting; kNN: k-Nearest Neighbour; LASSO: Least Absolute Shrinkage and Selection Operator; LDA: Linear Discriminant Analysis; LR: Logistic Regression; MARS: Multivariate Adaptive Regression Splines; NB: Naive Bayes; RF: Random Forest; SVM: Support Vector Machine; XGB: eXtreme Gradient Boosting

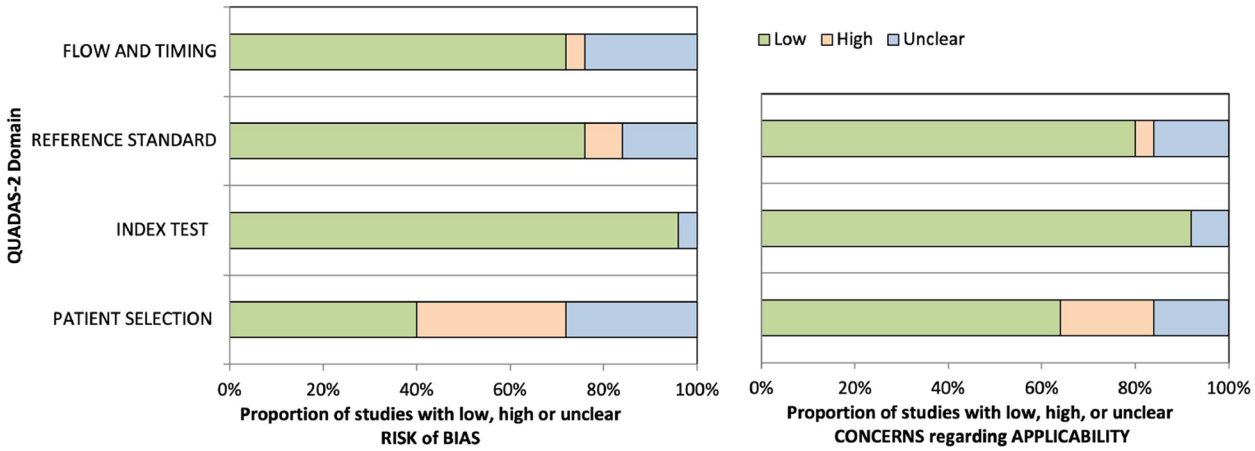

Fig. 4 Summary of the methodological quality of included studies regarding the 4 domains assessing the risk of bias (**a**) and the 3 domains assessing applicability concerns (**b**) of the QUADAS-2 score

different class indexes of LBP. Among utilized variables, spinopelvic parameters, disc height, disc signal intensity, and sociodemographic features were considered. Interestingly, they found 3 different clusters: the incidence of LBP was greater in cluster 3, which was characterized by the lowest disc height, compared to clusters 1 and 2, in which lumbar lordosis and lumbosacral angle were significantly higher. In another study, Page et al. [28] used ML models to create a predictive score for lumbar synovial cyst recurrence following surgical decompression. They managed to identify patients at high risk of reoperation with a validation AUC of 0.83, and the main risk factors were facet inclination angle > 45°, canal stenosis > 50%, T2 joint space hyperintensity and grade I spondylolisthesis. In another study, Roller et al. [19] fed 141 sagittal MRI images to a Convolutional Neural Network (CNN), namely SpineNet, to generate an aggregate score able to predict the surgical decompression level based on central canal stenosis, disc narrowing, Pfirrmann grading, spondylolisthesis, and endplate/bone marrow changes. Overall, the model showed an accuracy of 65%. Yu et al. [29] employed clinical records and MRI data of 200 patients with LDH to predict the most appropriate treatment strategy. They extracted several radiomics features after manual segmentation of the herniated disc and, thanks to least absolute shrinkage and selection operator (LASSO) and logistic regression (LR) models, they defined an overall quantitative radiomics score which, together with selected sociodemographic factors, allowed classification AUC performances up to 0.93 in predicting surgical vs. conservative treatment. Xiong et al. [14] used gradient boosting (GB) machine to predict inter-body cage subsidence after posterior lumbar fusion surgery. The authors introduced a cage subsidence score based on the postoperative disc height, achieving an AUC of 0.889 (data split 75/25%).

Clinical assessment

Another common task related to AI-based DSS pertained the evaluation of the patient clinical conditions. Aggarwal [11] assessed the LBP prediction performances of ten different AI classification models, utilizing quantitative measurements from MRI and CT lumbar scans as input data. The author obtained the best AUC score using an LR model (0.836), showing a positive association between LBP, decreased disc height and ligamentum flavum hypertrophy. In their study, Cheung et al. [24] trained a CNN on MRI images of 1343 patients (acquired at baseline and at 5 years) to predict lumbar IDD progression and achieved a prediction accuracy of 90.2%, 90.4%, and 89.9% for Schneiderman score, disc bulging, and Pfirrmann grading, respectively. Similarly, in another study [35], the same authors used a first net to segment the disc area and a CNN to predict the 5-year progression of Schmorl, HIZ and Modic, achieving weighted accuracy scores of 89.5, 91.8, 87.5%, respectively, for the three tasks. Darvishi et al. [12] trained a Feedforward NN on occupational, demographic, and psychological data from 160 workers to identify LBP risk factors, achieving a 88% test accuracy. In another study, D. Liu et al. [25] developed an active-matrix sensing array for real-time human plantar pressure monitoring and employed it as a ML-assisted platform for motion recognition, as well as lumbar degenerative disease diagnosis and postoperative recovery assessment. For each patient and control subject, they preprocessed the recorded pressure timeseries to get feasible training data and fed them into several kernel-based support vector machine (SVM)

classifiers, ultimately achieving a significant forecasting performance (AUC = 0.998). Oude Nijeweme-d'Hollosy et al. [13] fed 3 different AI models, namely Decision Tree (DT), boosted DT, and RF, with clinical data of LBP patients to predict a timely self-referral to primary care and prevent the transition from acute to chronic LBP. The models were trained with 1288 fictious LBP cases and validated on 38 real-life cases. The results showed that all the models provided a referral advice that was better than just a random guess, although the boosted DT demonstrated the highest accuracy (71%). Abbas et al. [22] employed RF to detect the essential variables that predict the development of symptomatic LSS. The authors listed the Gini index for each predictor showing that lumbar spine characteristics, rather than the demographic (e.g., age and BMI) and health data, were far more important factors in leading to development of symptomatic degenerative LSS. Unfortunately, no classification performance results were reported. Tagliaferri et al. [16] conducted a data-driven pilot study classifying chronic LBP patients into clusters based on multidimensional factors, employing ML models to assess the accuracy of classification to sub-groups. The authors showed that four factors (cognitive function, depressive symptoms, general self-efficacy and anxiety symptoms) and two phenotypes (normal versus impaired psychosocial profiles) optimally classified participants.

Eligibility prediction

Another challenging task reported in the literature concerns the support of appropriate treatment choices, especially when deciding between conservative and surgical options. Agarwal et al. [17] employed an RF model to identify a body mass index (BMI) threshold for obese patients undergoing surgical intervention for grade 1 lumbar spondylolisthesis and thus reliably identify optimal surgical candidates. They found that a BMI $\leq 37.5 kg/m^2$ was associated with improved patient outcomes following surgical intervention, with an accuracy rate of 77.7%. In another study, Chen et al. [31] applied an artificial neural network to assess the best surgical approach between transforaminal and interlaminar lumbar spinal endoscopy. Thanks to a four-layer fully connected NN, they achieved test accuracy of 85% and AUC score of 0.91, identifying as the most important factors the disc level, lesion location on the axial plane, and the direction of LDH migration. Furthermore, they designed a graphical user interface (GUI) to interact with the trained model so to provide surgeons with a tool for surgical planning. Durand et al. [18] tested several ML models to predict between conservative and surgical treatment in 1503 patients affected by ASD based on demographic, clinical, and X-ray data. They achieved AUC scores greater than 0.9 and in particular the SVM model showed an AUC of 0.91 and an accuracy of 86%, with Scoliosis Research Society (SRS), Oswestry Disability Index (ODI), and Short Form (SF)−36 being the most relevant variables for the SVM model. Similarly, Tseng et al. [20] investigated the appropriateness of conservative vs. surgical treatment in patients with LSS by means of several combinations of ML models. They applied LR and DT to extract the most significant features from demographic, clinical and MRI data, and combined them with NN and SVM to predict the factors affecting treatment choice. The prediction model which performed the best (AUC = 0.952) showed that disc height, age, difference between systolic and diastolic blood pressure, and leg pain were the most important factors involved. In their study, Wilson et al. [32] used a U-net on MRI data of 240 subjects to predict surgical candidacy among patients affected by LSS. After performing automatic segmentation of the spinal canal area, the model was able to anticipate surgical treatment based on the severity of LSS achieving an AUC of 0.88. Likewise, Xie et al. [27] developed a NN model able to predict surgical candidacy based on clinical and imaging features of patients affected by different lumbar degenerative disorders (LBP, lower limb radiculopathy, and/or neurogenic claudication). The employed model was able to predict surgical progression with 92.1% accuracy and an excellent discriminative ability (AUC = 0.90). Interestingly, the most relevant factors for recommending surgery as identified by the model were severe LSS, nerve root compression, neurogenic claudication, dermatomal distribution of symptoms, functional deterioration, presence of spondylolisthesis, pain severity, and symptom progression. Bui et al. [15] used LASSO regression and SVM to predict the cage height and the degree of lumbar lordosis subtraction from the pelvic incidence after TLIF surgery, utilizing preoperative X-ray images of 311 patients using 5-fold CV. The authors achieved an RMSE of 1.01 for cage height prediction, and 88.75% of the prediction fell within a 1mm margin from the real value. Chiu et al. [34] developed machine learning based radiomic models based on pre-treatment clinical and imaging data for predicting pain improvements after lumbar nucleoplasty in subjects with lumbar degenerative disc desease (LDDD). The authors compared different models, achieving accuracy values of 76% and an AUC score of 0.77, proving the potential of AI in supporting clinicians' decision-making about treatment strategies for LDDD. The study of Fan et al. [21] used machine learning to extract radiomic features from CT myelography images and predict decompression levels in LSS patients. Six feature selection methods combined with 12 ML algorithms were employed, resulting in a total of 72 ML classifiers

with EmbeddingLSVC and SVM showing the best performance (AUC scores over 0.80). Ren et al. [30] leveraged various AI-models to predict re-herniation before Percutaneous Endoscopic Lumbar Discectomy, aiming at optimizing the surgical decision-making. XGB outperformed other models (AUC = 0.93), identifying BMI, facet orientation, Modic changes, and disc calcification as significant predictors for recurrent LDH. Pedersen et al. [23] developed a DSS to detect pros and cons of surgical decompression in LSS patients. They employed MARS algorithm to predict the improvement in walking distance, disability, pain, and quality of life in terms of reaching the minimal clinically important difference (MCID) at 1-year follow-up after spinal surgery.

Discussion

The use of AI in the spine field has been increasingly reported in the last decade, with groundbreaking and potentially revolutionary applications. These include localization of specific objects in radiological images (e.g., detection and segmentation of intervertebral discs, spinal canal, vertebral bodies, etc.) [5], classification and regression tasks with regard to diagnostic processes (e.g., identification of vertebral fractures, estimation of the severity of LSS, etc.) [6], and supportive tools to predict outcomes and anticipate the most advantageous treatment and prevention strategies [37]. The tasks included in the last category namely pertain to DSS applications. In this study, we have systematically reviewed the available evidence on DSS systems able to support the decision-making process in terms of prevention and management of degenerative lumbar spine disorders. The earliest study included in this review dates back to 2017 [12], indicating that while this field is still relatively young, it shows promising potential for future advancements. The analyzed DSS models included datasets from patients with several degenerative conditions (nonspecific LBP, DLS, ASD, LDH, and facet joint synovial cysts), undergoing both conservative and surgical treatments (endoscopic decompression, laminectomy, interbody fusion, etc.), and were fed with a wide range of different inputs (demographic and psychosocial data, clinical outcomes, imaging data, and biomarkers). Numerous AI algorithms were employed, comprising both ML and DL; not surprisingly, the vast majority of the included studies adopted supervised learning approaches to implement support systems that replicate human performances in a faster and automatic way. Conversely, both Jin et al. [26] and Tagliaferri et al. [16] investigated the phenotypic characteristics of LBP patients by means of unsupervised techniques, trying to identify intrinsic scores being discriminative for a novel classification system. Broadly speaking, supervised learning is more focused on automatically labelling new data according to hand-made scoring patterns, thus being more suitable for supporting specific decision processes. On the other hand, unsupervised methods are domain-agnostic because they investigate the hidden patterns, structures, or relationships within the data without explicit guidance or predefined labels. In general, one of the main advantages of AI is the unprecedented capacity to elaborate large and heterogeneous datasets in order to seek correlations and causative relationships which would be otherwise unidentifiable. Apart from computing power, the ability to integrate different sources of multimodal inputs may provide a new understanding of pathophysiological processes and promote the identification of novel factors that could potentially affect the diagnosis, treatment, and prognosis of several conditions [38]. In accordance with the biopsychosocial model, LBP results from a spectrum of diverse factors (biological, psychosocial, demographic, etc.) that may differentially contribute to the clinical scenario in every single case [39]. In this regard, AI may help identify distinct disease phenotypes to develop a personalized medicine approach to LBP and degenerative spine disorders [40]. In this review, the included studies concerned three main areas: definition of clinical scores, clinical assessment, and eligibility prediction. The first domain found different applications such as prediction of surgical invasiveness based on perioperative outcomes and serum biomarkers [33], prediction of LBP according to spinopelvic parameters [26], estimation of the risk of recurrence of surgically treated facet synovial cysts based on facet morphology and canal stenosis [28], and prediction of conservative vs. surgical treatment for LSS [19] and LDH [29] using radiological and psychosocial data. Overall, these studies showed AUCs ranging from 0.76 to 0.93 thus demonstrating a considerably high degree of accuracy, with a slightly lower rate of 65% in the study of Roller and colleagues [19]. With regard to clinical assessment, included studies evaluated the risk of LBP [11, 12], IDD diagnosis [25] and progression [24], as well as the need for referral to primary care to prevent the risk of LBP chronicity [13]. Utilized AI algorithms were fed by imaging data, demographic and psychosocial characteristics, biosensor inputs, and clinical notes. Again, these studies showed accuracies and AUCs in the ranges of 71–90.4% and 0.84–0.99 respectively, demonstrating the significant potential of AI-based DSS models. These results confirm the emerging role of lifestyle, occupation, underlying systemic conditions, biomarkers, and imaging data patterns in the definition and management of degenerative lumbar disorders using a "precise" approach [41]. Concerning eligibility prediction, included studies evaluated surgical candidacy for DLS [17], ASD [18] and LSS [20, 32] and other degenerative disorders [27]

according to demographics, biometrics, MRI data, and patient-reported outcome measures, resulting in AUCs between 0.88 and 0.95 and accuracy rates between 77.7% and 94.9%. Additionally, another study employed a DSS system to support preoperative planning for endoscopic lumbar decompression based on MRI features, with a similar high accuracy and AUC of 85% and 0.91, respectively [31]. These results are in line with other studies in different surgical specialties, which have repeatedly underlined the promising value of DSS models in surgical decision-making [42].

We identified several research gaps among the studies reported in this review. First, there is still a lack of multimodal approaches to AI predictions in this field, combining both visual features extracted through convolutional models and numerical features clinically reported. Indeed, half of the articles employing NNs fed fully connected models with numerical features, being they clinical, demographic, psychosocial or radiology-derived measures [12, 27, 31]; the other half instead directly fed radiological images into CNN models [19, 24, 32]. A hybrid approach combining clinical and visual features deserves to be explored, having already proved to be a flexible and robust method to improve predictive capacity in healthcare as compared to single-modality approaches [43]. Indeed, by boosting data acquisition and elaboration, providing multimodal acquisition and processing, which ultimately improves accuracy, the use of AI may overcome the subjective aspects of traditional decision-making and eventually optimize patient outcomes [44]. Another issue concerns the evaluation metrics adopted by the different research papers. While most studies report performance metrics such as AUC and accuracy, these metrics alone may not provide a comprehensive view of model performance, particularly when dealing with imbalanced datasets, as often happens in the clinical domain. A high AUC, for instance, may mask poor performance in identifying minority class instances, such as true positives or true negatives, which are critical in clinical decision-making. Therefore, additional metrics such as recall, precision, and F1 score for classification models would offer a complete understanding of model performance. Furthermore, some of the reviewed studies [18, 21, 24, 27, 28] did not adequately address the issue of class imbalance, which can significantly skew accuracy metrics. Providing more information on the class distribution within the datasets would enable a clearer assessment of how well the models perform across different class instances. Finally, many studies lacked the use of proper scoring rules like the Log score or Brier score, and did not include calibration plots, which are essential for evaluating the reliability of probabilistic models. Incorporating these tools in future studies would allow for a more accurate evaluation of the model's ability to produce well-calibrated probability estimates, which is crucial in clinical applications.

A set of practical challenges common to the application of AI to healthcare are also associated with implementing DSS in LBP. Effective integration of structured and unstructured data, the need of training healthcare professionals to the utilization of the AI tools, the compliance with healthcare regulations including transparency and accountability in decision-making processes, the ethical and privacy concerns surrounding the use of AI in healthcare, and the secure handling of sensitive health data are essential to build trust and ensure the responsible use of AI in healthcare [45].

This study has some limitations. First, the average QUADAS-2 score was indicative of a moderate risk of bias, which may have an impact on the reliability of the data that were reported. Second, as the included studies exhibited a high number of different variables, the heterogeneity among diverse disorders, input data and AI models used may reduce the generalizability of the findings relative to the single investigations. Moreover, while the aim of this study was to review the applications of DSS to prevention and management tasks, a clear distinction with the domain of outcome prediction was not always possible (e.g., eligibility for a certain treatment may be provided on the basis of multiple outcome prediction scores). Therefore, some studies might have been included or excluded based on the definition of their principal endpoint while possibly performing both tasks.

Conclusions

Artificial intelligence is progressively revolutionizing the healthcare field, including spine surgery. In this systematic review, we have summarized the available evidence regarding the use of DSS models to support the prevention and management of lumbar degenerative disorders. Overall, the included studies demonstrated a high degree of accuracy in performing a wide set of different tasks, such as the definition of clinical scores, clinical assessment, and eligibility prediction. Collectively, these results pose the basis for further studies to expand the actual knowledge base and promote a wider implementation of AI in the decision-making process in the clinical setting.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12891-025-08356-x.

Supplementary Material 1.

Acknowledgements
Not applicable.

Authors' contributions
PG, FDA, FR, MM developed the initial hypothesis and designed the study framework. PG, FDA, FR, LA, GFP, OdA were involved in the gathering of the data utilized in the study. PG, FDA, LA, OdA, AC, LV performed the data analysis and interpreted the findings. PG, FDA, FR, LA, GFP drafted the initial version of the manuscript, incorporating the core ideas and findings. GV, AC, LV, MM, RP, VD provided significant revisions and input to enhance the manuscript's intellectual depth. FR, GV, RP, VD secured the funding that facilitated the execution of the research. FR, GV, AC, LV, MM, RP, VD oversaw the research process, ensuring adherence to methodological and ethical standards. All authors read and approved the final manuscript.

Funding
This research was funded by the Research Grants (BRIC-2022 ID28 - ID30) of the Italian Workers' Compensation Authority (INAIL) and by the European Union - Next Generation EU - NRRP M6C2 - Investment 2.1 Enhancement and strengthening of biomedical research in the NHS, project n. PNRR-MAD-2022-12376692_VADALA.

Data availability
All data generated or analysed during this study are included in this published article.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Fondazione Policlinico Universitario Campus Bio-Medico, Via Alvaro del Portillo, 200, Rome 00128, Italy. [2]Research Unit of Intelligent Technology for Health and Wellbeing, Department of Engineering, Università Campus Bio-Medico di Roma, Via Alvaro del Portillo, 21, Rome 00128, Italy. [3]Operative Research Unit of Orthopaedic and Trauma Surgery, Fondazione Policlinico Universitario Campus Bio-Medico, Via Alvaro del Portillo, 200, Rome 00128, Italy. [4]Research Unit of Orthopaedic and Trauma Surgery, Department of Medicine and Surgery, Università Campus Bio-Medico di Roma, Via Alvaro del Portillo, 21, Rome 00128, Italy. [5]Research Unit of Computer Systems and Bioinformatics, Department of Engineering, Università Campus Bio-Medico di Roma, Via Alvaro del Portillo, 21, Rome 00128, Italy. [6]Ri.MED Foundation, Via Bandiera, 11, Palermo 90133, Italy.

Received: 8 May 2024 Accepted: 24 January 2025
Published online: 07 February 2025

References

1. Tagliaferri SD, Angelova M, Zhao X, Owen PJ, Miller CT, Wilkin T, et al. Artificial intelligence to improve back pain outcomes and lessons learnt from clinical classification approaches: three systematic reviews. NPJ Digit Med. 2020;3(1):93.
2. Russo F, Papalia GF, Vadalà G, Fontana L, Iavicoli S, Papalia R, et al. The effects of workplace interventions on low back pain in workers: a systematic review and meta-analysis. Int J Environ Res Public Health. 2021;18(23):12614.
3. Faiella E, Santucci D, Calabrese A, Russo F, Vadalà G, Zobel BB, et al. Artificial intelligence in bone metastases: an MRI and CT imaging review. Int J Environ Res Public Health. 2022;19(3):1880.
4. Galbusera F, Casaroli G, Bassani T. Artificial intelligence and machine learning in spine research. JOR Spine. 2019;2(1):e1044. https://doi.org/10.1002/jsp2.1044.
5. D'Antoni F, Russo F, Ambrosio L, Vollero L, Vadalà G, Merone M, et al. Artificial Intelligence and Computer Vision in Low Back Pain: A Systematic Review. Int J Environ Res Public Health. 2021;18(20):10909. https://doi.org/10.3390/ijerph182010909.
6. D'Antoni F, Russo F, Ambrosio L, Bacco L, Vollero L, Vadalà G, et al. Artificial intelligence and computer aided diagnosis in chronic low back pain: A systematic review. Int J Environ Res Public Health. 2022;19(10):5971.
7. Bacco L, Russo F, Ambrosio L, D'Antoni F, Vollero L, Vadalà G, et al. Natural language processing in low back pain and spine diseases: A systematic review. Front Surg. 2022;9:957085.
8. Downie AS, Hancock M, Abdel Shaheed C, McLachlan AJ, Kocaballi AB, Williams CM, et al. An electronic clinical decision support system for the management of low back pain in community pharmacy: development and mixed methods feasibility study. JMIR Med Inform. 2020;8(5):e17203.
9. Kubben P, Van Santbrink H, Cornips E, Vaccaro AR, Dvorak M, Van Rhijn L, et al. An evidence-based mobile decision support system for subaxial cervical spine injury treatment. Surg Neurol Int. 2011;2:32
10. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.
11. Aggarwal N, et al. Prediction of low back pain using artificial intelligence modeling. JMAI J Released March. 2021.
12. Darvishi E, Khotanlou H, Khoubi J, Giahi O, Mahdavi N. Prediction effects of personal, psychosocial, and occupational risk factors on low back pain severity using artificial neural networks approach in industrial workers. J Manipulative Physiol Ther. 2017;40(7):486–93. https://doi.org/10.1016/j.jmpt.2017.03.012.
13. Nijeweme-d'Hollosy WO, van Velsen L, Poel M, Groothuis-Oudshoorn CG, Soer R, Hermens H. Evaluation of three machine learning models for self-referral decision support on low back pain in primary care. Int J Med Inform. 2018;110:31–41. https://doi.org/10.1016/j.ijmedinf.2017.11.010.
14. Xiong T, Wang B, Qin W, Yang L, Ou Y. Development and validation of a risk prediction model for cage subsidence after instrumented posterior lumbar fusion based on machine learning: a retrospective observational cohort study. Front Med. 2023;10:1196384.
15. Bui AT, Le H, Hoang TT, Trinh GM, Shao HC, Tsai PI, et al. Development of End-to-End Artificial Intelligence Models for Surgical Planning in Transforaminal Lumbar Interbody Fusion. Bioengineering. 2024;11(2):164.
16. Tagliaferri SD, Owen PJ, Miller CT, Angelova M, Fitzgibbon BM, Wilkin T, et al. Towards data-driven biopsychosocial classification of non-specific chronic low back pain: a pilot study. Sci Rep. 2023;13(1):13112.
17. Agarwal N, Aabedi AA, Chan AK, Letchuman V, Shabani S, Bisson EF, et al. Leveraging machine learning to ascertain the implications of preoperative body mass index on surgical outcomes for 282 patients with preoperative obesity and lumbar spondylolisthesis in the Quality Outcomes Database. J Neurosurg Spine. 2022;1(aop):1–10.
18. Durand WM, Daniels AH, Hamilton DK, Passias P, Kim HJ, Protopsaltis T, et al. Artificial intelligence models predict operative versus nonoperative management of patients with adult spinal deformity with 86% accuracy. World Neurosurg. 2020;141:e239–53.
19. Roller BL, Boutin RD, O'Gara TJ, Knio ZO, Jamaludin A, Tan J, et al. Accurate prediction of lumbar microdecompression level with an automated MRI grading system. Skeletal Radiol. 2021;50(1):69–78. https://doi.org/10.1007/s00256-020-03505-w.
20. Tseng LP, Pei YC, Chen YS, Hou TH, Ou YK. Choice between Surgery and Conservative Treatment for Patients with Lumbar Spinal Stenosis: Predicting Results through Data Mining Technology. Appl Sci. 2020;10(18):6406.
21. Fan G, Wang D, Li Y, Xu Z, Wang H, Liu H, et al. Machine Learning Predicts Decompression Levels for Lumbar Spinal Stenosis Using Canal Radiomic Features from Computed Tomography Myelography. Diagnostics. 2023;14(1):53.

22. Abbas J, Yousef M, Peled N, Hershkovitz I, Hamoud K. Predictive factors for degenerative lumbar spinal stenosis: a model obtained from a machine learning algorithm technique. BMC Musculoskelet Disord. 2023;24(1):218.
23. Pedersen CF, Andersen MØ, Carreon LY, Skov ST, Doering P, Eiskjær S. PROPOSE. Development and validation of a prediction model for shared decision making for patients with lumbar spinal stenosis. N Am Spine Soc J (NASSJ). 2024;17:100309.
24. Cheung JPY, Kuang X, Lai MKL, Cheung KMC, Karppinen J, Samartzis D, et al. Learning-based fully automated prediction of lumbar disc degeneration progression with specified clinical parameters and preliminary validation. Eur Spine J. 2021;1–9. https://doi.org/10.1007/s00586-021-07020-x.
25. Liu D, Zhang D, Sun Z, Zhou S, Li W, Li C, et al. Active-matrix sensing array assisted with machine-learning approach for lumbar degenerative disease diagnosis and postoperative assessment. Adv Funct Mater. 2022;32(21):2113008.
26. Jin L, Jiang C, Gu L, Jiang M, Shi Y, Qu Q, et al. Predictive Classification System for Low Back Pain Based on Unsupervised Clustering. Glob Spine J. 2021;21925682211001813. https://doi.org/10.1177/21925682211001813.
27. Xie N, Wilson PJ, Reddy R. Use of machine learning to model surgical decision-making in lumbar spine surgery. Eur Spine J. 2022;31(8):2000–6.
28. Page PS, Greeneway GP, Ammanuel SG, Resnick DK. Creation and validation of a predictive model for lumbar synovial cyst recurrence following decompression without fusion. J Neurosurg Spine. 2022;1(aop):1–4.
29. Yu G, Yang W, Zhang J, Zhang Q, Zhou J, Hong Y, et al. Application of a nomogram to radiomics labels in the treatment prediction scheme for lumbar disc herniation. BMC Med Imaging. 2022;22(1):1–12.
30. Ren G, Liu L, Zhang P, Xie Z, Wang P, Zhang W, et al. Machine learning predicts recurrent lumbar disc herniation following percutaneous endoscopic lumbar discectomy. Global Spine J. 2024;14(1):146–52.
31. Chen CM, Chen PC, Chen YC, Wang GC. Use artificial neural network to recommend the lumbar spinal endoscopic surgical corridor. Tzu Chi Med J. 2022;34(4):434–40.
32. Wilson B, Gaonkar B, Yoo B, Salehi B, Attiah M, Villaroman D, et al. Predicting spinal surgery candidacy from imaging data using machine learning. Neurosurgery. 2021;89(1):116–21. https://doi.org/10.1093/neuros/nyab085.
33. Campagner A, Berjano P, Lamartina C, Langella F, Lombardi G, Cabitza F. Assessment and prediction of spine surgery invasiveness with machine learning techniques. Comput Biol Med. 2020;121:103796.
34. Chiu PF, Chang RCH, Lai YC, Wu KC, Wang KP, Chiu YP, et al. Machine Learning Assisting the Prediction of Clinical Outcomes following Nucleoplasty for Lumbar Degenerative Disc Disease. Diagnostics. 2023;13(11):1863.
35. Cheung JPY, Kuang X, Zhang T, Wang K, Yang C. 5-Year progression prediction of endplate defects: Utilizing the EDPP-Flow convolutional neural network based on unbalanced data. J Orthop. 2023;38:7–13.
36. Chauhan NK, Singh K. A review on conventional machine learning vs deep learning. In: 2018 International conference on computing, power and communication technologies (GUCON). Greater Noida: IEEE; 2018. p. 347–52.
37. Hornung AL, Hornung CM, Mallow GM, Barajas JN, Rush A III, Sayari AJ, et al. Artificial intelligence in spine care: current applications and future utility. Eur Spine J. 2022;31(8):2057–81.
38. Young RR. Emerging role of artificial intelligence and big data in spine care. Int J Spine Surg. 2023;17(S1):S3–10.
39. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, et al. What low back pain is and why we need to pay attention. Lancet. 2018;391(10137):2356–67.
40. Khan O, Badhiwala JH, Grasso G, Fehlings MG. Use of machine learning and artificial intelligence to drive personalized medicine approaches for spine care. World Neurosurg. 2020;140:512–8.
41. Samartzis D, Alini M, An HS, Karppinen J, Rajasekaran S, Vialle L, et al. Precision spine care: a new era of discovery, innovation, and global impact. Los Angeles: SAGE Publications Sage CA; 2018.
42. Dudchenko A, Kopanitsa GD. Decision Support Systems in Cardiology: A Systematic Review. pHealth. 2017;209–14.
43. Soenksen LR, Ma Y, Zeng C, Boussioux L, Villalobos Carballo K, Na L, et al. Integrated multimodal artificial intelligence framework for healthcare applications. NPJ Digit Med. 2022;5(1):149.
44. Loftus TJ, Tighe PJ, Filiberto AC, Efron PA, Brakenridge SC, Mohr AM, et al. Artificial intelligence and surgical decision-making. JAMA Surg. 2020;155(2):148–58.
45. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People-an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach. 2018;28:689–707.

Publisher's Note