Brain and Spine 4 (2024) 102809

Contents lists available at ScienceDirect

Brain and Spine

journal homepage: www.journals.elsevier.com/brain-and-spine

Sensitivity and specificity of machine learning and deep learning algorithms in the diagnosis of thoracolumbar injuries resulting in vertebral fractures: A systematic review and meta-analysis

Hakija Bečulić [a,b], Emir Begagić [c,*], Amina Džidić-Krivić [d], Ragib Pugonja [b], Namira Softić [a], Binasa Bašić [e], Simon Balogun [f], Adem Nuhović [g], Emir Softić [h], Adnana Ljevaković [e], Haso Sefo [i], Sabina Šegalo [j], Rasim Skomorac [b,k], Mirza Pojskić [l]

[a] Department of Neurosurgery, Cantonal Hospital Zenica, Crkvice 67, 72000, Zenica, Bosnia and Herzegovina
[b] Department of Anatomy, School of Medicine, University of Zenica, Travnička 1, 72000, Zenica, Bosnia and Herzegovina
[c] Department of General Medicine, School of Medicine, University of Zenica, Travnička 1, 72000, Zenica, Bosnia and Herzegovina
[d] Department of Neurology, Cantonal Hospital Zenica, Crkvice 67, 72000, Zenica, Bosnia and Herzegovina
[e] Department of Neurology, General Hospital Travnik, Kalibunar Bb, 72270, Travnik, Bosnia and Herzegovina
[f] Division of Neurosurgery, Department of Surgery, Obafemi Awolowo University Teaching Hospitals Complex, Ilesa Road PMB 5538, 220282, Ile-Ife, Nigeria
[g] Department of General Medicine, School of Medicine, University of Sarajevo, Univerzitetska 1, 71000, Sarajevo, Bosnia and Herzegovina
[h] Department of Patophysiology, School of Medicine, University of Zenica, Travnička 1, 72000, Zenica, Bosnia and Herzegovina
[i] Neurosurgery Clinic, University Clinical Center Sarajevo, Bolnička 25, 71000, Sarajevo, Bosnia and Herzegovina
[j] Department of Laboratory Technologies, Faculty of Health Siences, University of Sarajevo, Stjepana Tomića 1, 71000, Sarajevo, Bosnia and Herzegovina
[k] Department of Surgery, School of Medicine, University of Zenica, Travnička 1, 72000, Zenica, Bosnia and Herzegovina
[l] Department of Neurosurgery, University Hospital Marburg, Baldingerstr., 35033, Marburg, Germany

ARTICLE INFO

Handling Editor: F Kandziora

Keywords:
Thoracolumbar injuries
Vertebral fractures
Machine learning
Deep learning
Artificial intelligence

ABSTRACT

Introduction: Clinicians encounter challenges in promptly diagnosing thoracolumbar injuries (TLIs) and fractures (VFs), motivating the exploration of Artificial Intelligence (AI) and Machine Learning (ML) and Deep Learning (DL) technologies to enhance diagnostic capabilities. Despite varying evidence, the noteworthy transformative potential of AI in healthcare, leveraging insights from daily healthcare data, persists.
Research question: This review investigates the utilization of ML and DL in TLIs causing VFs.
Materials and methods: Employing Preferred Reporting Items for Systematic Reviews and Meta-Analyzes (PRISMA) methodology, a systematic review was conducted in PubMed and Scopus databases, identifying 793 studies. Seventeen were included in the systematic review, and 11 in the meta-analysis. Variables considered encompassed publication years, geographical location, study design, total participants (14,524), gender distribution, ML or DL methods, specific pathology, diagnostic modality, test analysis variables, validation details, and key study conclusions. Meta-analysis assessed specificity, sensitivity, and conducted hierarchical summary receiver operating characteristic curve (HSROC) analysis.
Results: Predominantly conducted in China (29.41%), the studies involved 14,524 participants. In the analysis, 11.76% (N = 2) focused on ML, while 88.24% (N = 15) were dedicated to deep DL. Meta-analysis revealed a sensitivity of 0.91 (95% CI = 0.86–0.95), consistent specificity of 0.90 (95% CI = 0.86–0.93), with a false positive rate of 0.097 (95% CI = 0.068–0.137).
Conclusion: The study underscores consistent specificity and sensitivity estimates, affirming the diagnostic test's robustness. However, the broader context of ML applications in TLIs emphasizes the critical need for standardization in methodologies to enhance clinical utility.

* Corresponding author.
E-mail address: begagicem@gmail.com (E. Begagić).

https://doi.org/10.1016/j.bas.2024.102809
Received 1 January 2024; Received in revised form 13 March 2024; Accepted 4 April 2024
Available online 17 April 2024

1. Introduction

The highest incidence of spinal injuries occurs at the transition from the thoracic to the lumbar spine and is often due to high-energy trauma (Dai, 2012). Spontaneous incidents can occur in patients with spinal disease and result in disruption of the ligamentous apparatus and compression of nerve structures (Singleton and Hefner, 2023). Thoracolumbar (TL) fractures are more common in men aged 20–40 years. Flexion is the primary force that causes injury, sometimes in combination with compression, distraction or splitting forces, while extension injuries are rare but life-threatening (Postma et al., 2015).

Globally, the occurrence of traumatic spinal injuries stands at 10.5 cases per 100,000 individuals annually, with approximately 37.3% of these incidents resulting in spinal cord injuries (Barbiellini Amidei et al., 2022). From 2010 to 2017, there was an increase in spinal fracture incidences, rising from 21.5 to 24.0 cases per 100,000 inhabitants. The primary causes were falls from the same level, categorized as low-energy, and traffic accidents, categorized as high-energy. Among all patients, 42% were elderly individuals aged 65 years and older, as noted by Smits et al. (2020). Predominantly, spinal fractures occurred in the thoracic spine, followed by the lumbar and cervical regions. Falls from height were the leading cause of injury, followed by traffic accidents. den Ouden et al. (2019) reported that spinal cord injury was observed in 8.5% of cases, with associated injuries documented in 73% of patients. The annual occurrence of spinal cord injuries across European nations ranges from 13.9 to 19.4 per million population, while in North America, it ranges from 43.3 to 51 per million, as per Lenehan et al. (2009). Early pre-hospital mortality rates vary from 48.3% to 79%, while inpatient mortality rates range from 4.4% to 16.7% (Lenehan et al., 2009). In the United States, the prevalence of spinal cord injury falls between 721 and 906 per million population, while in Australia and Europe, it ranges from 681 to 280 per million, respectively (Lenehan et al., 2009). The reported annual incidence rate of traumatic spinal fractures, excluding those due to osteoporosis, varies between 19 and 88 per 100,000 individuals (Lenehan et al., 2009).

Males are affected 3.37 times more frequently than women, with the cervical spine being the most frequently affected (46.02%) and the lumbar spine the least (24.8%). The most common mechanisms of injury include road traffic accidents (39.5%) and falls (38.8%), while reported mortality ranges from 0% to 60%, and 36.4–59.1% of patients undergo surgery (Kumar et al., 2018). TL spine fractures account for 10% of skeletal injuries and are commonly observed at the junction of the thoracic and sacral regions (Fernández-de Thomas and De Jesus, 2023). Clinicians face the challenge of diagnosing TL fractures in a timely manner, which has prompted the integration of artificial intelligence (AI) and machine learning (ML) technologies into clinical practice (Sharma, 2023). Despite varying evidence, the potential for AI to transform healthcare by extracting insights from everyday healthcare data is significant (Bečulić et al., 2024).

ML involves computer learning and problem solving through algorithms categorized as 'supervised'', 'unsupervised' and 'reinforcement learning' (Sarker, 2021). "Big Data" has driven AI in medical diagnostics, particularly in spinal imaging, with promising results (Young, 2023). ML, including deep learning (DL) with neural networks, shows potential in the assessment of spinal disorders. In traumatic TL spinal injuries (TLI), ML supports personalized medicine and improves diagnoses, treatment prognoses and cost calculations (Karabacak and Margetis, 2023). Despite the recognized benefits, studies are limited due to the lack of efficient ML algorithms. This review examines ML and DL for the rapid diagnosis of TL spinal injury, comparing them with conventional methods and highlighting the potential clinical benefits.

2. Material and methods

2.1. Study methodology and registration

A comprehensive review was systematically conducted to evaluate the current use of ML and DL in diagnostic procedures for VF associated with TLI. The research methodology adhered to the established procedural framework described in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyzes) guidelines (Page et al., 2021). This systematic review was registered in the Open Science Framework (OSF) registry under the unique identifier OSF-REGISTRATIONS-RE5YP-V1.

2.2. Search strategy

On September 15, 2023, a thorough review of English-language publications was performed using the PubMed and Scopus databases. The search utilized key terms such as "deep learning," "machine learning," and "thoracolumbar injuries," or "thoracolumbar vertebral fractures," employing the PICOS strategy outlined in Table 1 for the PubMed search. A similar search approach was executed in the Scopus database. Further elaboration on the search methodology is available in Table 2.

2.3. Inclusion and exclusion criteria

Rigorous inclusion and exclusion criteria were applied when conducting this study to ensure a methodical and targeted selection of articles. Articles that were eligible for inclusion had to fulfill certain criteria: they had to be written in English, be directly related to the convergence of DL and ML in the detection of TLI and contain relevant data that meet the objectives of the study. Conversely, exclusion criteria were used to further refine the selection process. Articles in categories as book chapters, conference papers, reviews, non-English language literature, animal studies, and original articles without relevant data were excluded.

A total of 793 entries were found in PubMed and Scopus, and 398 duplicates were removed. After screening 395 unique entries, 21 were excluded due to non-discoverability. A total of 374 records were screened for eligibility, resulting in the exclusion of 357 reports based on specific criteria, such as book or book chapters, conference papers, reviews, non-English language literature, animal studies, and missing relevant data. Finally, 17 studies were included in the review, reflecting a systematic and careful approach to ensure the selection of articles that directly aligned with the aims and criteria of the study (Fig. 1).

2.4. Data extraction, synthesis and statistical analysis

The data extracted from the studies that meet the criteria include information on the authors and year of publication, the geographical location of the study, the study design used, the total number of patients with gender distribution, the machine learning or deep learning methods used, the specific pathology or research focus, the diagnostic modality, the variables considered in the test analysis, the details of the internal and external validation and the main conclusions of the study.

Table 1
PICOS search strategy.

Acronym	Search strategy
P (*population or problem*)	Thoracolumbar injuries OR thoracolumbar vertebral fractures
I (*intervention*)	Machine learning or deep learning – assisted radiological analysis
C (*comparison*)	None
O (*outcome*)	None
S (*study design*)	Original research studies

Table 2
Search strategy.

Search	(Machine learning OR deep learning) AND (thoracolumbar injuries OR thoracolumbar vertebral fractures OR spine)
Filter	none
Search details	("machine learning" [MeSH Terms] OR ("machine" [All Fields] AND "learning" [All Fields]) OR "machine learning" [All Fields] OR ("deep learning" [MeSH Terms] OR ("deep" [All Fields] AND "learning" [All Fields]) OR "deep learning" [All Fields])) AND (("thoracolumbar" [All Fields] AND ("injurie" [All Fields] OR "injuried" [All Fields] OR "injuries" [MeSH Subheading] OR "injuries" [All Fields] OR "wounds and injuries" [MeSH Terms] OR ("wounds" [All Fields] AND "injuries" [All Fields]) OR "wounds and injuries" [All Fields] OR "injurious" [All Fields] OR "injury s" [All Fields] OR "injuryed" [All Fields] OR "injurys" [All Fields] OR "injury" [All Fields])) OR ("thoracolumbar" [All Fields] AND ("spinal fractures" [MeSH Terms] OR ("spinal" [All Fields] AND "fractures" [All Fields]) OR "spinal fractures" [All Fields] OR ("vertebral" [All Fields] AND "fractures" [All Fields]) OR "vertebral fractures" [All Fields])) OR ("spine" [MeSH Terms] OR "spine" [All Fields] OR "spines" [All Fields] OR "spine s" [All Fields]))

The studies included in the meta-analysis had to provide data on false positive (FP), false negative (FN), true negative (TN) and true positive (TP) results to enable statistical analysis. In the absence of this information, the data required to calculate FP, FN, TN and TP were based on prevalence, sensitivity, specificity and sample size, and the calculation followed the guidelines of Rosner (2015). This approach allows the estimation of key parameters that are crucial for meta-analysis and improves the quality of statistical analysis in the absence of direct FP, FN, TN and TP data. After data processing, a random-effects meta-analysis was performed to assess the sensitivity and specificity of the included studies using Meta-DiSc 2.0. This analysis step was performed using the Shiny R application developed by Plana et al. (2022), which provides an integrated platform for the analysis and visualization of results. The MetaDTA shiny R application by Nyaga and Arbyn (2022) was used to analyze the hierarchical summary receiver operating characteristic curve (HSROC). The same approach was used to calculate logit-transformed sensitivity, logit-transformed specificity and measures such as diagnostic odds ratio (DOR) and false positive rate (FPR).

2.5. Risk of bias and applicability assessment

Risk of bias and applicability were assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) instrument (Whiting et al., 2011). Two authors (E.B. & H.B.) performed the assessment independently, and discrepancies in scores were resolved by consensus of all authors. To investigate the presence of weighted publication bias, Deek's funnel plot was used (Mizutani et al., 2023).

3. Results

3.1. Main research findings and trends

The total number of included studies was 17, and all were retrospective (Table 3). Of these studies, 52.9% (N = 9) were published in 2023, with smaller proportions in 2020 (N = 1; 5.88%), 2021 (N = 2; 11.76%) and 2022 (N = 5; 29.41%) (Fig. 2). Most studies were conducted in China (N = 5; 29.41%), followed by Taiwan and South Korea (N = 3; 17.65%) with the same proportion. Japan contributed with 11.76 % (N = 2), while Australia, Switzerland, the USA and Italy each accounted for 5.88 % (N = 1) (Fig. 3). The total number of participants was 14,524, and three studies reported the number of men and women in the cohorts studied, giving a male to female ratio of 0.52:1. In the analysis, 11.76% (N = 2) of the studies focused on ML, while the majority, 88.24% (N = 15), were devoted to DL. All included studies dealt with fractures of the TL spine.

Three studies (17.65%) investigated fractures associated with osteoporosis, while five studies (29.41%) investigated different types of fractures and the ability of ML and DL to differentiate between them. One study (5.88%) addressed sports injuries, while the remaining studies (47.06%) investigated various forms and causes of TL injuries. Six studies (35.29%) used radiographs to analyze with ML and DL

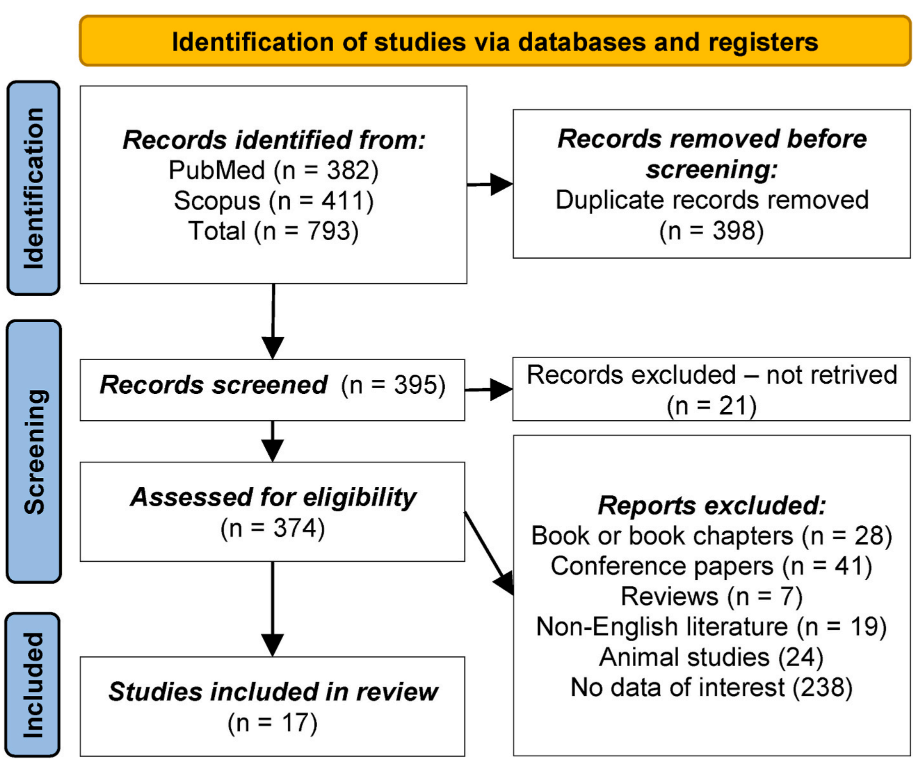

Fig. 1. PRISMA flowchart.

Table 3
Data summary of included studies.

Author (Year)	Country	Study design	Number of patients (Male/ Female)	ML/DL algorithm or Method	Pathology or focus	Diagnostic modality	Diagnostic test analysis related variables	Validation (internal and external)	Conclusions
Murata et al. (2020)	Japan	Retro	300 (n/d)	DCNN	TL VFs	RTG	Acc: 0.86 Se: 0.847 Sp: 0.873	I: Yes E: No	The DCNN algorithm identifies VF on PTLR with high accuracy and sensitivity
Li et al. (2021)	Taiwan	Retro	941 (n/d)	DL	TL VFs	CT, RTG, MRI	Acc: 0.89 Se: 0.83 Sp: 0.95	I: No E: Yes	Artificial intelligence model detected vertebral fractures on plain lateral radiographs with high accuracy, sensitivity and specificity, especially for osteoporotic lumbar fractures
Chen et al. (2021)	Taiwan	Retro	438 (n/d)	DL, DCNN	VF	RTG	Acc: 0.7359 Se: 0.7381 Sp: 0.7302 AUC: 0.72	I: Yes E: Yes	The algorithm trained by a DCNN to identify VFs on PARs showed the potential of delivering a highly accurate and acceptable specific rate and is expected to be useful as a screening tool
Ma et al. (2023)	China	Retro	529 (n/d)	ML	OVCF after PKP (NVCF)	MRI	Se: 0.907 Sp: 0.939 AUC: 0.923	I: Yes E: No	Machine learning performed better than logistic regression in predicting mew fractures after OVCF.
Chen and Liu (2022)	China	Retro	198 (n/d)	DL, RCNN	TL VFs	CT	Acc: 0.864 Se: L typ A 0.967 L typ B 0.777 T typ A 0.902 T typ B 0.786 Sp: 0.730 L typ A 0.957 L typ B 1.000 T typ A 0.920 T typ B 0.998	I: Yes (kappa 0.815) E: No	Classification accuracy based on deep learning
Doerr et al. (2022)	USA	Retro	111 (n/d)	DL, R–CNN	Trauma TL injuries (compression, burst fractures, translation or rotation)	CT	Acc: Compression f. 0.814 Burst fracture 0.686 Translation or rotation 0.801	I: No E: No	RCNN is accurate in analyzing CT scans
Yeh et al. (2022)	Taiwan	Retro	190 (n/d)	DL	Benign or malignant spinal fractures	MRI	Se: 0.94 Sp: 0.91	I: Yes E: Yes	ResNet50 DL model may provide information to assist less experienced clinicians in the diagnosis of VF on MRI.
Rosenberg et al. (2022)	Italy	Retro	151 (n/d)	DL, R–CNN	TL fractures	RTG, CT, MRI	Acc: 0.88 (ResNet) 0.86 (VGG16) Se: 0.91 (ResNet); 0.90 (VGG16) Sp: 0.89 (ResNet). 0.83 (VGG16)	I: Yes E: No	DL models can be adapted to accurately detect
Iyer et al. (2023)	Australia	Retro	308 (n/d)	DL	TL VFs	CT	Acc: 0.8595 Se: 0.881 Sp: 0.842	I: No E: No	Analysis of CT images with DRL and IL methods for more precise localization of the pathological process
Jo et al. (2023)	South Korea	Retro	400 (n/d)	DL (InResNetV2)	PLC, TL fractures	MRI	Se: 0.820 Sp: 0.940 AUC: 0.916	I: Yes E: Yes	The DL algorithm detected PLC injury in patients with TL fracture with high diagnostic efficiency, comparable to that of an experienced radiologist.
Li et al. (2023)	China	Retro	57 (21/ 36)	DL	Occult vertebral fractures	CT	Acc: 0.846 Se: 0.846 Sp: 0.846 AUC: 0.692; 0.775; 0.680; (SVM; LR;	I: Yes E: Yes	CT radiomics, combined with machine learning, allows for the identification of OVFs not readily appreciable on CT.

(continued on next page)

Table 3 (*continued*)

Author (Year)	Country	Study design	Number of patients (Male/ Female)	ML/DL algorithm or Method	Pathology or focus	Diagnostic modality	Diagnostic test analysis related variables	Validation (internal and external)	Conclusions
							Bayes) – axial 0.805.0.882 I 0.834 (SVM. LR Bayes) - sagittal		
Ono et al. (2023)	Japan	Retro	552 (n/d)	DL	Osteoporotic lumbar vertebral fractures (OLVF)	RTG	Acc: 0.894; Se: 0.836; Sp: 0.920;	I: Yes E: Yes	The proposed CNN-based method demonstrated high performance in determining the presence of OLVF and classifying old or fresh OLVF on radiography
Ryu et al. (2023)	South Korea	Retro	198 (n/d)	DL	Lumbar VCFs - vertebral compression fractures	RTG (LSLR)	Acc: 0.929 Se: 0.944 Sp: 0.917	I: Yes E: Yes	High accuracy of the DL model for VCF detection with the help of LSLR
Germann et al. (2023)	Switzerland	Retro	200 (n/d)	DCNN	Lumbar VFs	MRI	Acc: 0.964 Se: 0.941 Sp: 0.969	I: Yes E: Yes	DCNN can achieve high diagnostic performance in vertebral body measurements and insufficiency fracture detection on heterogenous lumbar spine MRI
Cheng et al. (2022)	China	Retro	390 (n/d)	ML	The difference between compression and burst fractures	RTG	Acc: 0.99 normal vertebral bodies 0.74 compression fractures 0.94 burst fractures	n/a	Assistance in the rapid detection of spinal fractures to emergency medicine physicians
Hong et al. (2023)	South Korea	Retro	9276 (3171/ 6105)	DL	VF and osteoporosis	RTG	Acc: 0.91 (0.92 – external) Se: 0.76 (0.75 – external) Sp: 0.94 (0.97 external) FP: 0.74 (0.82 external) FN: 0.95 (0.96 external)	I: Yes E: Yes	Spine radiography from DCNN models detected prevalent vertebral fractures and showed better detection performance than clinical models.
Zhang et al. (2023)	China	Retro	285 (119/ 166)	DL	VFs	CT	Acc: 0.9793 Se: 0.9523 Sp: 0.9835	n/a	Multilevel AO system automatically classifies acute vertebral body fractures in TL on CT images with high

Legend: DL, deep learning; ML, machine learning; TL, thoracolumbar; OVF, occult vertebral fractures; PKP, percutaneous kyphoplasty; OVCF, osteoporotic vertebral compression fracture; NVCF, new vertebral compression fracture; PARs, Plain abdominal frontal radiographs; DRL, deep reinforcement learning; IL, imitation learning; DCNN, deep convolutional neural network; VF, vertebral fracture; PTLR, plain thoracolumbar radiography; PLC, posterior ligamentous complex; AI, artificial intelligence, OLVF, osteoporotic lumbar vertebral fractures; CNN, conolutional neural networks; VCF, vertebral compression fracture; LSLR, lumbal spine lateral radiographs; R–CNN, region based-convolutional neural network; PLC, posterior ligamentous complex; TLICS, Thoracolumbar Injury Classification and Severity Score; AO, AO Spine thoraculumbar spine injury classification system.

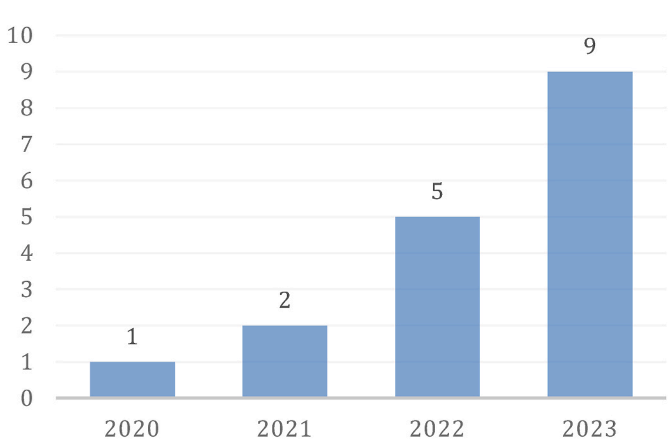

Fig. 2. Temporal distribution of included studies.

algorithms, five studies (29.41%) used CT, four (23.53%) used MRI, and two (11.76%) used a combination of diagnostic modalities. Of the total, 12 studies (70.85%) performed internal validation of results, while nine studies (52.94%) reported external validation.

3.2. Risk of bias assessment

The assessment of risk of bias and applicability is shown in Fig. 4a and b. The assessment shows that only one study had an unclear risk of bias, while the other studies had a low risk of bias (Fig. 4c). The same relationship was observed for the applicability of the studies (Fig. 4d). The Deek's funnel plot test for diagnostic odds ratios was performed (Fig. 4e) and yielded a non-statistically significant result with a t-statistic of −1.53 and a p-value of 0.1369.

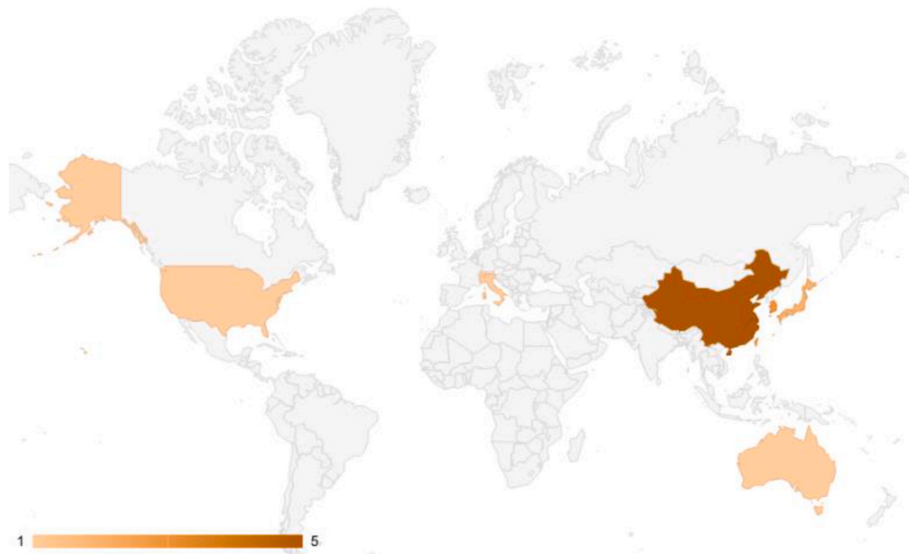

Fig. 3. Geographical distribution of included studies.

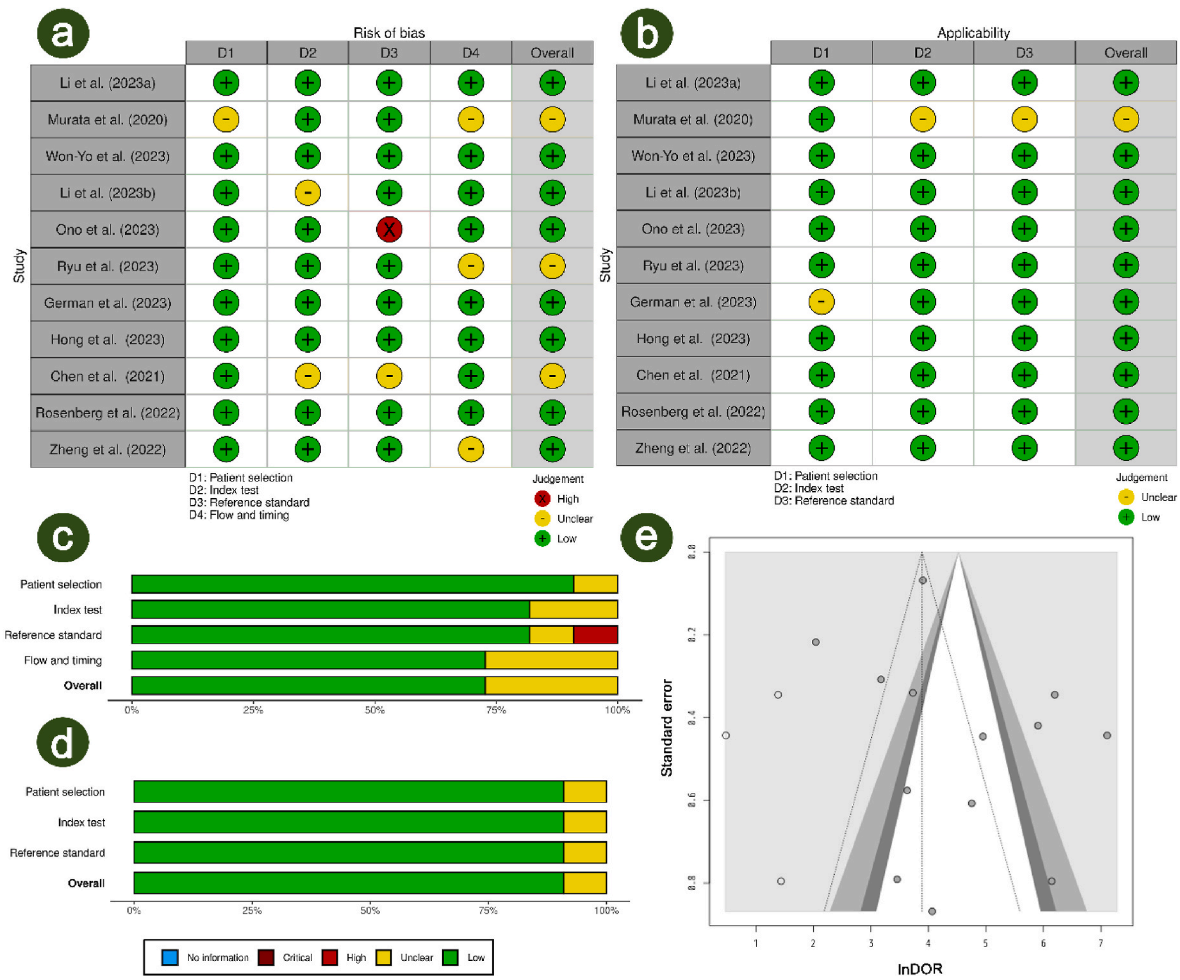

Fig. 4. Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2): a) analysis of bias risk; b) analysis of applicability; c) summary of bias risk assessment; d) summary of applicability analysis; e) Deek's funnel plot depicting publication bias.

3.3. Results of meta-analysis

The total number of studies that met the criteria for meta-analysis was 11 (64.1%), with 94.1% (N = 13,673) of the sample included in the study. The number of true positives (TP) accounted to 4865 (35.58%), false positives (FP) to 870 (6.36%), false negatives (FN) to

475 (3.47%) and true negatives (TN) to 7843 (57.36%).

The meta-analysis showed a sensitivity of 0.91 (95% CI = 0.86–0.95) (Fig. 5). The study by Zhang et al. (2023) had the highest sensitivity with an estimated value of 0.99 (95% CI = 0.98–1.00), while Chen et al. (2021) reported the lowest sensitivity of 0.70 (95% CI = 0.63–0.76). The specificity was an estimated value of 0.90 (95% CI = 0.86–0.93). Further analysis resulted in a DOR value of 94.603 (95% CI = 49.215–181.85). The LR+ was calculated to be 9.36 (95% CI = 6.575–13.325). In contrast, the LR-was low at 0.099 (95% CI = 0.061–0.16). The FPR had a value of 0.097 (95% CI = 0.068–0.137).

Fig. 6 shows the HSROC analysis of the studies included in the meta-analysis. The logit-transformed sensitivity results in a value of 2.327, which means a high probability of accurately identifying true positive cases. At the same time, the logit-transformed specificity is estimated to be 2.199, indicating a high probability of correctly identifying true negative cases. The estimated variances for the logit-transformed sensitivity and specificity are 0.070 and 0.037, respectively, illustrating the extent of variability of these parameters between studies. In addition, the estimated covariance between the logit-transformed sensitivity and specificity is 0.021, indicating the DOR between the two measures. These results emphasize the precision and reliability of the diagnostic test, and accounting for variances and covariances provides valuable insight into the uniformity and potential heterogeneity of performance between studies.

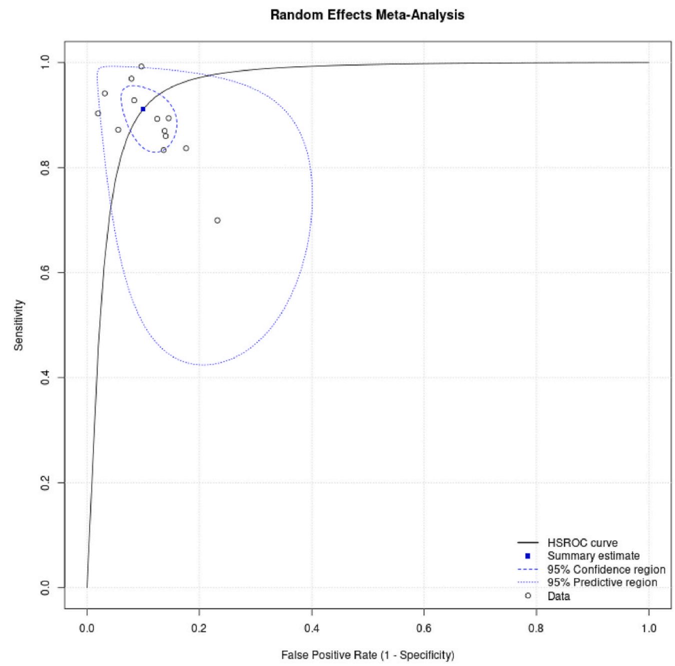

Fig. 6. Hierarchical summary receiver operating characteristic (HSROC) curve of included studies.

4. Discussion

The total number of studies included in our systematic review was 17. These studies investigated the applicability of ML and DL in the diagnostic evaluation of TLI-related VF. The chronological onset of increased research efforts in this area underscores the transformative path that these computational methods have taken.

The analysis of the included studies revealed a notable leadership position of China in the field of AI applications for the diagnosis of TL spinal injuries and fractures. Taiwan ranked second, followed by South Korea. The heightened incidence of TLI has led to an increased interest in the publication of articles in this field, a phenomenon that is particularly emphasized in China (Li et al., 2019). In addition, there is a noticeable trend towards the escalating utilization of modern neurosurgical technologies in this country (Dewan et al., 2019; Zhou et al., 2023). China's remarkable strides in research capacity and scientific activities is reflected in the significant increase in investment in research and development, accompanied by a notable increase in the number of research personnel and publications (Marginson, 2022)

The gold standard for the diagnosis of fractures and injuries of the TL spine is CT or MRI, particularly in the context of human interpretation. The first diagnostic step includes sagittal and anteroposterior radiographs, which offer the advantage of a lower radiation dose compared to CT or MRI (Rutsch et al., 2023). However, this method has its limitations, particularly in the detection of VF (Li et al., 2021). This limitation is particularly pronounced in OVF, where the bone marrow edema crucial for diagnosis remains invisible on X-ray images but is visible on MRI (Li et al., 2023; Ono et al., 2023). The increasing number of patients with pain in the TL spine region has led to an increase in referrals for further diagnostics. In 48% of cases, fractures are most localized between vertebrae Th12-L2 (Rosenberg et al., 2022). Studies indicate a higher prevalence in patients over 50 years of age and in postmenopausal women (Iyer et al., 2023; Ryu et al., 2023). Furthermore, most studies on this topic were conducted between 2022 and 2023, with a decline observed in 2020. The possible cause of this decline is attributed to the global COVID-19 pandemic (Kuo et al., 2023). AI is expected

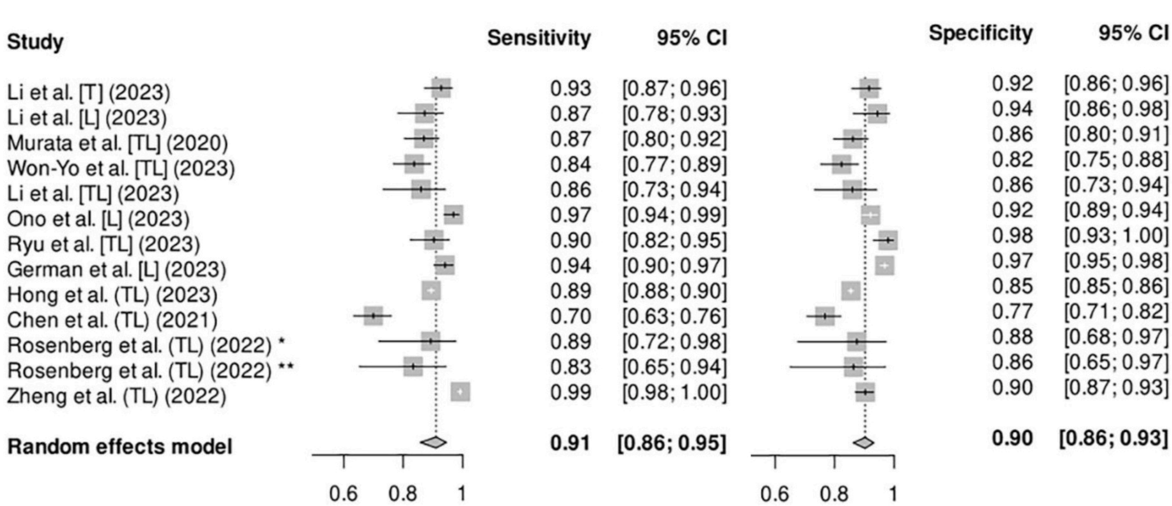

Fig. 5. Estimated values of sensitivity and specificity for included studies in meta-analysis
Legend: T – thoracal; L – lumbar; TL – thoracolumbar; * and ** - different DL models used od same cohort.

to be particularly beneficial for doctors, especially in the emergency room and primary care. Studies have shown that physicians, including radiologists and neurosurgeons, cannot always recognize VF (Murata et al., 2020). The application of AI in this context promises to improve diagnostic accuracy and patient outcomes (Krishnan et al., 2023).

Pizones et al. (2011) conducted a study on a cohort of 30 patients (15 men and 15 women) with traumatic injuries of TL spine. The primary objective was to demonstrate the role and efficacy of MRI in the diagnosis of spinal injuries. While X-ray and CT diagnosed 41 VF, MRI detected 50 VF and nine contusions. This raises questions about the diagnostic efficacy of radiographs and CT in VF or possible errors in interpretation by radiologists. The study conducted by Levi et al. (2006) investigated unrecognized VF in trauma centers, focusing on 24 patients who experienced neurological deterioration due to unrecognized conditions. Five patients developed radiculopathies, 16 suffered spinal cord injuries and three died. In the study, these outcomes were attributed to incorrect measurements, insufficiently specific diagnoses or poor-quality X-rays. Consistent with these findings, a study of 585 patients showed delays in making the correct diagnosis, with the longest delay being 115 days. Seven patients were X-rayed in the medical emergency department and the doctors did not recognize the VF (Levi et al., 2006).

In addition, the identification of VF was delayed in 22 cases (Aso-Escario et al., 2019). Physicians face challenges in diagnosing spinal injuries quickly and accurately, particularly in recognizing TL fractures and differentiating between burst and compression fractures from radiographs. These diagnostic difficulties have a significant impact on patient prognosis. In response, AI has emerged as a promising avenue over the past decade. Current AI methods, while still in the early stages, involve analyzing X-ray, CT or MRI data to make a definitive diagnosis, determine the appropriate treatment approach (conservative or surgical) and assess the risk of potential disability to the patient (Cheng et al., 2022; Rosenberg et al., 2022). AI in orthopedics is already making initial progress in overcoming specific orthopedic challenges, for example image recognition, preoperative risk assessment, clinical decision-making and the analysis of large data sets (Myers et al., 2020). ML helps in predicting patient-specific postoperative complications, assessing patterns of injury risk and clinical decision making (Han and Tian, 2019).

The use of ML and DL in diagnostics has been shown to be beneficial in various pathologic conditions, e.g., vertebral compression fractures (VCF) (Ryu et al., 2023), occult vertebral fractures (OVF) (Li et al., 2023), osteoporotic lumbar vertebral fractures (OLVF) (Ono et al., 2023), posterior ligament complex (PLC) injuries (Jo et al., 2023) and secondary vertebral fractures (VF) caused by pre-existing osteoporosis, neoplasms or traumatic injuries (Li et al., 2021). In the field of neurosurgery, a systematic literature review by Danilov et al. (2021) notes the broad applicability of AI, with approximately 41% of studies focusing on neuro-oncology and 19% on functional neurosurgery, including epilepsy surgery. Research on the use of AI technologies in neurosurgery is mainly focused on neuro-oncology, functional, vascular and spinal neurosurgery, and traumatic brain injury (Danilov et al., 2021). Among the predominant algorithms, Deep Convolutional Neural Network (DCNN) and Region-based Convolutional Neural Networks (RCNN) stand out. In the study by Wu-Gen Li et al. (2023), which focused on CT diagnosis in conjunction with ML for OVF, three algorithmic models were presented: Support Vector Machine (SVM), logistic regression (LR) and Bayesian model. Logistic regression (LR) was found to be the best performing model in the sagittal plane of CT with an accuracy of 0.846, a sensitivity of 0.846 and a specificity of 0.846. In contrast, the SVM model performed best in the axillary plane of CT with an accuracy of 0.731, a sensitivity of 0.462 and a specificity of 1.000.

AI models based on DL and ML show significant potential for diagnostic procedures related to VF due to traumatic brain injury. The included studies show that the DCNN is an algorithm with high diagnostic accuracy and precision in the detection of TL spinal fractures. Murata et al. (2020) claim that human knowledge, experience and intelligence are not interchangeable with AI. They report higher accuracy, sensitivity and specificity for spinal surgeons (98.4%, 96% and 100%) compared to DCNN (86%, 84.7% and 87.3%). In contrast, Jo et al. (2023) finds that DL and DCNN perform equally as well as radiologists in terms of diagnostic performance, with similar results in terms of accuracy, sensitivity and specificity. Germann et al. (2023) reported no significant differences between radiologists and DL/DCNN in the interpretation of findings. However, both studies used MRI, which is known for its high accuracy in the diagnosis of certain pathologies.

A robust classification system is essential for effective communication, treatment guidance and accurate prognosis in spinal surgery (Bajamal et al., 2021). The Thoracolumbar Injury Classification and Severity Score (TLICS) has become well known due to its wide acceptance (Gamanagatti et al., 2015; Nataraj et al., 2018). Despite the introduction of a new AO (*Arbeitsgemeinschaft für Osteosynthesefragen*) spine classification that incorporates elements of Magerl/AO and TLICS, further standardization and empirical validation is needed (Joaquim and Patel, 2013). Radiologic modalities such as X-ray, CT or MRI play a central role in diagnosis, although TLICS remains the preferred tool for the assessment of thoracic and lumbar spine injuries (Reinhold et al., 2013). The WFNS Spine Committee endorses the validity and applicability of both the AO and TLICS classifications in clinical practice for traumatic thoracolumbar fractures (Bajamal et al., 2021). New research suggests potential advantages of the complicated AO classification, particularly in the absence of CT/MRI scans (Park et al., 2016). MRI is recommended primarily for its accuracy in visualizing the disco ligamentous complex and detecting associated pathology in spinal trauma (Bajamal et al., 2021).

When investigating the precision of ML and DL algorithms in TLI, it is important to consider several radiologic abnormalities. Kyphotic lesions characterized by a reduction in anterior vertebral height of more than 50% play a crucial role in surgical planning and evaluation of the results of the procedure. Proper assessment of the posterior ligamentous complex (PLC) is essential, as an inadequate PLC directly affects the extent of the fracture. MRI is recommended when the interspinous gap widens by 20% or more to rule out unhealthy PLC (Jo et al., 2023). In the evaluation of upper spinal cord fractures with DL algorithms, the studies by Chen et al. (13) and Zhang et al. (2023) found varying degrees of sensitivity, with CT showing higher sensitivity than radiography with DL. In addition, the AO system was suggested to be superior to the DL method in the clinical setting. Specificity varied between studies, with Chen et al. (2021) reporting the lowest specificity and Li et al. (2021) reporting the highest specificity, suggesting that X-ray with DL performs better than CT without DL or MRI without DL in detecting vertebral fractures, possibly due to the overall advantages of CT/MRI over X-ray.

The estimated value for the specificity of the ML and DL algorithms in this meta-analysis is 0.90 (95% CI = 0.86–0.93), indicating high accuracy. A value of 0.91 (95% CI = 0.86–0.95) was determined for the sensitivity. These calculations represent the first synthesis of studies addressing the specificity and sensitivity of ML and DL algorithms in identifying VFs caused by TLIs. Yang et al. (2020) found a sensitivity of 0.87 (95% CI: 0.78–0.93) in different orthopedic fractures using ML and DL models for identification, which closely agrees with our results. In the same study, the specificity was 0.91 (95% CI: 0.85–0.95), which is consistent with the results of our assessment. When analyzing hip fractures, a slightly lower sensitivity of 0.844 (95% CI: 0.791–0.885) was found (Rahim et al., 2023).

The field of ML and DL research has a high publication rate, and new articles appear frequently. Consequently, the literature landscape may have evolved considerably at the time of publication and may contain relevant articles not yet included in this review. The AI models developed to analyze plain lateral radiographs had limitations in terms of their scope and diagnostic capabilities, particularly in detecting VF in TLI (Ryu et al., 2023). The models only recognized eight vertebrae and had difficulty identifying fractures above T9. Furthermore, their development was aimed at detecting VF without investigating underlying causes such as neoplasms, osteomyelitis or multiple myeloma. Clinical evaluation and confirmation by other imaging modalities remain paramount for the accurate diagnosis of pathologic VF in TLI. In addition, the models failed to distinguish between acute and subacute stages of VF, degenerative spondylolisthesis and disk degeneration, which may lead to misinterpretation. These shortcomings highlight the importance of using the models with critical awareness and ensuring comprehensive clinical assessments for a definitive diagnosis (Li et al., 2021). In addition, the performance of ML and DL models for detecting VF may vary in different age groups due to the heterogeneity of training data in some studies that include fractures in pediatric and geriatric populations. This variability in baseline data suggests potential limitations in generalizing the performance of the models to younger or older patients (Iyer et al., 2023). Other limitations cited include the exclusion of old fractures and the lack of assessment of the functional prognosis of fractures (Murata et al., 2020).

ML algorithms are not limited to linear data; they can also be used to handle non-linear relationships between variables and outcomes. This is particularly useful where risk factors and outcomes observed in patients may exhibit complex patterns. Furthermore, ML algorithms can analyze large amounts of data and identify which factors are most relevant to the outcome. Also, ML algorithms can achieve higher accuracy than conventional models and even handle missing data quite efficiently (Doerr et al., 2022). DL models for VF detection often comprise millions of parameters and are primarily used for data fitting. However, this complexity can lead to overfitting, hindering the models' ability to classify unseen data. Reducing the number of parameters through techniques like model compression and architectural optimization can foster improved generalization and robustness. Training DL models for VF detection necessitates the inclusion of images with diverse characteristics to enhance generalizability. Future studies should include more CT and MRI images. It is necessary to conduct training using images with various characteristics, performance, and applicability (Begagić et al., 2023b). Evaluating ML and DL performance and applicability in primary care settings, beyond secondary and tertiary care, is crucial for real-world implementation, especially in Low- and Middle-Income Countries (Begagić et al., 2023a).

Like any study, this one has its limitations, which are primarily due to the relatively small number of studies included. In addition, standardization of the reporting procedure for accuracy scores would be essential in the near future to enable the inclusion of studies in a meta-analysis to obtain a more comprehensive and in-depth overview of the applicability of ML and DL in the review of VF caused by TLI.

5. Conclusion

In our systematic review of diagnostic approaches for thoracolumbar spine fractures, deep learning was predominantly used, while machine learning was only explored to a limited extent. The study showed consistent specificity and sensitivity estimated in the meta-analysis, highlighting the robustness of the diagnostic test. However, the broader context of ML applications in TLIs suggests that there is a critical need for standardization of methods. The report highlights the importance of rigorous modeling techniques, clear criteria for model selection, and internal and external validation to ensure the reliability of machine learning models for clinical integration. Future research should address the identified limitations, expand modalities and prioritize robust methods to strengthen the evidence base for informed decision making between clinician and patient and ultimately improve patient care and clinical outcomes.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aso-Escario, J., Sebastián, C., Aso-Vizán, A., Martínez-Quiñones, J.V., Consolini, F., Arregui, R., 2019. Delay in diagnosis of thoracolumbar fractures. Orthop. Rev. 11 (2), 7774. https://doi.org/10.4081/or.2019.7774.
Bajamal, A.H., Permana, K.R., Faris, M., Zileli, M., Peev, N.A., 2021. Classification and radiological diagnosis of thoracolumbar spine fractures: WFNS spine Committee recommendations. Neurospine 18 (4), 656–666. https://doi.org/10.14245/ns.2142650.325.
Barbiellini Amidei, C., Salmaso, L., Bellio, S., Saia, M., 2022. Epidemiology of traumatic spinal cord injury: a large population-based study. Spinal Cord 60 (9), 812–819. https://doi.org/10.1038/s41393-022-00795-w.
Bečulić, H., Begagić, E., Skomorac, R., Mašović, A., Selimović, E., Pojskić, M., 2024. ChatGPT's contributions to the evolution of neurosurgical practice and education: a systematic review of benefits, concerns and limitations. Med. Glas. 21 (1) https://doi.org/10.17392/1661-23.
Begagić, E., Bečulić, H., Skomorac, R., Pojskić, M., 2023a. Accessible spinal surgery: transformation through the implementation of exoscopes as substitutes for conventional microsurgery in low- and middle-income settings. Cureus 15 (9), e45350. https://doi.org/10.7759/cureus.45350.
Begagić, E., Pugonja, R., Bečulić, H., Selimović, E., Skomorac, R., Saß, B., Pojskić, M., 2023b. The new era of spinal surgery: exploring the use of exoscopes as a viable alternative to operative microscopes-A systematic review and meta-analysis. World Neurosurg. https://doi.org/10.1016/j.wneu.2023.11.026.
Chen, H.-Y., Hsu, B.W.-Y., Yin, Y.-K., Lin, F.-H., Yang, T.-H., Yang, R.-S., Lee, C.-K., Tseng, V.S., 2021. Application of deep learning algorithm to detect and visualize vertebral fractures on plain frontal radiographs. PLoS One 16 (1), e0245992. https://doi.org/10.1371/journal.pone.0245992.
Chen, X., Liu, Y., 2022. A classification method for thoracolumbar vertebral fractures due to basketball sports injury based on deep learning. Comput. Math. Methods Med. 2022, 8747487 https://doi.org/10.1155/2022/8747487.
Cheng, L.-W., Chou, H.-H., Huang, K.-Y., Hsieh, C.-C., Chu, P.-L., Hsieh, S.-Y., 2022. Automated diagnosis of vertebral fractures using radiographs and machine learning intelligent computing theories and application. In: 18th International Conference, ICIC 2022, Xi'an, China, August 7–11, 2022, Proceedings, Part I, Xi'an, China. https://doi.org/10.1007/978-3-031-13870-6_59.
Dai, L.Y., 2012. Principles of management of thoracolumbar fractures. Orthop. Surg. 4 (2), 67–70. https://doi.org/10.1111/j.1757-7861.2012.00174.x.
Danilov, G.V., Shifrin, M.A., Kotik, K.V., Ishankulov, T.A., Orlov, Y.N., Kulikov, A.S., Potapov, A.A., 2021. Artificial intelligence technologies in neurosurgery: a systematic literature review using topic modeling. Part II: research objectives and perspectives. Sovrem Tekhnologii Med 12 (6), 111–118. https://doi.org/10.17691/stm2020.12.6.12.
den Ouden, L.P., Smits, A.J., Stadhouder, A., Feller, R., Deunk, J., Bloemers, F.W., 2019. Epidemiology of spinal fractures in a level one trauma center in The Netherlands: a 10 Years review. Spine 44 (10), 732–739. https://doi.org/10.1097/BRS.0000000000002923 (Phila Pa 1976).
Dewan, M.C., Rattani, A., Fieggen, G., Arraez, M.A., Servadei, F., Boop, F.A., Johnson, W. D., Warf, B.C., Park, K.B., 2019. Global neurosurgery: the current capacity and deficit in the provision of essential neurosurgical care. Executive summary of the global neurosurgery initiative at the program in global surgery and social change. J. Neurosurg. 130 (4), 1055–1064. https://doi.org/10.3171/2017.11.JNS171500.
Doerr, S.A., Weber-Levine, C., Hersh, A.M., Awosika, T., Judy, B., Jin, Y., Raj, D., Liu, A., Lubelski, D., Jones, C.K., Sair, H.I., Theodore, N., 2022. Automated prediction of the Thoracolumbar Injury Classification and Severity Score from CT using a novel deep learning algorithm. Neurosurg. Focus 52 (4), E5. https://doi.org/10.3171/2022.1.FOCUS21745.
Fernández-de Thomas, R.J., De Jesus, O., 2023. Thoracolumbar spine fracture. In: StatPearls. StatPearls Publishing.
Gamanagatti, S., Rathinam, D., Rangarajan, K., Kumar, A., Farooque, K., Sharma, V., 2015. Imaging evaluation of traumatic thoracolumbar spine injuries: radiological review. World J. Radiol. 7 (9), 253–265. https://doi.org/10.4329/wjr.v7.i9.253.
Germann, C., Meyer, A.N., Staib, M., Sutter, R., Fritz, B., 2023. Performance of a deep convolutional neural network for MRI-based vertebral body measurements and insufficiency fracture detection. Eur. Radiol. 33 (5), 3188–3199. https://doi.org/10.1007/s00330-022-09354-6.
Han, X.G., Tian, W., 2019. Artificial intelligence in orthopedic surgery: current state and future perspective. Chin. Med. J. 132 (21), 2521–2523. https://doi.org/10.1097/cm9.0000000000000479.
Hong, N., Cho, S.W., Shin, S., Lee, S., Jang, S.A., Roh, S., Lee, Y.H., Rhee, Y., Cummings, S.R., Kim, H., Kim, K.M., 2023. Deep-learning-based detection of vertebral fracture and osteoporosis using lateral spine X-ray radiography. J. Bone Miner. Res. 38 (6), 887–895. https://doi.org/10.1002/jbmr.4814.
Iyer, S., Blair, A., White, C., Dawes, L., Moses, D., Sowmya, A., 2023. Vertebral compression fracture detection using imitation learning, patch based convolutional neural networks and majority voting. Inform. Med. Unlocked 38, 101238. https://doi.org/10.1016/j.imu.2023.101238.
Jo, S.W., Khil, E.K., Lee, K.Y., Choi, I., Yoon, Y.S., Cha, J.G., Lee, J.H., Kim, H., Lee, S.Y., 2023. Deep learning system for automated detection of posterior ligamentous complex injury in patients with thoracolumbar fracture on MRI. Sci. Rep. 13 (1), 19017 https://doi.org/10.1038/s41598-023-46208-7.

Joaquim, A.F., Patel, A.A., 2013. Thoracolumbar spine trauma: evaluation and surgical decision-making. J. Craniovertebral Junction Spine 4 (1), 3–9. https://doi.org/10.4103/0974-8237.121616.
Karabacak, M., Margetis, K., 2023. Precision medicine for traumatic cervical spinal cord injuries: accessible and interpretable machine learning models to predict individualized in-hospital outcomes. Spine J. 23 (12), 1750–1763. https://doi.org/10.1016/j.spinee.2023.08.009.
Krishnan, G., Singh, S., Pathania, M., Gosavi, S., Abhishek, S., Parchani, A., Dhar, M., 2023. Artificial intelligence in clinical medicine: catalyzing a sustainable global healthcare paradigm. Front Artif Intell 6, 1227091. https://doi.org/10.3389/frai.2023.1227091.
Kumar, R., Lim, J., Mekary, R.A., Rattani, A., Dewan, M.C., Sharif, S.Y., Osorio-Fonseca, E., Park, K.B., 2018. Traumatic spinal injury: global epidemiology and worldwide volume. World Neurosurg 113, e345–e363. https://doi.org/10.1016/j.wneu.2018.02.033.
Kuo, C.C., Aguirre, A.O., Kassay, A., Donnelly, B.M., Bakr, H., Aly, M., Ezzat, A.A.M., Soliman, M.A.R., 2023. A look at the global impact of COVID-19 pandemic on neurosurgical services and residency training. Scientific African 19, e01504. https://doi.org/10.1016/j.sciaf.2022.e01504.
Lenehan, B., Boran, S., Street, J., Higgins, T., McCormack, D., Poynton, A.R., 2009. Demographics of acute admissions to a national spinal injuries unit. Eur. Spine J. 18 (7), 938–942. https://doi.org/10.1007/s00586-009-0923-y.
Levi, A.D., Hurlbert, R.J., Anderson, P., Fehlings, M., Rampersaud, R., Massicotte, E.M., France, J.C., Le Huec, J.C., Hedlund, R., Arnold, P., 2006. Neurologic deterioration secondary to unrecognized spinal instability following trauma–A multicenter study. Spine 31 (4). https://journals.lww.com/spinejournal/fulltext/2006/02150/neurologic_deterioration_secondary_to_unrecognized.14.aspx.
Li, W.-G., Zeng, R., Lu, Y., Li, W.-X., Wang, T.-T., Lin, H., Peng, Y., Gong, L.-G., 2023. The value of radiomics-based CT combined with machine learning in the diagnosis of occult vertebral fractures. BMC Muscoskel. Disord. 24 (1), 819. https://doi.org/10.1186/s12891-023-06939-0.
Li, Y., Zheng, S., Wu, Y., Liu, X., Dang, G., Sun, Y., Chen, Z., Wang, J., Li, J., Liu, Z., 2019. Trends of surgical treatment for spinal degenerative disease in China: a cohort of 37,897 inpatients from 2003 to 2016. Clin. Interv. Aging 14, 361–366. https://doi.org/10.2147/cia.S191449.
Li, Y.-C., Chen, H.-H., Horng-Shing Lu, H., Hondar Wu, H.-T., Chang, M.-C., Chou, P.-H., 2021. Can a deep-learning model for the automated detection of vertebral fractures approach the performance level of human subspecialists? Clin. Orthop. Relat. Res. 479 (7). https://journals.lww.com/clinorthop/fulltext/2021/07000/can_a_deep_learning_model_for_the_automated.32.aspx.
Ma, Y., Lu, Q., Yuan, F., Chen, H., 2023. Comparison of the effectiveness of different machine learning algorithms in predicting new fractures after PKP for osteoporotic vertebral compression fractures. J. Orthop. Surg. Res. 18 (1), 62. https://doi.org/10.1186/s13018-023-03551-9.
Marginson, S., 2022. 'All things are in flux': China in global science. High Educ. 83 (4), 881–910. https://doi.org/10.1007/s10734-021-00712-9.
Mizutani, S., Zhou, Y., Tian, Y.S., Takagi, T., Ohkubo, T., Hattori, S., 2023. DTAmetasa: an R shiny application for meta-analysis of diagnostic test accuracy and sensitivity analysis of publication bias. Res. Synth. Methods 14 (6), 916–925. https://doi.org/10.1002/jrsm.1666.
Murata, K., Endo, K., Aihara, T., Suzuki, H., Sawaji, Y., Matsuoka, Y., Nishimura, H., Takamatsu, T., Konishi, T., Maekawa, A., Yamauchi, H., Kanazawa, K., Endo, H., Tsuji, H., Inoue, S., Fukushima, N., Kikuchi, H., Sato, H., Yamamoto, K., 2020. Artificial intelligence for the detection of vertebral fractures on plain spinal radiography. Sci. Rep. 10 (1), 20031 https://doi.org/10.1038/s41598-020-76866-w.
Myers, T.G., Ramkumar, P.N., Ricciardi, B.F., Urish, K.L., Kipper, J., Ketonis, C., 2020. Artificial intelligence and orthopaedics: an introduction for clinicians. J Bone Joint Surg Am 102 (9), 830–840. https://doi.org/10.2106/jbjs.19.01128.
Nataraj, A., Jack, A.S., Ihsanullah, I., Nomani, S., Kortbeek, F., Fox, R., 2018. Outcomes in thoracolumbar burst fractures with a thoracolumbar injury classification score (TLICS) of 4 treated with surgery versus initial conservative management. Clin Spine Surg 31 (6), E317–e321. https://doi.org/10.1097/bsd.0000000000000656.
Nyaga, V.N., Arbyn, M., 2022. Metadta: a Stata command for meta-analysis and meta-regression of diagnostic test accuracy data – a tutorial. Arch. Publ. Health 80 (1), 95. https://doi.org/10.1186/s13690-021-00747-5.
Ono, Y., Suzuki, N., Sakano, R., Kikuchi, Y., Kimura, T., Sutherland, K., Kamishima, T., 2023. A deep learning-based model for classifying osteoporotic lumbar vertebral fractures on radiographs: a retrospective model development and validation study. Journal of Imaging 9 (9), 187. https://www.mdpi.com/2313-433X/9/9/187.
Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., et al., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Br. Med. J. 372, n71. https://doi.org/10.1136/bmj.n71.
Park, H.J., Lee, S.Y., Park, N.H., Shin, H.G., Chung, E. C.Rho, et al., 2016. Modified thoracolumbar injury classification and severity score (TLICS) and its clinical usefulness. Acta Radiol. 57 (1), 74–81. https://doi.org/10.1177/0284185115580487.
Pizones, J., Izquierdo, E., Álvarez, P., Sánchez-Mariscal, F., Zúñiga, L., Chimeno, P., Benza, E., Castillo, E., 2011. Impact of magnetic resonance imaging on decision making for thoracolumbar traumatic fracture diagnosis and treatment. Eur. Spine J. 20 (3), 390. https://doi.org/10.1007/s00586-011-1913-4.
Plana, M.N., Arevalo-Rodriguez, I., Fernández-García, S., Soto, J., Fabregate, M., Pérez, T., Roqué, M., Zamora, J., 2022. Meta-DiSc 2.0: a web application for meta-analysis of diagnostic test accuracy data. BMC Med. Res. Methodol. 22 (1), 306. https://doi.org/10.1186/s12874-022-01788-2.
Postma, I.L.E., Oner, F.C., Bijlsma, T.S., Heetveld, M.J., Goslings, J.C., Bloemers, F.W., 2015. Spinal injuries in an airplane crash: a description of incidence, morphology, and injury mechanism. Spine 40 (8). https://journals.lww.com/spinejournal/fulltext/2015/04150/spinal_injuries_in_an_airplane_crash_a.9.aspx.
Rahim, F., Zaki Zadeh, A., Javanmardi, P., Emmanuel Komolafe, T., Khalafi, M., Arjomandi, A., Ghofrani, H.A., Shirbandi, K., 2023. Machine learning algorithms for diagnosis of hip bone osteoporosis: a systematic review and meta-analysis study. Biomed. Eng. Online 22 (1), 68. https://doi.org/10.1186/s12938-023-01132-9.
Reinhold, M., Audigé, L., Schnake, K.J., Bellabarba, C., Dai, L.Y., Oner, F.C., 2013. AO spine injury classification system: a revision proposal for the thoracic and lumbar spine. Eur. Spine J. 22 (10), 2184–2201. https://doi.org/10.1007/s00586-013-2738-0.
Rosenberg, G.S., Cina, A., Schiró, G.R., Giorgi, P.D., Gueorguiev, B., Alini, M., Varga, P., Galbusera, F., Gallazzi, E., 2022. Artificial intelligence accurately detects traumatic thoracolumbar fractures on sagittal radiographs. Medicina 58 (8), 998. https://www.mdpi.com/1648-9144/58/8/998.
Rosner, B., 2015. Fundamentals of Biostatistics. Cengage learning.
Rutsch, N., Amrein, P., Exadaktylos, A.K., Benneker, L.M., Schmaranzer, F., Müller, M., Albers, C.E., Bigdon, S.F., 2023. Cervical spine trauma - evaluating the diagnostic power of CT, MRI, X-Ray and LODOX. Injury 54 (7), 110771. https://doi.org/10.1016/j.injury.2023.05.003.
Ryu, S.M., Lee, S., Jang, M., Koh, J.-M., Bae, S.J., Jegal, S.G., Shin, K., Kim, N., 2023. Diagnosis of osteoporotic vertebral compression fractures and fracture level detection using multitask learning with U-Net in lumbar spine lateral radiographs. Comput. Struct. Biotechnol. J. 21, 3452–3458. https://doi.org/10.1016/j.csbj.2023.06.017.
Sarker, I.H., 2021. Machine learning: algorithms, real-world applications and research directions. SN Comput Sci 2 (3), 160. https://doi.org/10.1007/s42979-021-00592-x.
Sharma, S., 2023. Artificial intelligence for fracture diagnosis in orthopedic X-rays: current developments and future potential. Sicot j 9, 21. https://doi.org/10.1051/sicotj/2023018.
Singleton, J.M., Hefner, M., 2023. Spinal cord compression. In: StatPearls. StatPearls Publishing Copyright © 2023. StatPearls Publishing LLC.
Smits, A.J., Ouden, L.P.D., Deunk, J., Bloemers, F.W., LNAZ Research Group, 2020. Incidence of traumatic spinal fractures in The Netherlands: analysis of a nationwide database. Spine 45 (23), 1639–1648. https://doi.org/10.1097/BRS.0000000000003658.
Whiting, P.F., Rutjes, A.W., Westwood, M.E., Mallett, S., Deeks, J.J., Reitsma, J.B., Leeflang, M.M., Sterne, J.A., Bossuyt, P.M., 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann. Intern. Med. 155 (8), 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.
Yang, S., Yin, B., Cao, W., Feng, C., Fan, G., He, S., 2020. Diagnostic accuracy of deep learning in orthopaedic fractures: a systematic review and meta-analysis. Clin. Radiol. 75 (9), 713.e717–713.e728. https://doi.org/10.1016/j.crad.2020.05.021.
Yeh, L.R., Zhang, Y., Chen, J.H., Liu, Y.L., Wang, A.C., Yang, J.Y., Yeh, W.C., Cheng, C.S., Chen, L.K., Su, M.Y., 2022. A deep learning-based method for the diagnosis of vertebral fractures on spine MRI: retrospective training and validation of ResNet. Eur. Spine J. 31 (8), 2022–2030. https://doi.org/10.1007/s00586-022-07121-1.
Young, R.R., 2023. Emerging role of artificial intelligence and big data in spine care. Internet J. Spine Surg. 17 (S1), S3–s10. https://doi.org/10.14444/8504.
Zhang, J., Liu, F., Xu, J., Zhao, Q., Huang, C., Yu, Y., Yuan, H., 2023. Automated detection and classification of acute vertebral body fractures using a convolutional neural network on computed tomography [Original Research]. Front. Endocrinol. 14 https://doi.org/10.3389/fendo.2023.1132725.
Zhou, S., Gao, Y., Li, R., Wang, H., Zhang, M., Guo, Y., Cui, W., Brown, K.G., Han, C., Shi, L., Liu, H., Zhang, J., Li, Y., Meng, F., 2023. Neurosurgical robots in China: state of the art and future prospect. iScience 26 (11), 107983. https://doi.org/10.1016/j.isci.2023.107983.