Machine learning in clinical diagnosis, prognostication, and management of acute traumatic spinal cord injury (SCI): A systematic review

Authors: Nicholas Dietz¹, MD, Vaitheesh Jaganathan¹, Victoria Alkin², Jersey Mettille³, DO, Maxwell Boakye¹, MD, Doniel Drazin⁴, MD

Affiliations:

¹Department of Neurosurgery, University of Louisville, 200 Abraham Flexner Hwy, Louisville, KY, U.S.A- 40202

²Cornell University, Ithaca, New York

³Department of Anesthesia, University of Louisville, Louisville, Kentucky, USA

⁴Department of Neurosurgery, Providence Regional Medical Center Everett, Everett, Washington

Dr. Nicholas Dietz
Department of Neurosurgery,
University of Louisville
200 Abraham Flexner Way, Louisville, KY 40202

Phone : 415-465-0245
Email- nkd25@georgetown.edu

Disclosures: The authors have no conflicts of interest or financial disclosures.
Running title: Machine Learning in Spinal Cord Injury Diagnosis and Management

Tables: 1, Figures: 1

Keywords: spinal cord injury, machine learning, artificial intelligence, prediction tool, spine surgery
Machine learning in clinical diagnosis, prognostication, and management of acute traumatic spinal cord injury (SCI): A systematic review

Abstract

Background: Machine learning has been applied to improve diagnosis and prognostication of acute traumatic spinal cord injury. We investigate potential for clinical integration of machine learning in this patient population to navigate variability in injury and recovery.

Materials and methods: We performed a systematic review using PRISMA guidelines through PubMed database to identify studies that use machine learning algorithms for clinical application toward improvements in diagnosis, management, and predictive modeling.

Results: Of the 132 records identified, a total of 13 articles met inclusion criteria and were included in final analysis. Of the 13 articles, 5 focused on diagnostic accuracy and 8 were related to prognostication or management of traumatic spinal cord injury. Across studies, 1,983 patients with spinal cord injury were evaluated with most classifying as ASIA C or D. Retrospective designs were used in 10 of 13 studies and 3 were prospective. Studies focused on MRI evaluation and segmentation for diagnostic accuracy and prognostication, investigation of mean arterial pressure in acute care and intraoperative settings, prediction of ambulatory and functional ability, chronic complication prevention, and psychological quality of life assessments. Decision tree, random forests (RF), support vector machines (SVM), hierarchical cluster tree analysis (HCTA), artificial neural networks (ANN), convolutional neural networks (CNN) machine learning subtypes were used.
Conclusions: Machine learning represents a platform technology with clinical application in traumatic spinal cord injury diagnosis, prognostication, management, rehabilitation, and risk prevention of chronic complications and mental illness. SVM models showed improved accuracy when compared to other ML subtypes surveyed. Inherent variability across patients with SCI offers unique opportunity for ML and personalized medicine to drive desired outcomes and assess risks in this patient population.
Introduction:

Artificial intelligence (AI), the domain of computer science that perform tasks through human intelligence simulation,¹ is a rapidly growing field that has broad application across industry, science, and technology.² Within AI, Machine Learning (ML) involves creation of algorithms in which machines dynamically learn from data and aim to improve their predictive and computational accuracy over time.¹ The nexus of ML technologies with advances in medicine has contributed to improvement in diagnosis, risk assessment, and predicting response to treatment.³⁻⁷ As advances in basic neuroscience meet technological innovation, neurosurgery has seen unique opportunities for ML in research and clinical application to optimize patient care.¹,⁴,⁸⁻¹⁰ However, challenges exist for complex and non-linear data to integrate clinical importance for prognostic modeling.¹¹⁻¹³

Acute traumatic spinal cord injury (SCI) results in temporary or permanent impairment of motor function and sensation,¹⁴ and has devastating short and long term sequela with significant morbidity and mortality.¹⁵⁻¹⁹ SCI has significant implications on longevity, functional ability, psychological well-being, and socioeconomic stability.¹⁷,¹⁸,²⁰ Prevalence of chronic spinal cord injury is estimated to be almost 300,000 in the United States.²¹ Given the importance of efficient diagnosis and treatment in management SCI, ML technologies have the potential to advance best practices and standard of care.¹,²² Further, personalized or precision medicine for patients with
SCI is useful to tailor expectation and management given the inherent variability among this population in outcome, functional prognosis, and rehabilitation journey. This review provides an overview of current machine learning methods used in diagnosis and management of acute traumatic spinal cord injury aimed at clinical applicability and treatment outcome. Emphasis was placed on studies that aimed to inform clinical improvement and prognostic models.

**Materials and Methods:**

*Data extraction*

We organized the systematic review around the PICOS model (Participants, Intervention, Comparison, Outcomes, Study Design) as previously described, to define the population with spinal cord injury and associated machine learning analysis.

**PICOS Outline**

**Participants:** Adult patients > 18 years with spinal cord injury. **Intervention:** Machine learning algorithms tailored to help SCI diagnosis, management, and were of primary focus. **Comparison:** Comparison of accuracies of machine learning models and clinical were assessed as it pertains to improving patient overall quality of life. **Outcomes:** Efficacy of various machine learning techniques by way of faster, more accurate diagnosis or improvements in clinical treatment and prediction of complications or functional ability were assessed. **Study Design:** Inclusive of prospective, retrospective, case reports, case series, observational studies, comparative studies and randomized clinical trials. No reviews or meta-analyses were included. **Search Criteria:** We followed the PRISMA guidelines (2009) to construct the framework of the systematic review and conducted the search on June 3rd, 2022 using PubMed databases. Additional articles used in
the references were incorporated from the references of those articles identified in the searches. We used Keyword and MeSH terms for predictive outcomes to include the following terms with numbered iterations for the database as follows: Spinal Cord Injury and Machine Learning. Eligibility criteria: Current publication in English language:


Risk of bias evaluation: Assessment of conflict of interest, funding for study and study design were assessed according to QUADAS criteria.

Exclusion criteria:

Systematic reviews and meta-analyses were excluded. Studies with nonhuman subjects, pediatric population, language other than English, case studies, brain-computer interface focus, and those without full text were also excluded.

Data Evaluation:

We followed the QUADAS tool to evaluate risk bias and result applicability of the studies according to 2003 guidelines. A total of 13 questions on the QUADAS survey that cover patient selection, index test, reference standard, and timing were addressed for each study and incorporated into the final analysis.

Results:

Of the 132 records identified, a total of 13 articles were included in final analysis and met inclusion criteria. Figure 1, Table 1, Table 2. Of the 13 articles, 5 focused on diagnostic accuracy and 8 were related to prognostication or management of traumatic spinal cord injury. Across studies, 1,983 patients with spinal cord injury were evaluated with most classifying as ASIA C or D. Retrospective designs were used in 10 of the 13 studies and 3 were prospective. Studies focused on MRI evaluation and segmentation for diagnostic accuracy and prognostication, investigation of mean arterial pressure in acute care and intraoperative settings, prediction of ambulatory and functional ability, long term complication prevention, and psychological quality of life assessments. Decision tree, random forests (RF), support vector
machines (SVM), hierarchical cluster tree analysis (HCTA), artificial neural networks (ANN), convolutional neural networks (CNN) machine learning subtypes were used. SVM was shown to have superior predictive ability.²⁸,³²

Diagnosis of SCI

Of the 5 diagnostic-related studies on ML in SCI, all were retrospectively designed with a total of 478 patients. Classifications were 77 with ASIA A, 65 with ASIA B, 142 ASIA C, and 139 ASIA D. MRI and other diagnostic modalities were investigated to improve assessment accuracy and segmentation of cord injury.

Management of SCI

Most of the studies (n=8) were focused on prognostication and management of SCI, with 3 prospective and 5 retrospectively designed. A total of 1,505 patients were included with classification breakdown of 387 ASIA A, 480 ASIA B, 195 ASIA C, 481 ASIA D. Topics included blood pressure management, prognosis with predictive modeling, and psychological outcome prediction.

Discussion:

The present review demonstrates ML contributions toward 1) diagnostic accuracy with advances in MRI segmentation and classification 2) acute blood pressure management for optimizing functional AISIA classification 3) quality of life clustering with predictive modeling of patient needs 4) prevention of long term complications and morbidity 5) functional prognostication with predictive modeling. A recognized purpose of ML is to identify key aspects and elements of
injury or patient characteristics that drive outcomes. Given the variability in injury type and
rehabilitative recovery, ML may meet the need for more personalized treatment approaches for
patients with SCI.\textsuperscript{13,40} Machine learning may have better efficacy in uniting datasets to create
accurate algorithmic models for clinical prediction in iterative generations of modeling.\textsuperscript{13}

\textit{Diagnosis}

Early and accurate diagnosis and classification of spinal cord injury (complete versus
incomplete) has implications on appropriate medical and surgical management for acute spinal
cord injury. To date, surgical decompression within 24 hours is the most effective intervention
for primary injury when indicated to limit progressive damage.\textsuperscript{17,41,42} As the gold standard for
spinal cord imaging, MRI aids in characterization of SCI through presence of hematoma and
edema, visualization of ligamentous and disc disruption, surgical considerations, and
implications for classification and prognosis.\textsuperscript{43,44} In the setting of polytrauma, patients may not
be able to participate in a reliable neurological exam, potentially masking the nature of SCI. In
such circumstances of suspected SCI without reliable exam, there is added emphasis on imaging
techniques to characterize the nature of a spinal cord injury. Arslan and colleagues have
demonstrated enhanced diagnosis of spinal cord injury (SCI) on unconscious and uncooperative
patients using dermatomal skin impedance analysis through artificial neural network (ANN) to
aid in diagnosis of SCI.\textsuperscript{28} In this study, classification methods of Support Vector Machines
(SVM) and hierarchical cluster tree analysis (HCTA) showed improved diagnostic rates.
Comparing three classification methods, the study showed that SVM and HCTA attained higher
recognition accuracies between SCI and control group using skin impedance than ANN, aiding
in SCI diagnosis with reduced need for reliance on patient clinical exam. Interestingly, Tay et al.
used ML to characterize diffusion tensor imaging (DTI) fractional anisotropy and fiber tracking
in the evaluation of white matter tract disruption in the setting of SCI for diagnosis. Pattern recognition algorithms were used to locate the region of cord injury with T2 and T1 sequences in different planes, and assess DTI diffusion anisotropy and fractional anisotropy to evaluate lesions and structural damage in the cord. Medical imaging will continue garner significant application of machine learning for diagnosis and prognosis, and may offer particular benefit to management of spinal cord injury given its acuity, complexity variability, and multimodal treatment paradigms.

Blood pressure management

A cornerstone of the acute clinical care of patients with traumatic spinal cord injury is maximizing spinal cord perfusion with mean arterial pressure (MAP) goals. Guidelines suggest maintaining a MAP of 85-90 for the first 7 days after spinal cord injury is related to optimal recovery of function, considering post-injury ASIA grading. However, little high-quality clinical data confirms the duration and blood pressure parameters following acute spinal cord injury. Several studies have explored blood pressure augmentation with one suggesting that MAP may be most beneficial in the first 2-3 days following injury in the intensive care unit. It has also been shown that certain vasopressor agents such as levophed or phenylephrine are preferred when compared to dopamine, given its higher risk of cardiac comorbidity. In 2021, Torres-Espín et al. used a topological network machine learning of patients to predict neurological recovery after spinal cord injury (SCI) on the basis of mean arterial pressure (MAP) during surgery. This similarity network was built using techniques including topological network extraction, dimensionality reduction, regression analysis, and prediction modeling and programming languages such as MATLAB and R. The network analysis showed that time outside of an optimal MAP range (hypotension or hypertension) during surgery was associated
with lower likelihood of neurological recovery. The optimum MAP range associated with neurological recovery was found to be 76 [104-117] mmHg, such that MAPs outside this range result in lower likelihood of neurological recovery. These findings suggest updated MAP targets for therapeutic intervention given the gap in knowledge from validation of clinical MAP guidelines.\textsuperscript{48} Another study associated with the Transforming Research and Clinical Knowledge for Spinal Cord Injury (TRACK-SCI) investigators published in 2022 by Chou et al. used a model-agnostic framework to further strengthen the automated machine learning (AutoML) applied to intraoperative hemodynamic monitoring and elucidated prognostic risks of intraoperative hypertension.\textsuperscript{29} Using retrospective intraoperative records and clinical data recordings, the group demonstrated a way to build an AutoML model that quantified feature importance and interpretation, recognized feature instability, and was designed for model reproducibility and translational application. TRACK-SCI retrospectively builds on previous work and via incorporating machine learning (ML) to procure clinically actionable thresholds for intraoperative MAP management intraoperatively with Agarwal and colleagues in 2022.\textsuperscript{27}

Seventy-four patients were retrospectively analyzed following SCI. All patients had hemodynamic monitoring with recordings at 5-minute intervals, which resulted in a total of 28,954 minutes and 5718 unique data points for each parameter, such as the type of vasopressor used, dosage, any drug related complications, the average intraoperative MAP and extreme MAP range (<76 mmHg or >104 mmHg). Outcomes were determined by ASIA grades. Factors related to improved AIS grade were determined through generation of random forests with 10,000 iterations. An average MAP value between 80 and 96 mmHg was associated with improved neurological function while an accumulated time of 93 minutes or longer outside of the MAP range of 76-104 mmHg was associated with worse neurological function.
Quality of Life and Psychological State

Nearly half of all patients with SCI experience mental health issues, with depression (37%) and anxiety representing the most common diagnoses. Complex factors underlying psycho-social state and emotional response for those with SCI affect quality of life and may present predictable trends that lend toward improved assistance from multidisciplinary teams. Gibert and colleagues use artificial intelligence (AI) to guide interpretation-oriented tools on quality of life for patients with SCI and create “Clustering based on rules” (CIBR)—a type of integral knowledge database that allows tracking and prediction of QoL. Four types of patterns have been identified that have been used to identify psychological distress or marked decreases in QoL. Using expert opinions and interviews of 109 participants with SCI, the authors cluster mental health profiles through “knowledge discovery from data” (KDD) that informs mental and psych-social health of patients with SCI observe over time. Older age and longer time since injury were found to be associated with decreased QoL and determined as higher risk of distress and need for social support.

Risk prevention

Chronic sequela of SCI contribute to notable morbidity and mortality and include development of pressure ulcers, pulmonary complications, dysautonomia, cardiovascular disease, metabolic derangements, sexual dysfunction, urinary and bowel dysregulation. Pressure ulcers are one of the most common complications of SCI found in as many as 85% of patients, and pose a significant risk for infection and sepsis that is responsible for 8% of deaths in those with SCI. Despite its significant cost to the healthcare system, pressure ulcers in SCI have not been recognized for the variability in individual need for positioning changes on a patient by patient basis. In 2014, Fu and colleagues create an intelligent model through machine learning
(artificial neural network (ANN), random forest (RF), and support vector machines (SVM)) to identify the optimal wheelchair tilt and recline to mitigate the risk of pressure ulcers for each patient.\textsuperscript{32} As wheelchair tilt and recline are the most efficacious biomechanical features of reducing ulcer risk through improving tissue blood flow, Doppler recordings of varying blood flow metrics at graded wheelchair angles allowed intelligent model algorithms to predict optimal patient positioning. Of the ML evaluations, SVM had the highest predictive accuracy for subset attributes at 78.8\% after comparison with ANN, RF, C4.5. Future models in machine learning may capitalize on opportunities to address the multitude of complications following spinal cord injury and stratify risk based on personal characteristics and differing patient needs.

\textit{Prognosis and Recovery}

Functional prognosis after spinal cord injury with potential to regain walking ability and functional independence is a focus of rehabilitation for clinicians and patients in the setting of traumatic SCI.\textsuperscript{55} While the natural history of SCI recovery may be difficult to predict, ASIA grade with injury severity determination of complete versus incomplete motor impairments has been most frequently used as the benchmark for clinical recovery and provided inference for long-term functional prognosis.\textsuperscript{55-57} However, nuances of motor-level recovery and functional differences within ASIA grades have been shown to clinically meaningful for patients with SCI.\textsuperscript{24} For example, improving a motor level from a C5 ASIA A grade to C7 or C8 ASIA A with improved upper extremity and hand function was shown to be perceived by patients as significant in value for QoL and independence.\textsuperscript{24} Several guidelines for prediction of ambulation have aided in prognostication after tSCI, including an accurate and user-friendly prediction rule from 2011\textsuperscript{58} that incorporates age, lower extremity muscle group function assessments and dermatomal assessments to predict ambulatory ability through logistic regression with high
accuracy. ML may further aid in ambulatory prognosis and personalize rehabilitative care and prediction of ability to regain independence for patients with SCI. In 2016, Zariffa et al. used ML to define the predictive value of impairment measures as through a Graded Redefined Assessment of Strength, Sensibility and Prehension (GRASSP) in traumatic cervical spinal cord injury (SCI).\textsuperscript{39} A dataset of 138 GRASSP assessments was analyzed, and machine learning classifiers were trained to predict scores on prehension performance tasks. The study found that upper extremity motor impairment measures were highly predictive of task performance (Spearman’s p of 0.84), indicating that automated assessments of upper extremity function after SCI can measure impairment and estimate other measures accordingly. In a prospective study by Facchinello et al., regression tree analysis of clinical and demographic parameters was used to predict functional outcome following traumatic spinal cord injury.\textsuperscript{31} Models using 4 and 11 predictors explained 51.4% and 62.3% of the variance of the Spinal Cord Independence Measure (SCIM) within the first year after injury after validation, respectively. More predictive variables enhanced accuracy and the most important predictor with discriminatory power was found to be the severity of the neurological deficit (ASIA grade) on admission, followed by Injury Severity Score (ISS), age, neurological level, and delay before surgery. In 2021, Okimatsu and colleagues developed a deep learning-based radiomics model (DLR) to quantify radiographic characteristics using a convolutional neural network (CNN) to stratify 1-month neurological outcomes of patients with acute spinal cord injury (SCI).\textsuperscript{36} The goal of the study was to parse injury heterogeneity across 215 medical records and predict short term functional prognosis of patients with cervical SCI. Accuracy of the final model was 71% in prediction of outcome. In 2019, DeVries et al. developed an unsupervised ML model that assessed the functional independence measure (FIM) locomotion score to predict ambulatory ability at 1 year following injury and
compare results with previous predictive models. The unsupervised model used complete neurological data from index hospitalization to inform the algorithm and assess if this improved accuracy of walking prediction compared to previously described non-ML algorithms. Ultimately, no significant improvement in accuracy was achieved with more data in the unsupervised ML model compared to prior studies. Additionally, it was shown that F1-score for accuracy and validation was less inclined to inaccuracies from false negatives that AUC failed to capture. In 2018, McCoy and colleagues as part of the TRACK-SCI group use a prospectively derived 2D convolutional neural network (CNN) for automatic segmentation and classification of cord and contusion injury in patients with acute traumatic SCI. For 47 patients within 24 hours of injury, an image-analysis pipeline with 2D CNNs was used for whole spinal cord and intramedullary spinal cord segmentation. Linear mixed modeling was used to compare the results of this study's CNN segmentation and current tools. The segmentation method using the BASICseg-3 algorithm developed in this study compares favorably with available state-of-the-art segmentation methods and correlated with lower extremity motor scores post-injury. As such, modeling of segmentation may stratify injury severity, classification, and contribute toward triage and prognosis in blunt SCI.

Future directions

In recent years, increased focus on molecular pathways of injury and recovery have elucidated the natural history of spinal cord injury and its management. Future studies may use machine learning in the basic science level to identify biological substrates, inflammatory markers, and pharmaceuticals to drive discovery in therapeutic developments. In 2015, Al-Ali used ML to locate kinases that aid in axonal regeneration following spinal cord injury in rat models. In coming years, with the potential advent of spinal cord perfusion pressure monitoring, ML will
surely have a role in validating optimal goals for perfusion pressures with consideration of the many injury and demographic characteristics. Many ML studies have evaluated functional and motor prognosis after SCI, however, an even more prominent concern for patient QoL is the need to address bowel and bladder dysregulation assessments and likelihood of recovery. ML may also aid in symptomatic management of spasticity to help navigate variability in dosing, efficacy, and complications related to anti-spastic agents such as baclofen. Further, as interventions such as spinal cord stimulation for spinal cord injury, closed loop feedback and integration of sensing evoked continuous action potentials to optimize therapeutic intervention for motor control and pain may be augmented by ML capability. SVM ML modeling may be used as it has been shown to be more accurate when compared to alternative forms of ML. SVM modeling creates linear regression through a hyperplane that manages large volumes of input variables and optimizes distance between outcomes. SVM also has improved accuracy for classification in small sample sizes and has been shown to have promise in clinical application for prediction of heart failure and cancer genomics. Given its early comparative success in predicting diagnostic and prognostic management in SCI, SVM will likely see a continued role for execution in the evolution of later clinical models. F1-score for efficacy assessment may also be considered to evaluate ML models given its predictive value against standard area under the curve. Clinical application of ML tools requires a judicious iterative process to evolve algorithms and build healthcare’s trust on reliability and efficacy to improve patient care.

**Strength and Limitations:**
While a few studies involved prospective methods, most of the studies included in final analysis were retrospective in design. A limitation of the analysis is the nature of retrospective data collection, especially for informing predictive modeling that is likely best created through prospective methods. Additionally, a small cohort of studies exists in the literature regarding applications of ML in diagnosis and management of SCI. While we anticipate its growth, current literature on this topic is limited and variable in topic and design, and not conducive to meta-analysis or direct comparison. Future studies may validate ML-derived predictive scoring systems, for example, or compare algorithms as more investigations are completed. Also, variability in ASIA score and type for assessment and prediction should be carefully assessed and may deviate from traditional linear regression modeling for predictive algorithms. Large numbers of variables derived from small datasets may also feed unstable model behavior without clinically meaningful outcome. Indeed, severity of injury was shown to be one of the most important factors in determining long-term functional outcome. Finally, models should be externally validated and carefully applied before clinical use and reliance.

**Conclusions:**
Machine learning represents a platform technology with clinical application in traumatic spinal cord injury diagnosis, prognostication, management, rehabilitation, and risk prevention of chronic complications and mental illness. SVM models showed improved accuracy when compared to other ML subtypes surveyed. Inherent variability across patients with SCI offers unique opportunity for ML and personalized medicine to drive desired outcomes and assess risks in this patient population.
References:


<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Number of Patients with SCI</th>
<th>ASIA Impairment Scale (AIS)</th>
<th>Topic</th>
<th>Cohort Study</th>
<th>Type of ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agarwal, et al.</td>
<td>2022</td>
<td>74</td>
<td>A (n=37), B (n=8), C (n=14), D (n=13), E (n=2)</td>
<td>D</td>
<td>Retrospective</td>
<td>Deep learning</td>
</tr>
<tr>
<td>Arslan, et al.</td>
<td>2012</td>
<td>15</td>
<td>NA</td>
<td>D</td>
<td>Retrospective</td>
<td>Support vector machine</td>
</tr>
<tr>
<td>Chou, et al.</td>
<td>2022</td>
<td>59</td>
<td>NA</td>
<td>M</td>
<td>Prospective</td>
<td>Auto-regression</td>
</tr>
<tr>
<td>DeVries, et al.</td>
<td>2020</td>
<td>862</td>
<td>A=257, B=447, C=149, D=379</td>
<td>M</td>
<td>Retrospective</td>
<td>Unsupervised and Supervised</td>
</tr>
<tr>
<td>Facchinello, et al.</td>
<td>2021</td>
<td>172</td>
<td>A=68, B=17 C=25 D=62</td>
<td>M</td>
<td>Prospective</td>
<td>Regression tree</td>
</tr>
<tr>
<td>Fu, et al.</td>
<td>2014</td>
<td>NA</td>
<td>NA</td>
<td>M</td>
<td>Retrospective</td>
<td>Artificial neural network (ANN)</td>
</tr>
<tr>
<td>Gibert, et al.</td>
<td>2009</td>
<td>109</td>
<td>NA</td>
<td>M</td>
<td>Retrospective</td>
<td>Knowledge extraction</td>
</tr>
<tr>
<td>Inoue, et al.</td>
<td>2020</td>
<td>165</td>
<td>A = 15, B = 38, C = 66, D = 44, E = 3</td>
<td>D</td>
<td>Retrospective</td>
<td>Extremity motor impairment tree</td>
</tr>
<tr>
<td>McCoy, et al.</td>
<td>2019</td>
<td>47</td>
<td>A (n=10), B (n=3), C (n=5), D (n=21), Unknown (n=8)</td>
<td>M</td>
<td>Prospective</td>
<td>Convolutional neural network</td>
</tr>
<tr>
<td>Okimatsu, et al.</td>
<td>2022</td>
<td>215</td>
<td>A (n=25), B (n=19), C (n=62), D (n=82), E (n=27)</td>
<td>D</td>
<td>Retrospective</td>
<td>Deep learning</td>
</tr>
<tr>
<td>Tay, et al.</td>
<td>2014</td>
<td>9</td>
<td>NA</td>
<td>D</td>
<td>Retrospective</td>
<td>Support vector machine</td>
</tr>
<tr>
<td>Torres-Espín, et al.</td>
<td>2021</td>
<td>118</td>
<td>A (n=52), B (n=13), C (n=16), D (n=19), E (n=4), Missing (n=14)</td>
<td>M</td>
<td>Retrospective</td>
<td>Topological regression</td>
</tr>
<tr>
<td>Zariffa, et al.</td>
<td>2016</td>
<td>138</td>
<td>NA</td>
<td>M</td>
<td>Retrospective</td>
<td>Classification</td>
</tr>
</tbody>
</table>
Table 2. QUADAS Bias Table

<table>
<thead>
<tr>
<th>QUADAS Question</th>
<th>Yes</th>
<th>No</th>
<th>Unclear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the spectrum of patients representative of the patients who will receive the test in practice?</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Were selection criteria clearly described?</td>
<td>11</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Is the reference standard likely to correctly classify the target condition?</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?</td>
<td>10</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Did patients receive the same reference standard regardless of the index test result?</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Was the execution of the index test described in sufficient detail to permit replication of the test?</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Was the execution of the reference standard described in sufficient detail to permit its replication?</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Were the index test results interpreted without knowledge of the results of the reference standard?</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Were the reference standard results interpreted without knowledge of the results of the index test?</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Were uninterpretable/intermediate test results reported?</td>
<td>9</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Were withdrawals from the study explained?</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
</tbody>
</table>
QUADAS criteria represented above with tally of ratings for all 13 studies included in the final analysis: (Y): Yes; (N): No; (U): Unclear. The "reference standard" was determined to be the control cohort for those studies that incorporated randomized control trials.
Figure 1: PRISMA flowchart of systemic search strategy
Declaration of interests

☒ 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: