



ORIGINAL

# Uso de inteligencia artificial para predecir complicaciones en cirugías de columna toracolumbar degenerativa: revisión sistemática

G. Ricciardi<sup>a,b,c,\*</sup>, J.I. Cirillo Toter<sup>d,e,f</sup>, R. Pons Belmonte<sup>g</sup>, L. Romero Valverde<sup>a</sup>, F. López Muñoz<sup>h</sup> y A. Manríquez Díaz<sup>i</sup>

<sup>a</sup> Sanatorio Güemes, Buenos Aires, Argentina

<sup>b</sup> Centro Médico Integral Fitz Roy, Buenos Aires, Argentina

<sup>c</sup> Hospital Álvarez, Buenos Aires, Argentina

<sup>d</sup> Facultad de Medicina, Hospital del Trabajador, Santiago, Chile

<sup>e</sup> Clínica Universidad de los Andes, Santiago, Chile

<sup>f</sup> Facultad de Medicina, Universidad Andrés Bello, Santiago, Chile

<sup>g</sup> Sanatorio Argentino, Hospital Marcial Quiroga, San Juan, Argentina

<sup>h</sup> Hospital del Trabajador, Santiago, Chile

<sup>i</sup> Clínica Francesa de Mendoza, Mendoza, Argentina

Recibido el 6 de septiembre de 2024; aceptado el 15 de febrero de 2025

Disponible en Internet el 18 de febrero de 2025

## PALABRAS CLAVE

«Inteligencia artificial»;  
«Aprendizaje mecanizado»;  
«Aprendizaje profundo»;  
«Redes neuronales artificiales»;  
«Patología degenerativa»;  
«Deformidad del adulto»

## Resumen

**Introducción:** El objetivo de los autores es realizar una revisión sistemática de la bibliografía para evaluar la efectividad de los modelos predictivos de inteligencia artificial en la predicción de complicaciones en pacientes adultos tratados mediante cirugía por enfermedad toracolumbar degenerativa, en comparación con otras técnicas predictivas de uso habitual.

**Materiales y métodos:** Se realizó una revisión sistemática de la bibliografía en Medline/Pubmed, Cochrane Library y Lilacs/Portal de la BVS sobre la efectividad del uso de modelos predictivos de inteligencia artificial para las posibles complicaciones en pacientes operados por enfermedad degenerativa de la columna toracolumbar durante el periodo de 1 de enero de 2000 y 1 de mayo de 2023. El riesgo de sesgo se evaluó con las herramientas ROBINS-I y PRO-BAST. Se registraron características de los estudios y resultados, contemplando como desenlace complicaciones generales o específicas.

**Resultados:** Se identificaron 2.321 títulos, 763 eran duplicados. Se realizó el cribado de 1.558 títulos; 22 fueron elegidos para su lectura completa con exclusión de 18 y elección final de 4 publicaciones para la siguiente revisión. Adicionalmente, se incluyeron 8 publicaciones desde

\* Autor para correspondencia.

Correo electrónico: [guillermoricciardi@gmail.com](mailto:guillermoricciardi@gmail.com) (G. Ricciardi).

otras fuentes (Biblioteca Asociación Argentina de Ortopedia y Traumatología, con búsqueda manual de citas). En 5 artículos (41,6%) se compararon la efectividad de modelos predictivos de inteligencia artificial frente a técnicas habituales. Todos fueron catalogados globalmente con muy alto riesgo de sesgo. Dada la heterogeneidad de las muestras, los resultados de interés y las métricas de evaluación de los algoritmos, no se realizó un metaanálisis.

**Conclusión:** Si bien la evidencia disponible es limitada y presenta un alto riesgo de sesgo, los estudios analizados indican que estos modelos pueden alcanzar un desempeño prometedor en la predicción de complicaciones, con valores del área bajo la curva que, en su mayoría, oscilan entre aceptables y excelentes.

© 2025 SECOT. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la CC BY-NC-ND licencia (<http://creativecommons.org/licencias/by-nc-nd/4.0/>).

## KEYWORDS

«Artificial intelligence»;  
«Machine learning»;  
«Deep learning»;  
«Artificial neural networks»;  
«Degenerative pathology»;  
«Adult deformity»

## Use of artificial intelligence to predict complications in degenerative thoracolumbar spine surgery: A systematic review

### Abstract

**Objective:** We aim to conduct a systematic review of the literature to evaluate the effectiveness of artificial intelligence prediction models in predicting complications in adult patients undergoing surgery for degenerative thoracolumbar pathology compared with other commonly used prediction techniques.

**Methods:** A systematic literature review was conducted in Medline/Pubmed, Cochrane Library, and Lilacs/Portal de la BVS to identify machine learning models in predicting complications in patients undergoing surgery for degenerative thoracolumbar spine pathology between January 1, 2000, and May 1, 2023. The risk of bias was assessed using the PROBAST tool. Study characteristics and outcomes focusing on general or specific complications were recorded.

**Results:** A total of 2,341 titles were identified (763 were duplicates). Screening was performed on 1,578 titles, and 22 were selected for full-text reading, with 18 exclusions and 4 publications selected for the subsequent review. Additionally, 8 publications were included from other sources (Argentine Association of Orthopedics and Traumatology Library; manual citation search). In 5 (41.6%) articles, the effectiveness of artificial intelligence predictive models was compared with conventional techniques. All were globally classified as having a very high risk of bias. Due to heterogeneity in samples, outcomes of interest, and algorithm evaluation metrics, a meta-analysis was not performed.

**Conclusion:** Although the available evidence is limited and carries a high risk of bias, the studies analysed suggest that these models may achieve promising performance in predicting complications, with area under the curve values mostly ranging from acceptable to excellent.

© 2025 SECOT. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introducción

Según estadísticas de Estados Unidos, el costo estimado de la enfermedad vertebral degenerativa es de alrededor de 100 billones de dólares anuales<sup>1</sup>. Se calcula que 2 de cada 3 adultos experimentarán de dolor lumbar bajo en algún momento de su vida<sup>2</sup>. La complejidad de los pacientes con enfermedad de la columna vertebral y las complicaciones asociadas con la cirugía han motivado la investigación de estrategias para una predicción precisa de estos episodios, así como para la estimación anticipada de los resultados clínicos. Tradicionalmente distintos modelos de análisis estadístico han permitido identificar factores predictivos para las complicaciones, con gran popularidad de los modelos de análisis multivariado, como la regresión logística, que produce una medida de riesgo (*odds ratio*) para variables independientes sobre un efecto o resultado específico<sup>3</sup>.

El campo de la inteligencia artificial (IA) ha impactado notablemente en múltiples áreas de la salud y la cirugía de columna no es la excepción<sup>3,4</sup>. La IA se ocupa no solo de comprender, sino también de construir «entidades inteligentes»: máquinas que pueden calcular cómo actuar de manera efectiva y segura<sup>4</sup>. La IA comprende diversas disciplinas que incluyen: el procesamiento natural del lenguaje, la representación del conocimiento, el razonamiento automatizado, el aprendizaje de la máquina (*machine learning* o ML) y la robótica. El ML es una subárea que permite al sistema aprender y retroalimentarse a sí mismo; es decir, desarrollar algoritmos que mejoran con la experiencia. El ML involucra numerosos métodos, como el aprendizaje profundo (*deep learning*), basado en redes neuronales artificiales<sup>3,4</sup>. El ML ha permitido desarrollar modelos predictivos y en la última década se han publicado numerosos artículos para su aplicación en áreas específicas, como la de la cirugía de columna vertebral<sup>3,4</sup>.

El objetivo de los autores es realizar una revisión sistemática de la bibliografía para evaluar la efectividad de los modelos predictivos de inteligencia artificial en la predicción de complicaciones en pacientes adultos tratados mediante cirugía por enfermedad toracolumbar degenerativa, en comparación con otras técnicas predictivas de uso habitual.

## Materiales y métodos

Se realizó una revisión sistemática de la bibliografía en las principales bases de datos biomédicos (Medline/Pubmed, Cochrane Library y Lilacs/Portal de la BVS) sobre la efectividad del uso de modelos predictivos de IA para predecir complicaciones en pacientes operados por enfermedad degenerativa de la columna toracolumbar durante el periodo comprendido entre el 1 de enero de 2000 y el 1 de mayo de 2023.

## Criterios de elegibilidad

Se seleccionaron estudios de acuerdo con los siguientes criterios de elegibilidad:

Diseños de estudios: ensayos clínicos controlados y aleatorizados, estudios prospectivos no aleatorizados, estudios observacionales de cohortes prospectivos y retrospectivos, estudios transversales y series descriptivas con más de 10 casos. Fueron excluidos los reportes de caso, revisiones (sistemáticas, narrativas), editoriales, cartas al editor y consensos.

Participantes: pacientes adultos (18-65 años) de ambos sexos tratados quirúrgicamente por enfermedad degenerativa de la columna toracolumbar (hernia de disco, canal estrecho lumbar y deformidad del adulto, sagital o coronal). Fueron excluidos los estudios de población con escoliosis idiopática, neuromuscular, congénita o sindrómica, fracturas por osteoporosis/enfermedad metabólica, artritis reumatoide, espondilitis anquilosante/hiperostosis esquelética idiopática difusa, enfermedad oncológica vertebral y los de pacientes tratados mediante bloqueos como procedimiento terapéutico único (sin cirugía).

Intervención: uso de IA para la creación de modelos predictivos de complicaciones contemplando modelos de aprendizaje profundo (*deep learning*), aprendizaje mecánico (*machine learning*), red neuronal artificial (*artificial neural network*) y otros métodos novedosos cuyo desarrollo implique el uso de inteligencia artificial. Se excluyeron los estudios que utilizaran modelos de IA para otros objetivos distintos de la predicción de complicaciones, como evaluación de pacientes e imágenes, clasificación, aplicación en cirugía navegada o robótica.

Comparador: otros métodos habituales para predecir complicaciones como métodos estadísticos o escalas. Debido a la novedad del tema, adicionalmente se contemplaron estudios sin comparador.

Resultados (*outcomes*): estudios que registran complicaciones en pacientes quirúrgicos por enfermedad toracolumbar degenerativa, contemplando, principalmente complicaciones intraoperatorias y postoperatorias tempranas (90 días desde la cirugía). Secundariamente, complicaciones en plazos mayores (6 meses, 1 y 2 años)

y otras variables de resultados, como dolor, discapacidad funcional, tiempo de internación, reingresos y morbilidad.

Tiempo: estudios con tiempo de seguimiento mayor o igual a 90 días.

Idioma: estudios en inglés, español y portugués.

En la [tabla 1](#) se resume la pregunta de investigación de acuerdo con el modelo PICO, que permitió estructurar el problema científico, describir los criterios de elegibilidad y guiar la búsqueda bibliográfica.

## Fuentes de información

Se desarrolló una estrategia de búsqueda bibliográfica en las bases de datos MEDLINE, Cochrane y LILACS (Literatura Latinoamericana y del Caribe en Ciencias de la Salud) a través de los buscadores Pubmed, Cochrane Library y el portal de la Biblioteca Virtual en Salud (BVS). Adicionalmente, se contemplaron otras fuentes de citas bibliográficas, como la consulta a la biblioteca de la Asociación Argentina de Ortopedia y Traumatología y la búsqueda manual en las listas de referencias de los estudios incluidos o revisiones (narrativas/sistemáticas) identificadas durante la búsqueda (*snowballing* o bola de nieve).

## Estrategia de búsqueda

Se desarrolló una estrategia de búsqueda con términos MESH y palabras clave sobre el uso de inteligencia artificial en la predicción de complicaciones de pacientes tratados con cirugías de columna toracolumbar degenerativa. La estrategia fue desarrollada por el equipo de investigadores y se describe a continuación: (((((artificial intelligence) OR (deep learning)) OR (machine learning)) OR (AI)) OR (inteligencia artificial)) AND (spine)) AND (((thoracolumbar) OR (lumbar)) OR (thoracic)) OR (lumbosacral)). La búsqueda bibliográfica se limitó, según los filtros de idioma (español, inglés y portugués) y fecha, contemplando como periodo de tiempo el comprendido entre el 1 de enero del 2000 y el 1 de mayo del 2023. No se utilizaron filtros de búsqueda sobre el diseño o tipo de estudio.

## Gestión de datos

Los resultados de la búsqueda bibliográfica fueron cargados en el programa Zotero, programa para administrar citas bibliográficas y facilitar la colaboración entre los revisores durante el proceso de selección de estudios. Se cargaron los resúmenes. Se eliminaron los duplicados. Antes del proceso de selección formal, se proporcionó capacitación a los miembros del equipo de revisión que no estaban familiarizados con el programa.

## Proceso de selección

Los revisores se agruparon en 2 grupos de 2 miembros cada uno; ambos grupos examinaron de forma independiente los títulos y resúmenes según los criterios de inclusión. Los desacuerdos fueron resueltos a través de la discusión entre los revisores y, eventualmente, por tercera opinión de un

**Tabla 1** Research question according to PICO model

| PICO         | Inclusion                                                                                                                                                                                                                                                                                    | Exclusion                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Patients     | Surgically treated adult patients (aged 18-65 years) of both sexes with degenerative thoracolumbar spine conditions, including herniated disc, lumbar stenosis, and adult spinal deformity (sagittal and/or coronal).                                                                        | Conditions such as idiopathic, neuromuscular, congenital or syndromic scoliosis, fractures caused by osteoporosis/metabolic disease, rheumatoid arthritis, ankylosing spondylitis/hyperostosis, diffuse idiopathic skeletal (DISH) disease, spinal oncologic conditions, patients who undergo blocks as a sole therapeutic procedure (without surgery). Studies that use artificial intelligence models for purposes other than prediction of complications were excluded. |
| Intervention | The utilization of artificial intelligence in developing predictive models for complications. We take into account methods such as deep learning, machine learning, artificial neural networks, and other new approaches that involve artificial intelligence.                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Comparison   | Other frequently used methods to predict complications, such as statistical models or measurement scales, are also considered. Due to the newness of the topic, studies without a comparison group were also included in the analysis.                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Outcome      | Studies reporting complications, with a focus on intraoperative and early postoperative complications (within 90 days of surgery). Furthermore, we examined complications beyond the 90-day period, up to 6 months, 1 year, and 2 years. Additionally, we considered specific complications. | No complications were recorded.                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Time         | Studies with a follow-up period of 90 days or longer.                                                                                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Study design | Controlled randomized clinical trials (RCTs), prospective non-randomized studies, prospective and retrospective cohort studies, cross-sectional studies, and descriptive series with over 10 cases.<br>Language: English, Spanish, and Portuguese.                                           | Case reports, systematic and narrative reviews, editorials, letters to the editor, and consensus papers were excluded.                                                                                                                                                                                                                                                                                                                                                     |

PICO: P=Patient; I=Intervention; C=Comparator; O=Outcome.

revisor adicional, miembro experimentado del equipo de investigadores. Después de la selección de artículos elegibles para la revisión de texto completo, todos los artículos de texto completo se recuperaron a través de fuentes de la biblioteca. Ambos grupos de revisores procedieron a evaluar los artículos de texto completo que habían sido seleccionados por el otro, y viceversa, (diseño cruzado) para limitar el posible sesgo de selección del revisor. Durante la revisión del texto completo, las referencias de los artículos también se revisaron para determinar su posible elegibilidad (bola de nieve). Nuevamente, cualquier posible conflicto fue resuelto primero por los revisores de cada grupo y, si era necesario, por la tercera opinión de un revisor adicional con experiencia.

### Extracción de datos

La extracción de datos fue realizada por duplicado y los revisores a cargo trabajaron de forma independiente. Se

registraron los datos en tablas. Una tabla sobre las características de los estudios incluidos contemplaba: autor, año, países participantes enfermedad en estudio, algoritmo utilizado, número de sitios participantes, tamaño de la muestra, variable de resultado (complicaciones en general o de tipo específico), origen de los datos (base de datos), validación, resultados reportados, precisión (porcentaje), área bajo la curva (AUC ROC), características operativas (sensibilidad, especificidad). También se registraron los criterios de inclusión y exclusión, características demográficas de los participantes, periodo de seguimiento, datos sobre financiación y posibles conflictos de interés.

### Evaluación del riesgo de sesgo

El riesgo de sesgo de los estudios observacionales no aleatorios se evaluó con la herramienta ROBINS-I<sup>5</sup>. Para evaluar el riesgo de sesgo en el uso de modelos predictivos de riesgo, se contempló la herramienta PROBAST<sup>6</sup>. La evaluación del

sesgo fue realizada por al menos 2 evaluadores en forma independiente. Los conflictos se resolvieron por consenso.

Para garantizar la consistencia, el autor principal examinó todos los resúmenes y textos completos para verificar su elegibilidad, extrajo los datos y evaluó el riesgo de sesgo en todos los estudios incluidos.

### Estrategia para la síntesis de datos

Posteriormente, todos los resultados de los revisores individuales se combinaron en una sola tabla de datos. Esta tabla fue discutida con el equipo completo de revisores para llegar a un consenso sobre los resultados de nuestra revisión.

Para la evaluación del rendimiento de los modelos predictivos se contemplará principalmente el AUC, para su categorización se contempló la siguiente clasificación: AUC = 0,5 inútil, AUC = 0,6-0,7 posiblemente útil; AUC = 0,7-0,8 aceptable; AUC = 0,8-0,9 excelente y AUC > 0,9 = excepcional<sup>7</sup>.

Por otro lado, se consideraron otros parámetros que reflejan el rendimiento de los modelos predictivos: exactitud (*accuracy*), sensibilidad (*recall*), especificidad, valor predictivo positivo (precisión).

Para la evaluación de la efectividad de los modelos predictivos frente a otros métodos, contemplamos como alternativas el uso de instrumentos como escalas o puntajes y la comparación con métodos estadísticos tradicionales, ya sea regresión lineal o logística multivariable. Estos métodos de análisis estadísticos mencionados se usan más típicamente para generar modelos clínicos predictivos o pronósticos y su uso puede considerarse como un indicador de rendimiento de referencia. Cabe aclarar que cualquier tipo de algoritmo más avanzado puede contemplarse como una forma de ML.

### Resultados

Se identificaron un total de 2.321 títulos, de los cuales 763 eran duplicados. Se realizó el cribado (*screening*) de 1.558 títulos, de los cuales 22 fueron elegidos para su lectura completa<sup>8-29</sup>. Se excluyeron 18 artículos de acuerdo con los criterios de selección propuestos<sup>9-17,20-28</sup>. Por último, 4 artículos fueron elegidos para la siguiente revisión<sup>8,18,19,29</sup>. Además, se recuperaron 8 publicaciones desde otras fuentes (Biblioteca de la Asociación Argentina de Ortopedia y Traumatología y búsqueda manual de citas o *snowballing*)<sup>30-37</sup>. La *figura 1* presenta el diagrama de flujo PRISMA.

Todos los estudios incluidos describen el desarrollo y la validación interna de modelos predictivos basados en el uso de IA para la predicción de complicaciones en cirugía de columna toracolumbar por enfermedad degenerativa. No se encontraron estudios que lleven a cabo la validación externa de modelos predictivos previamente desarrollados.

De acuerdo al tipo de enfermedad degenerativa, 7 publicaciones (58,3%) incluyen pacientes con escoliosis del adulto<sup>30-34,37</sup>; 4 (33%) incluyen pacientes con enfermedad degenerativa en general (no escoliosis)<sup>18,19,29,36</sup> y una (8,33%) pacientes con espondilolistesis degenerativa, exclusivamente<sup>8</sup>.

Si bien todas las publicaciones evalúan complicaciones como resultado primario, la variable «complicaciones» regis-

tra diferentes definiciones en todas las publicaciones. En 5 artículos (41,6%) se evalúan complicaciones perioperatorias como desenlace primario, incluyendo clínicas y quirúrgicas, sin consenso en la definición<sup>8,31,37</sup>. En 2 artículos (16,6%) se contempla infección de sitio quirúrgico<sup>19,29</sup>, en 2 (16,6%) la cifosis o falla de la unión proximal<sup>32,33</sup>, en uno (8,33%) complicaciones mecánicas agrupadas (falla de unión proximal, cifosis de unión proximal, complicaciones del implante, rotura de barra)<sup>30</sup>, en otro (8,33%) pseudoartrosis<sup>34</sup> y en otro (8,33%) trombosis venosa profunda/tromboembolia de pulmón<sup>18</sup>. En la *tabla 2* se resumen las características de los estudios incluidos.

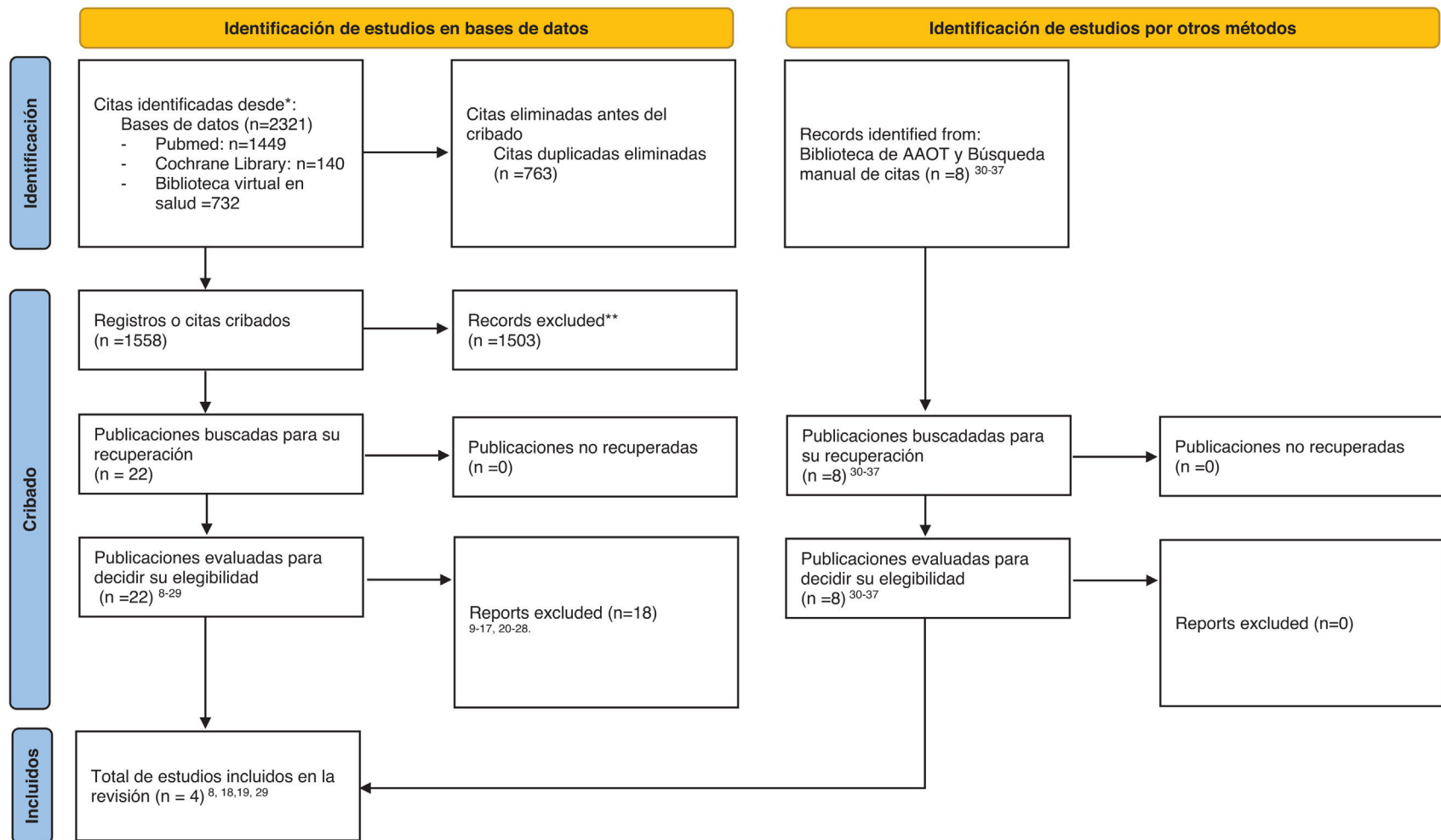
Las medidas comúnmente utilizadas para evaluar el rendimiento de los modelos predictivos fueron el área bajo la curva (n = 12; 100%) y la exactitud (*accuracy*) del modelo (n = 7; 58,3%). En menor medida, fueron reportados sensibilidad (*recall*; n = 4; 33%, especificidad n = 3; 25%) y, rara vez, el valor predictivo positivo (precisión). El rendimiento de los modelos predictivos fue variable, según el resultado contemplado (complicaciones generales versus específicas) y el tipo modelo de aprendizaje mecánico utilizado. Considerando el modelo con mejor rendimiento de cada publicación, el área bajo la curva (AUC) osciló entre 0,6 y 1,0; fue excelente o excepcional (AUC > 0,8) en más de la mitad de las publicaciones (n = 7; 58,3%)<sup>19,29-34</sup>. En las otras 5 publicaciones el rendimiento según el AUC fue aceptable (AUC = 0,7-0,8) en al menos una de las variables de resultado analizadas<sup>8,18,35-37</sup>. La mitad de los estudios no reportaron la precisión (IC95%) del AUC estimada. Los resultados de los estudios se describen en la *tabla 3*.

### Efectividad frente a otros métodos predictivos

En 5 publicaciones (41,6%) se comparó la efectividad de modelos predictivos de IA para la predicción de complicaciones generales o específicas<sup>8,18,19,30,37</sup>.

Kim et al. compararon el rendimiento del algoritmo predictivo de aprendizaje mecánico basado en una red neuronal artificial (ANN) con la regresión logística y la escala de la American Society of Anesthesiologists (ASA) de evaluación preanestésica para la predicción de 3 variables de resultados (complicaciones cardíacas, trombosis venosa profunda/tromboembolia de pulmón/complicaciones de la herida); el desempeño según el AUC del algoritmo predictivo de IA fue superior en 2 de los 3 resultados estimados a la regresión logística (excepto en la predicción de trombosis venosa profunda/tromboembolia de pulmón) y en todos con respecto a la escala ASA. Adicionalmente, la sensibilidad de ANN fue mayor que la de la regresión logística en la predicción de complicaciones de heridas<sup>37</sup>: [ANN AUC: complicaciones cardíacas 0,768 (IC95%: 0,76-0,77); TVP/TEP: 0,542 (IC95%: 0,53-0,55); complicaciones de la herida 0,606 (IC95%: 0,60-0,61). Regresión logística AUC: complicaciones cardíacas 0,690 (IC95%: 0,68-0,69); TVP/TEP: 0,547 (IC95%: 0,54-0,55); complicaciones de la herida 0,575 (IC95%: 0,56-0,58); ASA AUC: complicaciones cardíacas 0,469 (IC95%: 0,46-0,47); TVP/TEP: 0,485 (IC95%: 0,47-0,49); complicaciones de la herida 0,508 (IC95%: 0,50-0,51)].

En la publicación de Wang et al. sobre la predicción de trombosis venosa profunda/tromboembolia de pulmón, el AUC del modelo predictivo (0,716; IC95%: 0,701-0,731)



\*Consider, if feasible to do so, reporting the number of records identified from each database or register searched (rather than the total number across all databases/registers).

\*\*If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools.

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: <http://www.prisma-statement.org/>

**Figura 1** Diagrama de flujo según PRISMA 2020. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. Para más información: <http://www.prisma-statement.org/>.

**Tabla 2** Characteristics of the studies

| Author (year); institutions, country.                      | Pathology | Selection criteria                                                                                                                                                                                                                                                                                                                                                                                                  | Machine learning Algorithm                | Demographic data                                                                                                                         | Follow-up                                   | Sample split % training:validation | Outcome                                                       | Funding and conflict of interest |
|------------------------------------------------------------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------|------------------------------------|---------------------------------------------------------------|----------------------------------|
| Kim et al. (2018) <sup>37</sup><br>Multicenter study; EEUU | ASD       | <i>Inclusion:</i> Patients aged over 18 years undergoing ASD surgery.<br><i>Exclusion:</i> Patients with missing preoperative data, emergency cases, class 2, 3 or 4 wounds, open wounds on the body, sepsis, pneumonia, previous surgeries within 30 days, cardiopulmonary resuscitation before surgery, or spinal neoplasm.                                                                                       | LR y ANN                                  | Sample: 5794 - M: 2376 (41%) - F: 3418 (59%) Age; average 59,5 (DE: NR)                                                                  | 2010 - 2014                                 | 70:30                              | Complications:<br>- Cardiac complications<br>- PE/DVT - Wound | No                               |
| Noh et al. (2023) <sup>30</sup><br>Single center; Korea    | ASD       | <i>Inclusion:</i> Spine surgery for ASD and one or more radiological criteria (Coronal Cobb angle greater than 20°; sagittal vertical axis greater than 5 cm; pelvic tilt greater than 25°; TK > 60°; PI-LL > 10°; fixation of at least 4 levels); Follow-up for a period of 2 years or more.<br><i>Exclusion:</i> Syndromic deformity, autoimmune disease, infection, tumor, or any other pathological conditions. | LR; Gradient boosting; Random forest; ANN | Sample: 238 - M: 34 (14%) - F: 204 (86%) Age; average: NR (training set: 67,8 ± 7,49; validation set: 66,94 ± 6,98 years old)            | 2009-2017; Follow-up >2 years               | 70:30                              | Mechanical complications                                      | No                               |
| Yagi et al. (2018) <sup>33</sup><br>Single center; Japan   | ASD       | <i>Inclusion:</i> ASD patients aged ≥ 50 years, meeting radiological criteria (Cobb angle ≥ 20°; C7 SVA ≥ 5 cm; PT ≥ 25°), with fusion of ≥ 5 levels, and minimum follow-up of ≥ 2 years.<br><i>Exclusion:</i> Poor quality radiographs; syndromic, neuromuscular or other spinal pathologies.                                                                                                                      | DNDT; Tobulida Decision-making Tree C5.0  | Sample n = 145 Sex and age NR. Grupo Entrenamiento: n = 112 sexo M:F (5:107); edad (63.9 ± 9.4). Grupo Validación: n = 33 Edad y sexo NR | Periodo de estudio: NR; Seguimiento: 2 años | 70:30                              | PJK/PJF                                                       | NO                               |
| Scheer et al. (2016) <sup>32</sup><br>Multicenter; EEUU    | ASD       | <i>Inclusion:</i> Patients aged over 18 years old; Radiological criteria: coronal Cobb angle ≥ 20°; C7 SVA ≥ 5 cm; PT ≥ 25°; and/or thoracic kyphosis greater than or equal to 60°; Fusion of 4 or more levels was required; A minimum follow-up period of 2 years was required.<br><i>Exclusion:</i> Patients with neuromuscular deformity, infection or malignancy were excluded from the study.                  | DNDT; Decision-making Tree C5.0           | Sample: 510; Sex F:M (396:114); Age, average 57,2 ± 13.9 years old.                                                                      | Period: NR; Follow-up: 2 years              | 70:30                              | PJK/PJF                                                       | Yes*                             |

**Tabla 2** (continuación)

| Author (year); institutions, country.                                                          | Pathology | Selection criteria                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Machine learning Algorithm                                                                                                                                                                                | Demographic data                                                                                                                                                             | Follow-up                      | Sample split % training:validation   | Outcome                 | Funding and conflict of interest |
|------------------------------------------------------------------------------------------------|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|--------------------------------------|-------------------------|----------------------------------|
| Scheer et al. (2018) <sup>34</sup><br>Multicenter; EEUU.                                       | ASD       | <i>Inclusion:</i> Participants aged over 18 years old<br>Radiological criteria: Cobb angle $\geq 20^\circ$ ; C7 SVA $\geq 5$ cm; PT $\geq 25^\circ$ ; and/or thoracic kyphosis greater than or equal to 60 degrees.<br>Fusion of 4 or more levels was required.<br>A minimum follow-up period of 2 years was required.<br><i>Exclusion:</i> neuromuscular deformities, infections, and malignancies. Revision surgery is indicated only if there are reasons other than pseudoarthrosis. | DNDT; Decision-making Tree C5.1                                                                                                                                                                           | Sample: 336; F:M = 268:68; Age, average $57,7 \pm 15,1$ years old.                                                                                                           | Period: NR; Follow-up: 2 years | Validation set n = 126 (randomized). | Pseudoarthrosis         | Yes*                             |
| Pellisé et al. (2019) <sup>35</sup><br>Multicenter: España, EEUU, Switzerland, Turkey, France. | ASD       | <i>Inclusión:</i> Age >18 años Radiological criteria: Cobb coronal $\geq 20^\circ$ ; SVA $\geq 5$ cm; PT $\geq 25^\circ$ ; and/or thoracic kyphosis greater than or equal to 60 degrees.<br><i>Exclusión:</i> NR                                                                                                                                                                                                                                                                         | Random forest                                                                                                                                                                                             | Sample n = 1612; F:M NR; Age, average NR.; Training (n = 1289; F:M 1000:289; Age, average $56,5 \pm 17,3$ ); Validation (n = 323; F:M 235:88; Age, average $57,6 \pm 17,8$ ) | 2008-2016; Follow-up 730 days  | 80-20                                | Major Complication      | Yes*                             |
| Author (year); institutions, country.                                                          | Pathology | Selection criteria                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Machine learning Algorithm                                                                                                                                                                                | Demographic data                                                                                                                                                             | Follow-up                      | Sample split % training:validation   | Outcome                 | Funding and conflict of interest |
| Xiong (2022) <sup>29</sup><br>Single center; China.                                            | DSD       | <i>Inclusion:</i> Patients aged 18 years or older with degenerative lumbar disease which includes herniated disc, lumbar stenosis, spondylolisthesis, or instability and have undergone posterior lumbar interbody fusion (at least one level).<br><i>Exclusion:</i> history of spine surgery, active infection or tumor, and deformity.                                                                                                                                                 | Boosted Classification Trees, Boosted Logistic Regression, Extreme Gradient Boosting, Stochastic Gradient Boosting, Generalized Linear Model, AdaBoost AdaBoost Classification Trees*, and Random Forest. | Sample: 584; F:M 321:263; Age, average $58,36 \pm 13,76$ years old; Disc herniation: 284; Lumbar stenosis: 137; spondylolisthesis/instability: 163.                          | 2019-2021 Follow-up: 90 days.  | 50:50                                | Surgical site infection | No                               |

**Tabla 2** (continuación)

| Author (year); institutions, country.                                      | Pathology | Selection criteria                                                                                                                                                                                                                                                         | Machine learning Algorithm                                                                                                  | Demographic data                                                                                                                                                                       | Follow-up                                        | Sample split % training:validation | Outcome                                            | Funding and conflict of interest |
|----------------------------------------------------------------------------|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|------------------------------------|----------------------------------------------------|----------------------------------|
| Fatima (2020) <sup>8</sup><br>Multicenter study; EEUU.                     | DSD       | <i>Inclusión:</i> Cirugía de descompresión, artrodesis o instrumentación de la columna lumbar; espondilolistesis degenerativa lumbar; Operados entre 2005 y 2016; por neurocirugía o traumatología, bajo anestesia general y pacientes internados.<br><i>Exclusión:</i> NR | LR y LASSO: least absolute shrinkage and selection operator                                                                 | Sample: 80610; Age, median 58 years old (range:18-89); F:M 38874:41654.                                                                                                                | 2005-2016<br>Follow-up: 30 days                  | 70:30                              | Adverse events                                     | No                               |
| Zehnder (2021) <sup>36</sup><br>Multicenter study. Switzerland, UK, Italy. | DSD       | <i>Inclusión:</i> cirugía de columna por enfermedad lumbar degenerativa; Edad 18-95 años. <i>Exclusión:</i> casos con datos perdidos.                                                                                                                                      | Shrinkage Algorithm (dfbeta method)                                                                                         | Sample: 23714; F:M 12264:11450; Age, average 58,9 ± 15,7 years old.                                                                                                                    | 2012-2017<br>Follow-up until hospital discharge. | NR                                 | Surgical complications: perioperative and general. | No                               |
| Scheer (2017) <sup>31</sup><br>Multicenter study; EEUU                     | ASD       | <i>Inclusión:</i> Edad >18 años Criterios radiológicos: Cobb coronal ≥ 20°; SVA ≥ 5 cm; PT ≥ 25°; o cifosis torácica ≥ 60°. <i>Exclusión:</i> deformidad neuromuscular, infección o neoplasia maligna.                                                                     | DNNT; Decision-making Tree C5.0                                                                                             | Sample: 557<br>F:M = 439:118; Age, average 57,5 ± 15,3 years old.                                                                                                                      | Period: NR; Follow-up: 6 weeks.                  | 70:30                              | Major Complication                                 | Yes*                             |
| Wang (2021) <sup>18</sup><br>Multicenter study. EEUU                       | DSD       | <i>Inclusión:</i> fusion lumbar posterior (1 nivel). <i>Exclusión:</i> trauma, tumores, cirugía de revisión.                                                                                                                                                               | XGBoost (extreme gradient boosting)                                                                                         | Sample n = 13500<br>Age, categories n(%): 19-34 years old = 490 (3,63); 35-49 years old = 2146 (15,9); 50-65 years old = 5050 (37,41); >65 years old = 5814 (43,07).<br>F:M 7516:5984. | 2010-2017<br>Follow-up: 30 days.                 | 80:20                              | PE/DVT                                             | No                               |
| Liu (2022) <sup>19</sup><br>Single center; China                           | DSD       | <i>Inclusión:</i> enfermedad lumbar degenerativa (estenosis de canal; hernia de disco; espondilolistesis degenerativa); Cirugía via posterior única; Cirugía electiva. <i>Exclusión:</i> cirugía de urgencia.                                                              | RL, multilayer perceptron, decision tree, random forest, gradient boosting machine, and XGBoost (extreme gradient boosting) | Sample: 288; Age, average: 55,3 ± 12,3<br>F:M NR                                                                                                                                       | 2010-2019<br>Follow-up: NR                       | 70:30                              | Surgical site infection                            | Yes*                             |

Abbreviations: ASD = Adult Spinal Deformity; ANN = Artificial Neural Network; DNNT = Deep Neural Decision Tree; F:M = Female:Male; NR = Not reported; SD = Standard deviation; PE/DVT = Pulmonary embolism/Deep venous thrombosis; PJK/PJF = Proximal junctional kyphosis/failure; SVA: Sagittal vertical alignment; PT = Pelvic tilt; PI = Pelvic incidence; PI-LL = Pelvic incidence minus lumbar lordosis.

\* Declare funding and/or at least 1 financial conflict of interest.

**Tabla 3** Results of the studies

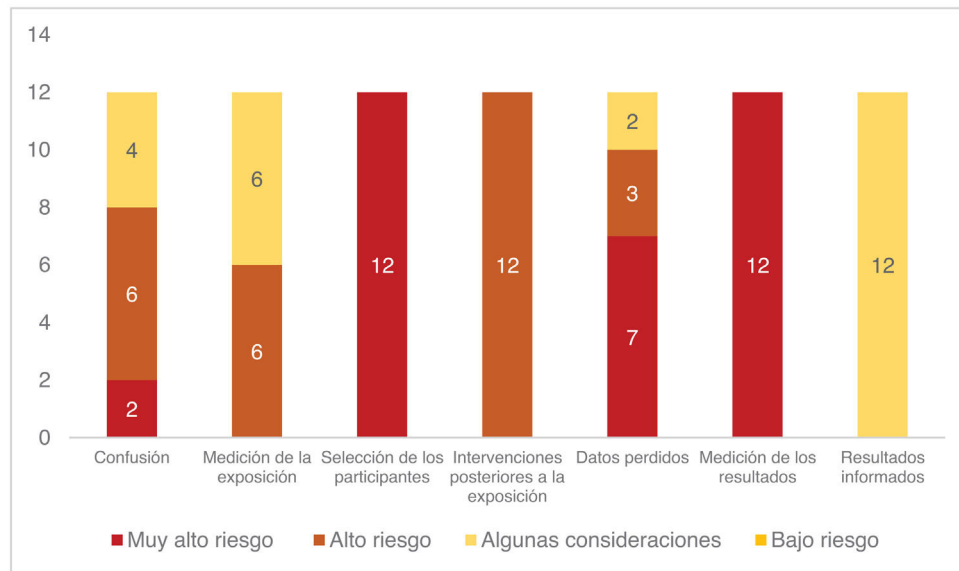
| Author (year).<br>Centers; Country.                                                                                          | Pathology | Data<br>origin | Algorithm                                       | Outcome                                                             | Model Performance [a]  |                                                                                                                        |                            |                            |                                                                                  |
|------------------------------------------------------------------------------------------------------------------------------|-----------|----------------|-------------------------------------------------|---------------------------------------------------------------------|------------------------|------------------------------------------------------------------------------------------------------------------------|----------------------------|----------------------------|----------------------------------------------------------------------------------|
|                                                                                                                              |           |                |                                                 |                                                                     | Accuracy<br>(IC95%)    | AUC-ROC<br>(IC95%)                                                                                                     | Recall<br>(IC95%)          | Specificity<br>(IC95%)     | Observations                                                                     |
| Kim et al. (2018) <sup>37</sup><br>Multicenter study;<br>United States                                                       | ASD       | NSQIP          | LR y ANN                                        | Complications:<br>- Cardiac<br>complications -<br>PE/DVT -<br>Wound | NR                     | Cardiac<br>complications<br>= 0,768<br>(0,76-0,77)<br>PE/DVT = 0,542<br>(0,53-0,55)<br>Wound<br>= 0,606<br>(0,60-0,61) | Wound<br>= 0,657(NR)       | Wound<br>= 0,587<br>(NR)   | Better results<br>with ANN<br>(Except for<br>PE/DVT).                            |
| Noh et al. (2023) <sup>30</sup><br>Single center;<br>Korea                                                                   | ASD       | RC             | LR; Gradient<br>boosting; Random<br>forest; DNN | Mechanical<br>complications                                         | 1,000<br>(1,000-1,000) | 1,000<br>(1,000-<br>1,000)                                                                                             | 1,000<br>(1,000-<br>1,000) | 1,000<br>(1,000-<br>1,000) | Better results<br>with Random<br>forest                                          |
| Yagi et al. (2018) <sup>33</sup><br>Single center;<br>Japan                                                                  | ASD       | RC             | DNDT; Tobulida<br>Decision-making Tree<br>C5.0  | PJK/PJF                                                             | 0,981 (NR)             | 1,0 (NR)                                                                                                               | NR                         | NR                         | Better results<br>including the<br>predictive<br>variable<br>«T-score ≤<br>-1.5» |
| Scheer et al. (2016) <sup>32</sup><br>Multicenter study.<br>United States                                                    | ASD       | RC             | DNDT;<br>Decision-making Tree<br>C5.0           | PJK/PJF                                                             | 0,863 (NR)             | 0,89 (NR)                                                                                                              | NR                         | NR                         | -                                                                                |
| Scheer et al. (2018) <sup>34</sup><br>Multicenter study;<br>United States.                                                   | ASD       | RC             | DNDT;<br>Decision-making Tree<br>C5.1           | Pseudoarthrosis                                                     | 0,876 (NR)             | 0,89 (NR)                                                                                                              | NR                         | NR                         | -                                                                                |
| Pellisé et al. (2019) <sup>35</sup><br>Multicenter study;<br>Spain, United<br>States,<br>Switzerland,<br>Turkey, and France. | ASD       | RC             | Random forest                                   | Major<br>Complications                                              | NR                     | 0,717<br>(0,68-0,75)                                                                                                   | NR                         | NR                         | -                                                                                |

**Tabla 3** (continuación)

| Author (year).<br>Centers; Country.                                              | Pathology | Data<br>origin             | Algorithm                                                                                                                                                                                                                                  | Outcome                                                     | Model performance [a] |                                                                                                                       |                   |                        |                                                            |
|----------------------------------------------------------------------------------|-----------|----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|-----------------------|-----------------------------------------------------------------------------------------------------------------------|-------------------|------------------------|------------------------------------------------------------|
|                                                                                  |           |                            |                                                                                                                                                                                                                                            |                                                             | Accuracy<br>(IC95%)   | AUC-ROC<br>(IC95%)                                                                                                    | Recall<br>(IC95%) | Specificity<br>(IC95%) | Observations                                               |
| Xiong (2022) <sup>29</sup> Single<br>center; China.                              | DSD       | RC                         | Boosted Classification<br>Trees, Boosted<br>Logistic Regression,<br>Extreme Gradient<br>Boosting, Stochastic<br>Gradient Boosting,<br>Generalized Linear<br>Model, <u>AdaBoost</u><br><u>Classification Trees*</u> ,<br>and Random Forest. | Surgical site<br>infection                                  | 0,8247 (NR)           | 0,906 (NR)                                                                                                            | 0,9375<br>(NR)    | 0,818<br>(NR)          | Better results<br>with AdaBoost<br>Classification<br>Tress |
| Fatima (2020) <sup>8</sup><br>Multicenter study;<br>EEUU.                        | ESD       | NSQIP                      | LR y LASSO: least<br>absolute shrinkage<br>and selection<br>operator                                                                                                                                                                       | Advetrse<br>events                                          | NR                    | General:<br>0.70<br>(0.62-0.74);<br>Surgical<br>complications<br>0,70 (NR);<br>Clinical<br>complications<br>0,70 (NR) | NR                | NR                     | Better results<br>with LR                                  |
| Zehnder (2021) <sup>36</sup><br>Multicenter study.<br>Switzerland, UK,<br>Italy. | DSD       | EUROSPINE<br>Spine<br>Tang | Shrinkage Algorithm<br>(dfbeta method)                                                                                                                                                                                                     | Surgical<br>complications:<br>perioperative<br>and general. | NR                    | Generales<br>0.74<br>(0.72-0.76);<br>Quirúrgicas<br>0.64<br>(0.62-0.65).                                              | NR                | NR                     | -                                                          |
| Scheer (2017) <sup>31</sup><br>Multicenter study;<br>EEUU                        | ASD       | RC                         | DNDT;<br>Decision-making Tree<br>C5.0                                                                                                                                                                                                      | Major<br>Complication                                       | 0,876 (NR)            | 0,89 (NR)                                                                                                             | NR                | NR                     | -                                                          |
| Wang (2021) <sup>18</sup><br>Multicenter study.<br>EEUU                          | DSD       | NSQIP                      | XGBoost (extreme<br>gradient boosting)                                                                                                                                                                                                     | PE/DVT                                                      | NR                    | 0,716<br>(0,701-<br>0,731)                                                                                            | NR                | NR                     | -                                                          |
| Liu (2022) <sup>19</sup> Single<br>center;<br>China                              | DSD       | RC                         | LR, multilayer<br>perceptron, decision<br>tree, random forest,<br>gradient boosting<br>machine, and<br>XGBoost (extreme<br>gradient boosting)                                                                                              | Surgical site<br>infection                                  | 0,860 (NR)            | 0,923 (NR)                                                                                                            | 0,834<br>(NR)     | NR                     | Better results<br>with XGBoost                             |

Abbreviations: ANN = Artificial Neural Network; ASD = Adult Spinal Deformity; AUC = Area Under the Curve; DNDT = Deep Neural Decision Tree; DNN = Deep Neural Network; DSD = Degenerative Spine disorders; LR = Logistic Regression; NR = Not reported; NSQIP = The National Surgical Quality Improvement Program; PE/DVT = Pulmonary Embolism/Deep Vein Thrombosis; PJK/PJF = Proximal Junctional Kyphosis/Proximal Junctional Failure; RC = retrospective cohort; SSIs = Surgical Site Infections.

[a] = In case of multiple predictive models, results of the best predictive model are reported.



**Figura 2** Gráfico de barras apiladas. Distribución de los artículos según dominios de la herramienta ROBINS-E para la valoración del riesgo de sesgo.

de aprendizaje mecánico fue significativamente mayor ( $p < 0,001$ ) que el AUC para el ASA y del Charlson Comorbidity Index<sup>18</sup>.

Noh et al. compararon 3 modelos de predictivos de aprendizaje mecánico (*gradient boosting*, *random forest* y *deep neural network*) con regresión logística. El modelo de IA *random forest* [AUC = 1.000 (IC95%: 1.000-1.000)] logró el mejor rendimiento predictivo<sup>30</sup>.

Fatima et al. compararon el modelo predictivo de aprendizaje mecánico (LASSO) con 2 índices de fragilidad (mFI-5 y mFI-11) y con el método de regresión logística. El rendimiento del modelo predictivo basado en IA [AUC: 0,65; IC95%: 0,61-0,69] fue menor que el de la regresión logística [AUC = 0,70; IC95%: 0,62-0,74] para la predicción general de episodios adversos y para los específicos. No obstante, el rendimiento fue significativamente mejor ( $p < 0,001$ ) que con los 2 índices de fragilidad [mFI-5 AUC = 0,50 (IC 95%: 0,47-0,53); mFI-11 AUC = 0,56 (IC95%: 0,54-0,59)]<sup>8</sup>.

Liu et al. compararon el rendimiento de 6 modelos predictivos incluyendo la regresión logística (AUC = 0,871) y determinaron que el modelo *extreme gradient boosting* tuvo el mejor rendimiento de predicción (AUC = 0,923)<sup>19</sup>.

## Riesgo de sesgo

Utilizando la herramienta Robins-E (*The Risk Of Bias In Non-randomized Studies of Exposure*) para la evaluación del riesgo de sesgo en estudios observacionales no aleatorizados, todos los artículos incluidos fueron catalogados globalmente con muy alto riesgo de sesgo, con riesgo alto o muy alto en casi todos los dominios de la herramienta (confusión, medición de la exposición, selección de los participantes, datos perdidos (fig. 2)).

Con la herramienta PROBAST (*Prediction Model Risk Of Bias Assessment Tool*) todos los estudios ( $n = 12$ ; 100%) presentaron alto riesgo de sesgo en al menos uno de los 4 dominios que conforman la escala (sesgo de selección;

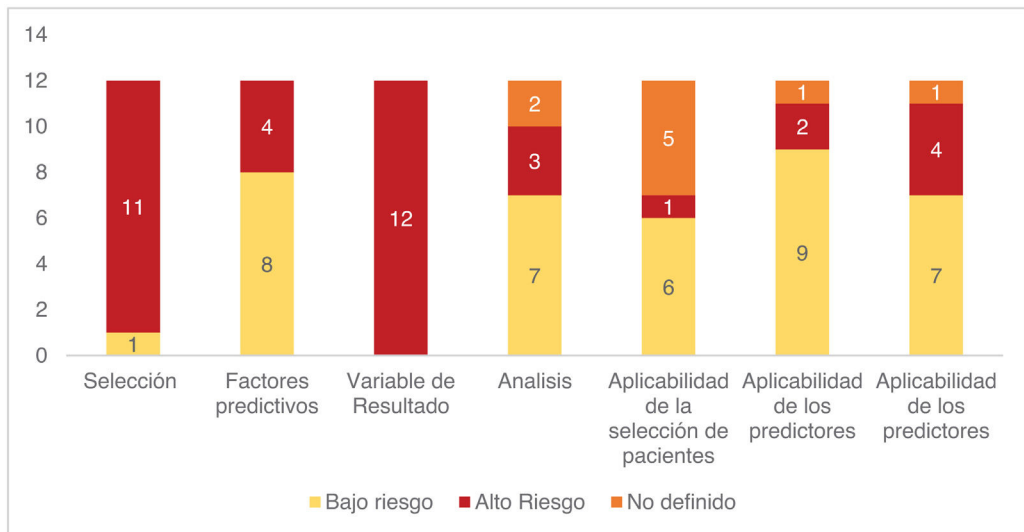
sesgo asociado con los factores predictivos; sesgo en la valoración del resultado; sesgo en el análisis). La selección de los pacientes y la valoración de la variable de resultado fueron los 2 dominios valorados más frecuentemente con alto riesgo de sesgo (fig. 3).

Dada la heterogeneidad de las muestras (cohortes o bases de datos), los resultados de interés (definición de complicaciones) y las métricas de evaluación de los algoritmos, no se realizó un metaanálisis.

## Discusión

El campo de la IA incluye diversas áreas con aplicaciones actuales o potenciales en el cuidado de la salud. Entre ellas, el ML (foco de esta revisión), el procesamiento del lenguaje natural utilizado en chatbots, la realidad aumentada, mixta y virtual, y la cirugía robótica. Estas tecnologías no solo impactan la cirugía de columna, sino también amplias áreas de la práctica médica y otras disciplinas<sup>3,4,38</sup>.

El aprendizaje mecanizado es una rama de la IA que permite a las computadoras aprender. Implica el desarrollo de algoritmos que mejoran su rendimiento con la experiencia y la incorporación de nuevos datos al sistema permite mejorar su rendimiento<sup>7</sup>. El aprendizaje mecanizado tiene una amplia gama de aplicaciones, una de ellas es el desarrollo de modelos predictivos multivariable<sup>3,4</sup>. Un modelo de predicción multivariable es una ecuación matemática que relaciona múltiples predictores (factores de riesgo, predictivos, variables independientes, covariables) para un individuo particular con la probabilidad o el riesgo de la presencia (diagnóstico) o la ocurrencia futura (pronóstico) de un resultado particular<sup>38</sup>. El desarrollo de modelos predictivos implica la selección de predictores y su combinación en un modelo multivariable. Tradicionalmente, la estimación de resultados pronósticos multivariables se basaba en técnicas estadísticas, como la regresión logística y la regresión de Cox<sup>37</sup>. El uso de técnicas de IA permite enfrentar



**Figura 3** Gráfico de barras apiladas. Distribución de los artículos según dominios de la herramienta PROBAST para la valoración del riesgo de sesgo en estudios de modelos predictivos.

un factor limitante de la metodología estadística tradicional, que es que la condición de que el poder estadístico disminuye a medida que crece la dimensión del análisis multivariado. Además, en el aprendizaje mecanizado no se propone necesariamente una hipótesis predeterminada al inicio del estudio y los algoritmos pueden correlacionar información y asociaciones, que de otro modo, podrían haberse pasado por alto o haber pasado desapercibidos debido a su complejidad y orígenes multifactoriales<sup>3</sup>.

En esta revisión, los autores nos propusimos evaluar la efectividad de los modelos predictivos basados en IA para la predicción de complicaciones en pacientes tratados mediante cirugías de columna vertebral toracolumbar degenerativa. Como resultado, no encontramos evidencias robustas en favor del rendimiento de algoritmos basados en IA frente a otros métodos predictivos tradicionales. Predominaron los estudios de desarrollo y validación interna de modelos predictivos con buen rendimiento según el AUC, que osciló mayoritariamente entre aceptable y excelente. No obstante, solo 5 (41%) estudios compararon su rendimiento con técnicas estadísticas tradicionales o con escalas o sistemas de puntajes<sup>8,18,19,30,37</sup>.

La evidencia fue débil, debido al alto riesgo de sesgo en todos de los estudios, con predominio de los sesgos en la evaluación de la variable de resultado y la selección de los pacientes. En las publicaciones recuperadas existe una heterogeneidad en la definición de la variable de resultado «complicaciones» que impide sintetizar los datos y guiar una recomendación. En ocasiones, la definición de complicación perioperatoria incluyó las acontecidas durante el intraoperatorio y el postoperatorio inmediato, lo cual, según criterio de los investigadores es una debilidad, ya que pueden estar condicionadas por diferentes variables de riesgo y agruparlas adhiere la posibilidad de sesgo de confusión<sup>8,31,35,36</sup>. Por otra parte, en algunas de las publicaciones, la estimación de la complicación fue realizada en función de la información disponible en bases de datos nacionales, previamente confeccionadas para otro fin y con tiempos de seguimiento limitados (30 días)<sup>18,37</sup>.

Cabe destacar que, en una especialidad quirúrgica cuyo desempeño puede verse condicionado por el medio, la experiencia de los cirujanos-instituciones, los recursos y características del sistema de salud de cada país o región, es difícil expresar los beneficios de algoritmos predictivos de complicaciones quirúrgicas sobre muestras conformadas por cohortes retrospectivas de un solo centro, cohortes multicéntricas no representativas, bases de datos confeccionadas con otra finalidad o muestras obtenidas por técnicas de muestreo no probabilístico sujetas al sesgo de selección. Por otro lado, podemos mencionar otras fuentes principales de sesgo en las publicaciones incluidas en esta revisión como: la falta de estudios prospectivos o muestras de casos seleccionados al azar o la ausencia de estudios de validación externa de los algoritmos predictivos que permitan estimar su rendimiento con datos ajenos a la base de datos utilizada para su desarrollo, entrenamiento y validación. Solo la mitad de los artículos publicaron los puntos estimados (ejemplo el AUC) con sus respectivos intervalos de confianza, lo que impide valorar la precisión de las estimaciones.

A pesar de lo dicho y a la evidente baja calidad de la evidencia disponible, los autores observamos una tendencia hacia el beneficio del uso de modelos predictivos basados en IA como herramienta para conocer el riesgo individual de complicaciones de la cirugía de columna en pacientes con enfermedad vertebral toracolumbar degenerativa. En un futuro cercano, estas técnicas podrían guiar la toma de decisiones de los cirujanos de columna. Estimar el riesgo quirúrgico en un paciente dado representa un verdadero desafío por la gran cantidad de variables que interactúan de forma compleja e impactan sobre el riesgo general. Variables que comprenden características que pueden generalizarse y otras que son propias del medio. Por ende, el registro de datos propios locales y regionales es la base para el desarrollo de futuros algoritmos predictivos que permitan reconocer el riesgo de nuestros pacientes con exactitud y precisión.

Esta revisión tiene como limitaciones predominantes que cierta bibliografía relevante puede no haberse recuperado

debido a que la búsqueda se hizo exclusivamente en las bases de datos MEDLINE, Cochrane Library y Lilacs. La búsqueda fue restringida a artículos en inglés, español y portugués. Además, no fue consultada la bibliografía gris. Existen consensos sobre el adecuado reporte de investigaciones de algoritmos predictivos, lo que permitiría una selección más rigurosa de los artículos para la síntesis de datos. Sin embargo, la escasez de estudios disponibles y la falta de revisiones sistemáticas previas sobre el tema llevaron a los autores de la presente revisión a adoptar criterios de elegibilidad más flexibles.

## Conclusiones

Esta revisión sistemática proporciona una visión actualizada sobre la aplicación de modelos predictivos de IA, en particular, el aprendizaje mecanizado, para la identificación del riesgo de complicaciones en pacientes tratados mediante cirugía por enfermedad degenerativa de la columna toracolumbar. Si bien la evidencia disponible es limitada y presenta un alto riesgo de sesgo, los estudios analizados indican que estos modelos pueden alcanzar un desempeño prometedor en la predicción de complicaciones, con valores del AUC que en su mayoría oscilan entre aceptables y excelentes. Son necesarias investigaciones futuras con bases de datos regionales, metodologías más robustas y validaciones externas que permitan mejorar la confiabilidad y aplicabilidad de estos modelos.

## Nivel de evidencia

Nivel de evidencia III.

## Financiación

Sin financiación externa.

## Consideraciones éticas

El siguiente trabajo es una revisión sistemática de la bibliografía, basada en datos de estudios primarios publicados, por lo cual esta exceptuada de evaluación por un comité de ética. No incluye datos primarios de pacientes ni animales.

## Conflicto de intereses

Los autores no tienen conflictos de interés que declarar.

## Agradecimientos

Los autores agradecen al Dr. Víctor Barrientos, del Hospital del Trabajador (Santiago, Chile) su ayuda con la metodología.

## Bibliografía

1. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. *Spine J.* 2008;8:8–20.
2. Andersson GB. Epidemiologic features of chronic low-back pain. *Lancet.* 1999;354:581–5.
3. Brownd SR, Park C, Donoho DA. Potential applications of artificial intelligence and machine learning in spine surgery across the continuum of care. *Int J Spine Surg.* 2023 Jun 8;8507, <http://dx.doi.org/10.14444/8507>. Epub ahead of print. PMID: 37291063.
4. Lee NJ, Lombardi JM, Lehman RA. Artificial intelligence and machine learning applications in spine surgery. *Int J Spine Surg.* 2023;16:8503, <http://dx.doi.org/10.14444/8503>. Epub ahead of print. PMID: 37193608.
5. Bero L, Chartres N, Diong J, Fabbri A, Ghersi D, Lam J, et al. The risk of bias in observational studies of exposures (ROBINS-E) tool: Concerns arising from application to observational studies of exposures. *Syst Rev.* 2018 Dic 21;7:242, <http://dx.doi.org/10.1186/s13643-018-0915-2>. PMID: 30577874; PMCID: PMC6302384.
6. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, et al., PROBAST Group. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med.* 2019 Ene 1;170:51–8, <http://dx.doi.org/10.7326/M18-1376>. PMID: 30596875.
7. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol.* 2010 Sep;5:1315–6, <http://dx.doi.org/10.1097/JTO.0b013e3181ec173d>. PMID: 20736804.
8. Fatima N, Zheng H, Massaad E, Hadzipasic M, Shankar GM, Shin JH. Development and validation of machine learning algorithms for predicting adverse events after surgery for lumbar degenerative spondylolisthesis. *World Neurosurg.* 2020 Ago;140:627–41.
9. Harada GK, Siyaji ZK, Mallow GM, Hornung AL, Hassan F, Basques BA, et al. Artificial intelligence predicts disk re-herniation following lumbar microdiscectomy: Development of the «RAD» risk profile. *Eur Spine J.* 2021 Ago;30:2167–75.
10. Karhade AV, Fogel HA, Cha TD, Hershman SH, Doorly TP, Kang JD, et al. Development of prediction models for clinically meaningful improvement in PROMIS scores after lumbar decompression. *Spine J.* 2021 Mar;21:397–404.
11. Müller D, Haschtmann D, Fekete TF, Kleinstück F, Reitmeir R, Loibl M, et al. Development of a machine-learning based model for predicting multidimensional outcome after surgery for degenerative disorders of the spine. *Eur Spine J.* 2022 Ago;31:2125–36.
12. Pedersen CF, Andersen MØ, Carreon LY, Eiskjær S. Applied machine learning for spine surgeons: Predicting outcome for patients undergoing treatment for lumbar disc herniation using PRO data. *Global Spine J.* 2022 Jun;12:866–76.
13. Ghogawala Z, Dunbar MR, Essa I. Lumbar spondylolisthesis: Modern registries and the development of artificial intelligence. *J Neurosurg Spine.* 2019 Jun 1;30:729–35.
14. Purohit G, Choudhary M, Sinha VD. Use of artificial intelligence for the development of predictive model to help in decision-making for patients with degenerative lumbar spine disease. *Asian J Neurosurg.* 2022 Jun;17:274–9.
15. Kim JS, Merrill RK, Arvind V, Kaji D, Pasik SD, Nwachukwu CC, et al. Examining the ability of artificial neural networks machine learning models to accurately predict complications following posterior lumbar spine fusion. *Spine (Phila Pa 1976).* 2018 Jun 15;43:853–60.
16. Wirries A, Geiger F, Hammad A, Oberkircher L, Blümcke I, Jabari S. Artificial intelligence facilitates decision-making in the treatment of lumbar disc herniations. *Eur Spine J.* 2021 Ago;30:2176–84.
17. Lewandrowski KU, Muraleedharan N, Eddy SA, Sobti V, Reece BD, Ramírez León JF, et al. Artificial intelligence comparison of the radiologist report with endoscopic predictors of successful transforaminal decompression for painful conditions of the lum-

- ber spine: Application of deep learning algorithm interpretation of routine lumbar magnetic resonance imaging scan. *Int J Spine Surg*. 2020 Dic;14(s3):S75–85.
18. Wang KY, Ikwuezunma I, Puvanesarajah V, Babu J, Margalit A, Raad M, et al. Using predictive modeling and supervised machine learning to identify patients at risk for venous thromboembolism following posterior lumbar fusion. *Global Spine J* [Internet]. 2021 May; 21925682211019360-21925682211019360. <https://doi.org/10.1177/21925682211019361>
  19. Liu WC, Ying H, Liao WJ, Li MP, Zhang Y, Luo K, et al. Using preoperative and intraoperative factors to predict the risk of surgical site infections after lumbar spinal surgery: A machine learning-based study. *World Neurosurg*. 2022 Jun;162:e553–60.
  20. Shah AA, Devana SK, Lee C, Bugarin A, Lord EL, Shamie AN, et al. Prediction of major complications and readmission after lumbar spinal fusion: A machine learning-driven approach. *World Neurosurg*. 2021 Ago;152:e227–34.
  21. 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*. 2022 May;2, 21925682221097650.
  22. 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*. 2023 Feb 1;38:182–91.
  23. Shamim MS, Enam SA, Qidwai U. Fuzzy Logic in neurosurgery: Predicting poor outcomes after lumbar disk surgery in 501 consecutive patients. *Surg Neurol*. 2009 Dic;72:565–72, discussion 572.
  24. Staartjes VE, Stumpo V, Ricciardi L, Maldaner N, Eversdijk HA, Vieli M, et al. FUSE-ML: Development and external validation of a clinical prediction model for mid-term outcomes after lumbar spinal fusion for degenerative disease. *Eur Spine J*. 2022 Oct;31:2629–38.
  25. Dong S, Zhu Y, Yang H, Tang N, Huang G, Li J, et al. Evaluation of the predictors for unfavorable clinical outcomes of degenerative lumbar spondylolisthesis after lumbar interbody fusion using machine learning. *Front Public Health*. 2022;10:835938.
  26. Yagi M, Michikawa T, Yamamoto T, Iga T, Ogura Y, Tachibana A, et al. Development and validation of machine learning-based predictive model for clinical outcome of decompression surgery for lumbar spinal canal stenosis. *Spine J*. 2022 Nov;22:1768–77.
  27. Staartjes VE, de Wispelaere MP, Vandertop WP, Schröder ML. Deep learning-based preoperative predictive analytics for patient-reported outcomes following lumbar discectomy: Feasibility of center-specific modeling. *Spine J*. 2019 May;19:853–61.
  28. Page PS, Greenway 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 Dec 1;37:851–4.
  29. Xiong C, Zhao R, Xu J, Liang H, Zhang C, Zhao Z, et al. Construct and validate a predictive model for surgical site infection after posterior lumbar interbody fusion based on machine learning algorithm. *Comput Math Methods Med*. 2022;2022:2697841.
  30. Noh SH, Lee HS, Park GE, Ha Y, Park JY, Kuh SU, et al. Predicting mechanical complications after adult spinal deformity operation using a machine learning based on modified global alignment and proportion scoring with body mass index and bone mineral density. *Neurospine*. 2023 Mar;20:265–74.
  31. Scheer JK, Smith JS, Schwab F, Lafage V, Shaffrey CI, Bess S, et al. Development of a preoperative predictive model for major complications following adult spinal deformity surgery. *J Neurosurg Spine*. 2017 Jun;26:736–43.
  32. Scheer JK, Osorio JA, Smith JS, Schwab F, Lafage V, Hart RA, et al. Development of validated computer-based preoperative predictive model for proximal junction failure (PJF) or clinically significant PJK with 86% accuracy based on 510 ASD patients with 2-year follow-up. *Spine (Phila Pa 1976)*. 2016 Nov 15;41:E1328–35.
  33. Yagi M, Fujita N, Okada E, Tsuji O, Nagoshi N, Asazuma T, et al. Fine-tuning the predictive model for proximal junctional failure in surgically treated patients with adult spinal deformity. *Spine (Phila Pa 1976)*. 2018 Jun 1;43:767–73.
  34. Scheer JK, Oh T, Smith JS, Shaffrey CI, Daniels AH, Sciubba DM, et al. Development of a validated computer-based preoperative predictive model for pseudarthrosis with 91% accuracy in 336 adult spinal deformity patients. *Neurosurg Focus*. 2018 Nov 1;45:E11.
  35. Pellisé F, Serra-Burriel M, Smith JS, Haddad S, Kelly MP, Vila-Casademunt A, et al. Development and validation of risk stratification models for adult spinal deformity surgery. *J Neurosurg Spine*. 2019 Jun;28:1–13.
  36. Zehnder P, Held U, Pigott T, Luca A, Loibl M, Reitmeir R, et al. Development of a model to predict the probability of incurring a complication during spine surgery. *Eur Spine J*. 2021 May;30:1337–54.
  37. Kim JS, Arvind V, Oermann EK, Kaji D, Ranson W, Ukogu C, et al. Predicting surgical complications in patients undergoing elective adult spinal deformity procedures using machine learning. *Spine Deform*. 2018;6:762–70.
  38. Combalia A, Sánchez-Vives MV, Donegan T. Immersive virtual reality in orthopaedics-a narrative review. *Int Orthop*. 2024 Ene;48:21–30, <http://dx.doi.org/10.1007/s00264-023-05911-w>.