Invasive Physiotherapy and Musculoskeletal Medicine

eISSN: 3101-0105 invasivephysiotherapy.com

doi: 10.63360/ipmm.v2.e25

Revisión de alcance

Uso de la Inteligencia Artificial como Sistema de Apoyo a la Decisión Clínica en el Manejo del Dolor Lumbar Crónico. Una Revisión de Alcance con Componente Conceptual Aplicado

Marcos Campillos Lázaro[1] 0009-0001-6347-7290

[1] Facultad de Ciencias de la Salud, Universidad de Zaragoza, Zaragoza, España.

RESUMEN

Antecedentes y Objetivos: La alta prevalencia del dolor lumbar crónico y la creciente demanda asistencial requieren modelos de gestión en fisioterapia más eficientes. La Inteligencia Artificial (IA) se presenta como una herramienta de soporte, aunque su integración clínica plantea dudas éticas y técnicas. El objetivo principal fue revisar la literatura reciente sobre sistemas de apoyo a la decisión clínica (CDSS, por sus siglas en inglés) en fisioterapia lumbar y explicar sus implicaciones para el diseño de asistentes clínicos supervisados.

Métodos: Se realizó una revisión de alcance (scoping review) con componente conceptual aplicado mediante búsqueda estructurada en PubMed, Web of Science, Scopus, Cochrane y PEDro. Se seleccionaron estudios originales publicados en los últimos 5 años que aplicaron algoritmos de aprendizaje automático o herramientas digitales en el manejo del dolor lumbar.

Resultados: La evidencia indica que los algoritmos de aprendizaje automático poseen una alta capacidad predictiva para el triaje y el pronóstico funcional, mostrando un rendimiento equivalente o, en ocasiones, superior a las herramientas convencionales. En el ámbito terapéutico, las intervenciones digitales muestran resultados positivos en adherencia y monitorización remota, aunque su eficacia directa sobre el dolor es variable. La falta de transparencia de los algoritmos se identificó como la principal barrera de confianza.

Discusión: Los resultados sugieren que la utilidad actual de la IA reside en el filtrado de seguridad y la estratificación de pacientes, más que en la prescripción autónoma. Se discute la necesidad de implementar modelos con supervisión humana obligatoria para mitigar riesgos. Como implicación práctica derivada de la revisión, se plantea la viabilidad teórica de asistentes basados en reglas clínicas estrictas para la detección de banderas rojas y arquitecturas RAG (Retrieval-Augmented Generation) para personalizar la educación sin alucinaciones.

Conclusiones: La IA es una tecnología válida para apoyar el triaje y optimizar recursos, pero requiere validación externa y transparencia. La profesión debe liderar la integración de estos "copilotos" digitales para descargar tareas burocráticas y reforzar la alianza terapéutica, evitando la automatización no supervisada.

Palabras clave: Inteligencia Artificial; Sistemas de Apoyo a la Decisión Clínica; Fisioterapia; Dolor Lumbar Crónico; Aprendizaje Automático; Telerrehabilitación.

4.0 BY-NC-SA

Correspondencia: 869739@unizar.es

1. Introducción

El dolor lumbar (DL) persiste como una de las principales causas de años vividos con discapacidad a nivel global, un desafío de salud pública cuya prevalencia ha escalado un 60% en las últimas tres décadas[1,2]. En su forma crónica (DLC), definida como aquella con una evolución superior a tres meses, representa el subgrupo de mayor impacto clínico, económico y funcional, y constituye el foco central del presente trabajo. Esta carga no se limita al sufrimiento individual o a la restricción funcional sino que genera una presión económica considerable sobre los sistemas sanitarios[3]. En Estados Unidos, los costes directos superan los 134.500 millones de dólares[1]. En España, el DL presenta una prevalencia del 17,1% en hombres y del 24,5% en mujeres, con un coste total atribuible de casi 9.000 millones de euros (0,68% del PIB)[4]. Esto conlleva una gran carga personal, un altísimo uso de recursos sanitarios y una notable pérdida de productividad laboral, un problema que se extiende a nivel continental, donde el 43% de los trabajadores europeos reportan padecer dolor de espalda[5]. Pese a esta inversión masiva, los indicadores de discapacidad mantienen una tendencia ascendente, lo que sugiere que los modelos de gestión de salud actuales requieren una revisión profunda bajo criterios de eficiencia y equidad[1,6,7].

La complejidad del abordaje del dolor lumbar crónico radica en la necesidad de articular un enfoque biopsicosocial que supere el paradigma biomédico y se adapte a la heterogeneidad de los pacientes[1,6,7]. Sin embargo, en la práctica diaria, el fisioterapeuta se enfrenta a una sobrecarga informativa (information overload)[3]. La actualización constante de la evidencia resulta inabarcable y se suman dificultades para cuantificar objetivamente la evolución del dolor o monitorizar la adherencia del paciente fuera de la consulta[2,8]. En este contexto de saturación y variabilidad clínica, la Inteligencia Artificial (IA), y específicamente el aprendizaje automático (Machine Learning, ML), emerge como una herramienta capaz de identificar patrones no lineales en grandes volúmenes de datos sin necesidad de programación manual explícita[1,9,10].

La aplicación de estas tecnologías suele materializarse en Sistemas de Apoyo a la Decisión Clínica (CDSS, por sus siglas en inglés), diseñados no para sustituir el juicio profesional, sino para organizar la información multimodal, como la historia clínica, imágenes, variables psicosociales, y respaldar decisiones diagnósticas o terapéuticas[3,11,12]. No obstante, la integración de la IA en fisioterapia plantea desafíos éticos y operativos, especialmente en lo referente a la supervisión de modelos que, aunque prometedores, carecen de razonamiento clínico profundo[10,13,14].

Dada la naturaleza incipiente de esta tecnología y la gran heterogeneidad de las herramientas disponibles (que abarcan desde algoritmos predictivos de aprendizaje automático hasta sistemas conversacionales basados en Modelos de Lenguaje de Gran Escala (Large Language Models, LLMs))[1,13], la literatura actual presenta una gran dispersión que dificulta la síntesis de resultados y la obtención de conclusiones firmes mediante métodos cuantitativos estrictos como el metaanálisis[15]. Por este motivo, el presente estudio adopta el diseño de una revisión de alcance (scoping review), tipología que permite mapear sistemáticamente la literatura sobre un campo emergente y heterogéneo sin restringirse a una pregunta clínica única. Este enfoque resulta idóneo para identificar tendencias, evaluar críticamente la utilidad clínica real de estas herramientas y, sobre esa base, proponer marcos conceptuales aplicados, en lugar de limitarse a una agregación estadística de datos que aún son prematuros.

El objetivo principal de este trabajo es analizar la evidencia científica reciente sobre las aplicaciones de IA y sistemas CDSS en el abordaje del dolor lumbar crónico . Más allá de realizar una síntesis descriptiva de la literatura, el propósito central es utilizar dicha evidencia como base para el diseño teórico de un marco conceptual de "copiloto clínico" totalmente novedoso, orientado a optimizar el triaje, la valoración y la prescripción de ejercicio, y que garantice la seguridad del paciente mediante la supervisión humana (Human-in-the-loop) frente a la automatización.

2. Métodos

Se ha realizado una revisión de alcance (scoping review) con un componente conceptual aplicado, con una estrategia de búsqueda sistematizada, orientada a integrar evidencia heterogénea sobre un tema tecnológico emergente. Aunque la síntesis de los resultados es narrativa, se aplicaron principios de transparencia y reproducibilidad adaptados de la declaración PRISMA 2020[17] para el proceso de identificación y selección de estudios.

2.1 Protocolo y Registro

El presente trabajo no cuenta con un protocolo registrado prospectivamente en plataformas como PROSPERO u OSF, dado su carácter de revisión de alcance en un campo emergente. No obstante, la metodología sigue un flujo de trabajo estructurado y replicable descrito a continuación, diseñado específicamente para mapear la literatura reciente y fundamentar la propuesta de intervención tecnológica, sin pretender el alcance estadístico de un metaanálisis.

2.2 Criterios de Elegibilidad

Los criterios de selección se definieron bajo el marco PICO (Población, Intervención, Comparación, Resultados):

- Población: Pacientes adultos (>18 años) diagnosticados con dolor lumbar crónico, definido clínicamente como aquel con una evolución superior a 3 meses. Excepcionalmente, se incluyeron estudios con pacientes en fase aguda únicamente cuando el objetivo principal de la herramienta de IA fuera predecir el riesgo de cronificación del dolor a largo plazo. De igual forma, se incluyeron estudios sobre dolor crónico no específicamente lumbar cuando su objetivo principal era el fenotipado o estratificación de pacientes mediante algoritmos de IA, dada la aplicabilidad directa de sus metodologías de agrupamiento al manejo del dolor lumbar crónico
- Intervención: Implementación de Sistemas de Apoyo a la Decisión Clínica (CDSS) basados en Inteligencia Artificial (IA), algoritmos de Machine Learning (aprendizaje automático) o Modelos de Lenguaje (LLMs).
- Comparación: Se incluyeron estudios que compararan la intervención con fisioterapia convencional, así como estudios de validación técnica sin grupo control clínico.
- Resultados: Se incluyeron estudios que evaluaran la efectividad de las herramientas de IA en las áreas de triaje, valoración y prescripción de ejercicio. Específicamente, se analizó un espectro heterogéneo de variables agrupadas en tres categorías: 1. Variables de precisión técnica y

predictiva (rendimiento de los algoritmos para triaje y pronóstico); 2. Variables de resultados clínicos ligadas a la valoración y tratamiento (intensidad del dolor, discapacidad funcional, calidad de vida y factores psicosociales); 3. Variables operativas y de proceso (adherencia a la prescripción de ejercicio, usabilidad del sistema y viabilidad de la implementación).

Respecto al resto de criterios de inclusión, se admitieron ensayos clínicos aleatorizados (ECA), estudios de cohortes (prospectivos y retrospectivos), estudios piloto y trabajos de validación de modelos predictivos que aportaran datos empíricos. Así mismo se limitó la búsqueda a artículos publicados en los últimos 5 años (2020-2025) en idioma inglés o español, excluyendo reportes de caso único, artículos de opinión sin datos y estudios de imagen médica oncológica no relacionados con rehabilitación.

2.3 Fuentes de Datos y Estrategia de Búsqueda

La búsqueda bibliográfica se llevó a cabo en noviembre de 2025 en cinco bases de datos internacionales: PubMed, Scopus, Web of Science, PEDro y Cochrane Library (véase estrategia detallada en Tabla S1 del Material Suplementario). Se diseñó una "Ecuación Maestra" combinando términos controlados (MeSH) y lenguaje libre mediante operadores booleanos, estructurada en tres bloques temáticos:

1. Tecnología: "Artificial Intelligence", "Machine Learning", "Deep Learning", "CDSS", "Generative AI".
2. Intervención: "Physiotherapy", "Physical Therapy", "Rehabilitation", "Telerehabilitation".
3. Patología: "Low Back Pain", "Chronic Low Back Pain". Adicionalmente, se realizó una búsqueda manual inversa en las referencias bibliográficas de los estudios seleccionados ("bola de nieve").

2.4 Selección de los Estudios incluidos en la Revisión

El proceso de selección se llevó a cabo por un revisor en dos fases secuenciales para garantizar el rigor metodológico. En una primera etapa, se realizó un cribado (screening) de los registros identificados mediante la lectura del título y el resumen, descartando aquellos que no se ajustaban a la temática de estudio. En la segunda etapa, se examinaron los textos completos de los artículos preseleccionados para verificar el cumplimiento estricto de los criterios de elegibilidad descritos anteriormente. La selección final y la extracción de datos fueron supervisadas por un segundo revisor independiente con el objetivo de verificar la consistente aplicación de los criterios de elegibilidad, asegurar la trazabilidad del proceso y garantizar que los estudios cumplían correctamente con los criterios de inclusión. Dicho revisor no figura como coautor del manuscrito por no cumplir los criterios de autoría establecidos por el ICMJE.

Cabe precisar que, en coherencia con el diseño de revisión de alcance con búsqueda estructurada, el cribado inicial no fue realizado de forma independiente por duplicado, por lo que el rol del segundo revisor se circunscribió a la supervisión del proceso y no al cálculo de concordancia interobservador.

2.5 Proceso de Extracción de Datos

La información relevante de cada estudio se extrajo y organizó en una tabla de síntesis (Tabla 1) considerando las siguientes variables: autoría/año, diseño metodológico, características sociodemográficas de la muestra (tamaño, perfil de los participantes, ocupación y otros detalles

relevantes) herramienta de IA utilizada (ej. visión artificial, aprendizaje automático), objetivo clínico (diagnóstico, pronóstico o tratamiento) y hallazgos principales.

Para la gestión y organización de los archivos de los estudios incluidos, así como para la revisión gramatical y estilística del borrador, se emplearon herramientas de asistencia basadas en IA Generativa (Google Gemini y NotebookLM de Google). En ningún caso se delegó en estas herramientas el análisis crítico de la evidencia, la interpretación de los resultados ni la toma de decisiones sobre la estructura o el contenido del manuscrito, responsabilidades que recaen exclusivamente sobre los autores. Todo el contenido fue verificado por los autores contra las fuentes originales antes de su inclusión en el trabajo

2.6 Proceso de Síntesis de Resultados

Dada la heterogeneidad clínica y tecnológica de los estudios (que incluyen desde validaciones de algoritmos hasta ensayos de telerrehabilitación), no fue posible realizar un metaanálisis. Se optó por una síntesis narrativa temática, agrupando los resultados en tres categorías principales: precisión diagnóstica y pronóstica de los modelos, efectividad terapéutica de las intervenciones respaldadas por IA y variables de proceso relacionadas con la adherencia y la usabilidad de las herramientas digitales. Esta aproximación permitió integrar evidencia cuantitativa y cualitativa para ofrecer una visión holística del estado del arte.

2.7 Variables del Estudio

Las variables de interés analizadas incluyeron:

- Capacidad diagnóstica y de valoración: precisión de los modelos para identificar patologías, clasificar fenotipos de dolor y realizar triaje inicial (medida mediante el área bajo la curva (AUC) y otras métricas de rendimiento predictivo).
- Precisión pronóstica: capacidad de la IA para anticipar la evolución del paciente y la respuesta al tratamiento (riesgo de cronificación, probabilidad de mejora funcional).
- Efectividad y adecuación del tratamiento: impacto de las intervenciones propuestas o guiadas por IA en términos de mejoras clínicas, incluyendo intensidad del dolor (Escala Visual Analógica, EVA / Numeric Rating Scale, NRS), discapacidad funcional (Oswestry Disability Index, ODI / Roland Morris Disability Questionnaire, RMDQ), calidad de vida y, cuando estuvo disponible, variables físicas y psicosociales relevantes (por ejemplo, fuerza o control motor, ansiedad, depresión o catastrofización del dolor).
- Variables de proceso y compromiso: nivel de adecuación de la propuesta tecnológica en términos de adherencia del paciente, tasas de finalización de los programas, patrones de uso (frecuencia y duración de las sesiones) y usabilidad/aceptabilidad de las herramientas digitales

2.8 Evaluación de la Calidad Metodológica/Riesgo de Sesgo

La evaluación de la calidad metodológica se adaptó al diseño específico de cada investigación. Para los ensayos clínicos aleatorizados, se evaluó la calidad metodológica mediante la escala PEDro[17] (desglose de puntuaciones en Tabla S2 del Material Suplementario). Para los estudios observacionales y de cohortes (retrospectivos y prospectivos), la calidad metodológica se eva-

luó utilizando la Escala de Newcastle-Ottawa (NOS)[18] (desglose de puntuaciones en Tabla S3 del Material Suplementario), la cual valora la selección de los grupos, la comparabilidad y la determinación de los resultados.

3. Resultados

3.1 Selección de Estudios.

El flujo de identificación y selección se ilustra en el diagrama PRISMA (Figura 1). La búsqueda inicial en las bases de datos arrojó un total de 661 registros. Tras la eliminación de 179 duplicados, se sometieron a cribado 482 artículos. De estos, se descartaron 401 tras la revisión por título y se evaluaron 81 resúmenes. Finalmente, tras la lectura a texto completo de 40 informes, se excluyeron 21 por motivos de diseño inadecuado o falta de resultados cuantitativos, resultando en una muestra final de 19 estudios incluidos en la revisión.

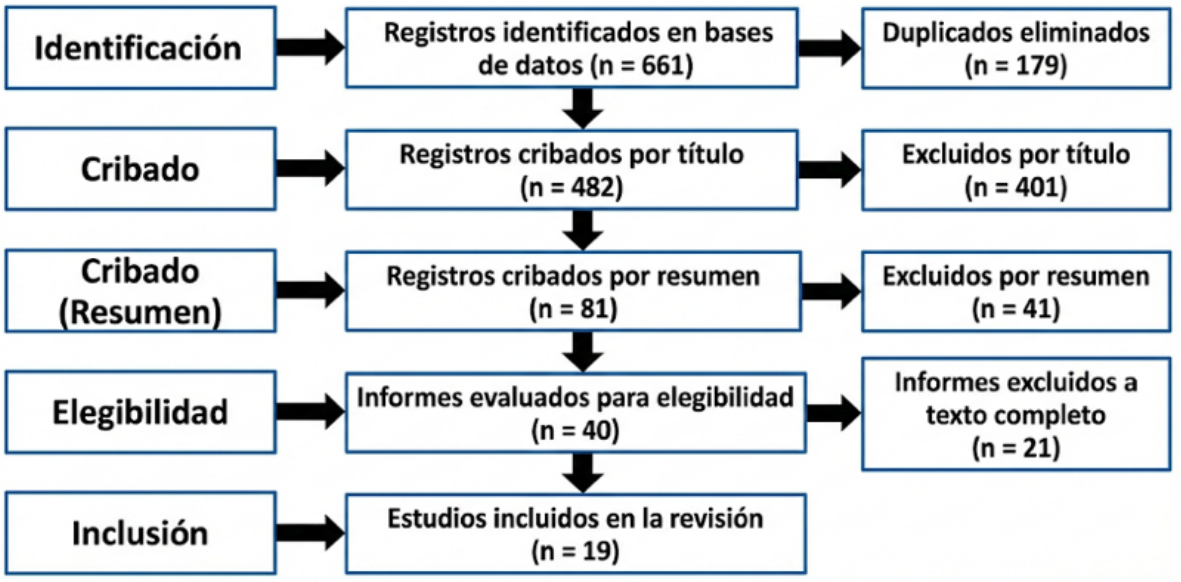

Figura 1. Diagrama de Flujo PRISMA 2020. Fuente: Elaboración propia asistida por herramientas de IA Generativa (Google Gemini)

3.2 Características de los Estudios

El análisis abarca diecinueve artículos originales publicados entre 2021 y 2025, lo que confirma el carácter emergente de esta línea de investigación. La distribución geográfica es heterogénea, con investigaciones desarrolladas principalmente en Europa, Estados Unidos y Asia. En cuanto al diseño metodológico, la evidencia se agrupa en dos grandes bloques:

- Estudios observacionales y de validación de modelos predictivos (n=12), centrados en el procesamiento de datos masivos mediante algoritmos de aprendizaje automático.
- Ensayos clínicos aleatorizados (n=7), orientados a comparar la eficacia terapéutica de las intervenciones digitales diseñadas o monitorizadas por IA frente al cuidado estándar.

La información detallada sobre el diseño, la intervención y los resultados de cada estudio se presenta en la Tabla 1.

Tabla 1: Resumen de las características y principales hallazgos de los estudios incluidos (2020-2025).

Autor Y Año	Diseño, Muestra y Perfil	Herramienta IA	Comparador	Objetivo Clínico	Resultados principales	Otros resultados
Xiao et al., 2025 [19]	ECA (n=38 aleatorizados, 34 completaron). Edad media 29 años, 76,5% mujeres. Alta escolaridad (mediana 16-17 años de educación) e IMC normal (~21 kg/m²). Dolor lumbar crónico inespecífico	Computer Vision. La aplicación Yun-fu emplea IA para identificar 30 puntos corporales clave mediante la cámara, reconociendo movimientos y proporcionando guía y retroalimentación en tiempo real durante los ejercicios	Grupo que recibió guía convencional por video	Tratamiento (Aplicación), ya que la IA se usa para dar guía y retroalimentación en tiempo real durante el ejercicio	Los pacientes tratados con IA mostraron mejoras significativas frente al control en la reducción del dolor máximo (p<0,001) y medio (p=0,01), discapacidad funcional (p=0,002), pensamientos catastrofistas (p=0,04) y grosor de la musculatura central (p<0,05) a las cuatro semanas	No hubo diferencias significativas entre los grupos respecto al dolor actual (p=0,83), la discapacidad severa (p=0,07) y la movilidad física (p=0,55) tras cuatro semanas de intervención
Cheema et al., 2025 [20]	Estudio de cohorte observacional retrospectivo (n=2497 analizados). Edad media 57 años. IMC > 30 presente en el 42.1% - 44.5% de los pacientes	Machine Learning. El algoritmo Random Forest analiza 47 variables clínicas de pacientes para predecir los factores pronósticos que determinan los resultados del tratamiento para el dolor lumbar crónico	N/A (Estudio predictivo)	Soporte a la decisión (Pronóstico), ya que la IA se usa para identificar variables pronósticas de resultados en dolor lumbar crónico	El modelo predictivo de inteligencia artificial obtuvo una precisión del 98,0% al identificar pacientes con mayores mejoras en su estado funcional, superando a los modelos de regresión lineal. Las principales variables pronósticas incluyeron el cambio funcional esperado, visitas previstas, edad y nivel educativo del profesional	No hubo diferencias significativas en la mejora clínica de los pacientes según la formación especializada del fisioterapeuta (p=0,7546) ni el historial de cirugías previas (p=0,6218) utilizando métodos estadísticos tradicionales
Griefahn et al., 2025 [21]	Análisis de cohorte retrospectiva (n=379). Edad media de 51 años, 63,6% mujeres. Ocupación: Mayoritariamente sedentaria o mixta (46,7% trabaja sentado y 43,5% alternando sentado y de pie). Pacientes con dolor espinal (zona lumbar, cervical, dorsal y glúteos)	Recommendation System. La aplicación medicalmotion usa IA para seleccionar dinámicamente ejercicios personalizados analizando la ubicación e intensidad del dolor, el historial médico y el feedback en tiempo real del paciente	N/A (Estudio de cohorte retrospectiva sin grupo control)	Tratamiento (Diseño), ya que la IA se usa para componer y personalizar programas de ejercicios para cada paciente	Los pacientes tratados con ejercicios compuestos por IA mostraron, sin disponer de un grupo control, una reducción significativa de la intensidad del dolor (p<0,001) y una mejora sustancial del bienestar (p<0,001) a las ocho semanas de intervención respecto a su evaluación inicial	No se halló una correlación estadísticamente significativa entre los ejercicios omitidos y la intensidad del dolor. Como hallazgo secundario, las sesiones medias duraron aproximadamente diez minutos
Alzouhayli et al., 2025 [22]	Subanálisis longitudinal de ECA (n=52). Edad media de 44 años, 51,9% mujeres. IMC: Presentaban un IMC medio de 28,5 kg/m² (sobrepeso). Pacientes con dolor lumbar inespecífico agudo o recurrente	Recommendation System. El dispositivo Tonal utiliza IA para medir rango de movimiento, potencia y velocidad, proporcionando al paciente feedback postural en tiempo real y ajustando la resistencia de los ejercicios	Grupo que recibió terapia guiada por IA en clínica	Tratamiento (Aplicación), ya que la IA se usa para dar retroalimentación en tiempo real y ajustar la resistencia	Los pacientes tratados con terapia guiada por IA en casa mostraron resultados clínicamente equivalentes frente al entorno clínico. Aunque estadísticamente evidenciaron menor interferencia del dolor, menor kinesiofobia (p<0,01) y mayor función física (p<0,001), la mejora al año fue similar y sin relevancia clínica entre grupos	No se observaron diferencias significativas en el cambio clínico general desde el fin de la intervención hasta la semana 52 (p≥0,23), ni respecto a la edad de los participantes (p=0,93)
Pfeifer et al., 2025 [23]	Estudio de cohorte (reanálisis de datos individuales de 2 ECA; n=1.511),. Edad media de 41 años, 57% mujeres. IMC y Ocupación: Presentaban un IMC medio de 24,8 kg/m² (normopeso rozando el sobrepeso leve) y la gran mayoría desempeñaba trabajos de "cuello blanco" u oficina (n=1.402 frente a n=109 de "cuello azul"). Pacientes con dolor lumbar inespecífico no agudo	Machine Learning. Gradient Boosting Machines y Random Forest analizan factores clínicos y psicosociales del paciente para predecir su nivel de adherencia a los programas de ejercicios para el dolor lumbar	N/A (Estudio predictivo)	Variables de gestión (Adherencia), ya que la IA se usa para identificar predictores de adherencia al ejercicio	El modelo predictivo obtuvo una precisión del 64% al 83% al clasificar la adherencia al tratamiento. Los principales factores pronósticos incluyeron las fluctuaciones en la intensidad del dolor, el control postural inicial y la reducción del nivel de ejercicio de los pacientes	No se observaron diferencias significativas en la adherencia respecto a la edad (p=0,31), sexo (p=0,26) o nivel educativo (p=0,81). La adherencia media global fue del 64%

Autor Y Año	Diseño, Muestra y Perfil	Herramienta IA	Comparador	Objetivo Clínico	Resultados principales	Otros resultados
Areias et al. (Scaling) 2024 [24]	Estudio controlado no aleatorizado (análisis post-hoc de cohortes prospectivas; n=36.186 iniciaron, 27.741 completaron). ~49 años, ~58% mujeres. Alta escolaridad (64-66% con educación universitaria o superior) y sobrepeso (IMC medio ~29,5 kg/m²). Dolor musculoesquelético crónico	Machine Learning. Un sistema de aprendizaje automático analiza el rendimiento, dolor y fatiga del paciente para recomendar al fisioterapeuta ajustes en el programa, como nuevos ejercicios o cambios de dosis	Grupo de atención digital estándar manejado manualmente	Variables de gestión (Escalabilidad), ya que la IA se usa para recomendar ajustes terapéuticos y permitir que cada fisioterapeuta atienda más pacientes	Los pacientes tratados con el programa respaldado por IA mostraron resultados equivalentes frente al control estándar. No se observaron diferencias significativas en la tasa de respuesta de mejora al dolor (p=0,399), ni en la reducción de síntomas de ansiedad (p=0,928) o depresión (p=0,187)	El grupo con IA logró una mayor adherencia terapéutica al programa (p<0,001). Específicamente, no hubo diferencias significativas en la tasa de eventos adversos (p=0,231) entre ambos grupos
Areias et al. (Tool) 2024 [25]	Estudio de pronóstico (análisis secundario de estudios prospectivos; n=6.125). 48,5 años, 56,2% mujeres. Alta escolaridad (~66% con educación universitaria o superior) y sobrepeso (IMC medio 29,5 kg/m²). Dolor lumbar (81% crónico)	Machine Learning. RNN y LightGBM analizan el movimiento, dolor y rendimiento del paciente por sesión para predecir su alivio del dolor, alertando al fisioterapeuta para ajustar el tratamiento	N/A (Estudio predictivo)	Soporte a la decisión (Pronóstico), ya que la IA se usa para predecir el alivio del dolor al final del programa.	El modelo predictivo obtuvo una capacidad de predicción que mejoró progresivamente (p<0,001), logrando un área bajo la curva de 0,71 y 0,70 en la séptima sesión para identificar el alivio del dolor. Las principales variables pronósticas incluyeron dolor, rendimiento físico, motivación y adherencia al tratamiento	No se hallaron diferencias significativas en el rendimiento predictivo entre los modelos de redes neuronales y árboles de decisión (p≥0,22). Secundariamente, la baja adherencia predecía una mayor probabilidad de fracaso
Zhao et al., 2024 [26] *	Estudio de cohorte basado en registro (n=11.995). 43,9 años, 77,6% mujeres. Escolaridad predominantemente secundaria (~58%) o terciaria (~29%). Dolor crónico complejo no maligno	Machine Learning. El algoritmo LVMM clasifica a pacientes con dolor crónico en cuatro subgrupos analizando cuestionarios clínicos y psicosociales para predecir la interferencia del dolor tras un año de tratamiento	N/A (Estudio predictivo)	Soporte a la decisión (Fenotipado), ya que la IA se usa para identificar subgrupos pronósticos en pacientes con dolor crónico	El modelo predictivo obtuvo una capacidad explicativa del 26% sobre la varianza de la interferencia del dolor a un año. Logró clasificar a los pacientes en cuatro subgrupos pronósticos distintos, mostrando diferencias estadísticamente significativas en el estado de salud general entre todas las clases identificadas (p<0,001). Los cuatro subgrupos fueron: Clase 1 (21,5%): bajo dolor pero alta carga emocional; Clase 2 (29,8%): mayor severidad clínica y psicológica; Clase 3 (30,3%): peor salud física con emociones más moderadas; Clase 4 (18,4%): subgrupo más saludable, menor dolor y mejor estado general.	No se observaron diferencias significativas en la interferencia del dolor al año de seguimiento entre el subgrupo con bajo dolor físico pero alta carga emocional (Clase 1) y el subgrupo más saludable (p=0,130)
Bates et al., 2023 [27]	ECA (n=108 aleatorizados, 69 completaron). 41,2 años, 65,2% mujeres. Sobrepeso (IMC medio 29,4 kg/m²). Dolor lumbar (excluyendo pacientes con cirugía de espalda reciente)	Machine Learning. Tonal AI analiza en tiempo real la potencia, rango de movimiento y velocidad del paciente para ajustar automáticamente la resistencia del entrenamiento y proporcionar feedback sobre su técnica	Grupo control que no recibió ningún tratamiento	Tratamiento (Aplicación), ya que la IA se usa para ajustar automáticamente la resistencia durante el ejercicio	Los pacientes tratados con IA mostraron mejoras significativas frente al control en la reducción del dolor (p≤0,05) y en la menor interferencia del dolor en actividades diarias (p≤0,03). A las ocho semanas, esta intervención logró resultados clínicos y funcionales subjetivos equivalentes a la atención terapéutica convencional.	El grupo con IA no logró mejoras significativas en la resistencia extensora isométrica (p≥0,17) ni en la capacidad de marcha (p≥0,39). Todos los participantes mantuvieron una adherencia superior al 66%.

Autor Y Año	Diseño, Muestra y Perfil	Herramienta IA	Comparador	Objetivo Clínico	Resultados principales	Otros resultados
Marcuzzi et al., 2023 [28]	ECA de 3 brazos (n=294 aleatorizados). 50,6 años, 58,8% mujeres. Alta escolaridad (61,2% con >12 años de educación) y sobrepeso (IMC medio 26,8 kg/m²). Pacientes con dolor de cuello y/o lumbar referidos a atención especializada	Case-Based Reasoning. SELFBACK utiliza razonamiento basado en casos para analizar los síntomas y pasos diarios del paciente, generando automáticamente planes semanales personalizados de ejercicios y educación basados en casos exitosos previos	Dos grupos de control: un grupo que recibió atención habitual exclusiva (examen clínico y oferta de tratamiento adecuado) y otro grupo que recibió soporte web genérico sin personalización (e-Help) junto con la atención habitual	Variables de gestión (Adherencia), ya que la IA se usa para enviar mensajes interactivos promoviendo la continuidad de los ejercicios	Los pacientes tratados con la aplicación de IA mostraron resultados equivalentes frente al cuidado habitual y al soporte web. No se observaron diferencias significativas en la mejora de la salud musculoesquelética a los tres meses respecto a la atención habitual (p=0,60) ni frente al soporte web (p=0,36)	No hubo diferencias significativas en discapacidad, dolor ni calidad de vida entre grupos. Sin embargo, el grupo con IA reportó una mejora significativamente mayor en el Efecto Global Percibido por el paciente a los 3 meses frente a la atención habitual (p=0,004). Secundariamente, el 29,3% del grupo con IA nunca accedió a la aplicación
Tun Firzara et al., 2023 [29]	ECA por conglomerados piloto (n=36 reclutados). 47,4 años, 38,9% mujeres. Alta escolaridad (el 41,7% posee un título universitario o de grado). Dolor lumbar inespecífico	Clinical Decision Support System. DeSSBack analiza automáticamente un cuestionario de nueve ítems del paciente para estratificar su nivel de riesgo y recomendar la vía de tratamiento correspondiente	Grupo manejado con la práctica clínica habitual	Soporte a la decisión (Triaje), ya que la IA se usa para estratificar el riesgo de pacientes y guiar su nivel de atención	Los pacientes tratados bajo el sistema electrónico de soporte a decisiones mostraron una mayor reducción de la discapacidad lumbar frente al control habitual, con un tamaño del efecto medio de 0,718. Secundariamente, se observaron tamaños del efecto pequeños para la reducción del dolor y depresión, mientras que la mejora en ansiedad obtuvo un tamaño del efecto medio [0,480].	La adherencia terapéutica de los pacientes al tratamiento prescrito fue deficiente. No se observaron diferencias basales significativas entre grupos respecto a la intensidad del dolor (p=0,318) o discapacidad (p=0,548)
Park et al., 2023 [30]	ECA (n=100 aleatorizados y completaron el estudio). 35,5 años, 40% mujeres. Dolor lumbar crónico	Deep Learning. Dr AI analiza mediante la cámara del móvil el movimiento del paciente para diagnosticar, prescribir ejercicios personalizados y proporcionar feedback audiovisual en tiempo real que corrige su postura	Grupo que recibió terapia física convencional	Tratamiento (Aplicación), ya que la IA se usa para dar retroalimentación audiovisual en tiempo real durante los ejercicios	Los pacientes tratados con IA mostraron resultados equivalentes frente a la fisioterapia convencional en la reducción del dolor (p=0,41) y discapacidad general (p=0,44). No obstante, la intervención con IA fue superior al mejorar la limitación en actividades cotidianas (cuestionario RMDQ, p=0,03), mostrando además ser más rentable y segura	No hubo diferencias significativas entre los grupos respecto a la mejora de la calidad de vida (p=0,18). Secundariamente, el grupo con IA reportó una significativa mayor accesibilidad (p=0,02), satisfacción con el costo (p=0,01) y con la retroalimentación en tiempo real (p=0,01)
Hartmann et al., 2023 [31]	Estudio de cohorte longitudinal (n=58 reclutados, 43 analizados). 41,4 años (grupo App) y 50,9 años (grupo Control), 59% mujeres (grupo App) y 50% mujeres (grupo Control). Dolor lumbar	Recommendation System. Medicalmotion analiza mediante IA los datos anamnésicos, el nivel de dolor y el feedback del usuario para recomendar automáticamente los ejercicios más beneficiosos entre 350 opciones.	Grupo de atención habitual con deporte de rehabilitación	Tratamiento (Diseño), ya que la IA se usa para elegir y adaptar recomendaciones de ejercicios al paciente	Los pacientes tratados con IA mostraron mejoras significativas frente al control en la reducción del dolor (p=0,003) y en la disminución de la discapacidad diaria (p<0,001) tras ocho semanas. Sin embargo, la reducción del dolor no alcanzó el umbral de relevancia clínica (estandarizado en 2 puntos en la escala NRS)	Como factores de confusión, no hubo diferencias significativas en la evolución del dolor respecto a la edad (p=0,23), sexo (p=0,65), situación laboral (p=0,22) o actividad deportiva (p=0,32)
Nordstoga et al., 2023 [32]	Análisis secundario de un ECA (n=461 aleatorizados). ~47 años, 55,3% mujeres. Alta escolaridad (entre el 58% y 72% de los subgrupos con >12 años de educación) y sobrepeso (IMC medio de ~27,6 kg/m²). Dolor lumbar inespecífico	Case-Based Reasoning. SelfBACK utiliza razonamiento basado en casos para analizar características, síntomas y progresión del paciente, adaptando semanalmente recomendaciones personalizadas de actividad física, ejercicios y educación	Grupo que recibió solo atención habitual	Tratamiento (Diseño), ya que la IA se usa para adaptar recomendaciones semanales de automanejo al paciente	El beneficio general de la IA frente al control se mantuvo independientemente de la duración o intensidad del dolor. Sin embargo, la IA fue significativamente superior al reducir la discapacidad a los tres meses específicamente en el subgrupo con dolor inicial severo (>5/10), logrando 1,8 puntos adicionales de mejora (p=0,03)	La superioridad en discapacidad observada a los tres meses no se mantuvo a los nueve meses. Tampoco hubo diferencias significativas globales en autoeficacia del dolor o mejora percibida

Autor Y Año	Diseño, Muestra y Perfil	Herramienta IA	Comparador	Objetivo Clínico	Resultados principales	Otros resultados
Strigo et al., 2023 [33] *	Estudio de cohorte retrospectivo (n=3.296 analizados). 50–51 años, ~64.6% mujeres. Obesidad (IMC medio de los subgrupos entre 30,5 y 33,0 kg/m²). Pacientes con dolor crónico complejo y refractario a intervenciones habituales.	Machine Learning. Mclust analiza nueve síntomas iniciales reportados por el paciente mediante agrupamiento no supervisado para clasificarlo en tres subgrupos, prediciendo su trayectoria y personalizando el tratamiento del dolor crónico	N/A (Estudio predictivo)	Soporte a la decisión (Fenotipado), ya que la IA se usa para identificar fenotipos de dolor crónico	El modelo predictivo identificó tres subgrupos clínicos. Todos mostraron una mejora clínica global (p=0,002) y una reducción del impacto del dolor (p<0,001) tras un año de tratamiento, aunque el subgrupo más severo evidenció una menor mejora al finalizar el seguimiento (p<0,05)	No hubo diferencias significativas entre subgrupos en la presión arterial inicial (p>0,05), la proporción de pacientes medicados con opioides (p>0,1), ni en los tipos de servicios médicos utilizados (p=1)
Zmudzki & Smeets, 2023 [34]	Estudio de cohorte retrospectivo (n=2.364 que completaron la intervención). 43,8 años, 72,5% mujeres. Escolaridad predominantemente secundaria (44,2% con educación secundaria vocacional) y sobrepeso u obesidad (65,7% con IMC ≥ 25 kg/m²). Dolor musculoesquelético crónico referidos a un programa de rehabilitación multimodal interdisciplinaria	Machine Learning. Un framework de machine learning analiza 30 variables iniciales del paciente para predecir su respuesta al tratamiento en 13 dimensiones, generando un perfil pronóstico que permite personalizar sus objetivos	N/A (Estudio predictivo)	Soporte a la decisión (Pronóstico), ya que la IA se usa para predecir resultados del tratamiento multimodal del dolor musculoesquelético crónico	El modelo predictivo obtuvo un rendimiento individual con un AUC entre 0,49 y 0,65 para diversas medidas clínicas. No obstante, al emplear el conjunto completo de algoritmos, el perfil multidimensional logró evaluar y predecir de forma consistente los resultados del tratamiento en el 75,3% de los pacientes evaluados.	El índice de masa corporal no se seleccionó como factor pronóstico en ningún modelo. Las variables predictivas universales incluyeron el control del tratamiento, el dolor máximo y la situación laboral.
Knoop et al., 2022 [35]	Estudio de cohorte prospectivo (n=247 con datos basales). 49 años, 41% mujeres. Escolaridad predominantemente baja (60% con nivel educativo bajo, correspondiente a educación primaria, secundaria o vocacional). Dolor lumbar agudo (≤1 mes de duración)	Machine Learning. XGBoost analiza tres variables del paciente (resiliencia, discapacidad previa y expectativa de recuperación) para predecir automáticamente su riesgo de no recuperarse del dolor lumbar a los tres meses	N/A (Estudio predictivo)	Soporte a la decisión (Pronóstico), ya que la IA se usa para predecir la no recuperación del dolor lumbar agudo	El modelo predictivo obtuvo un área bajo la curva de 0,66 y una precisión del 63% al predecir la falta de recuperación del dolor lumbar. Aunque no mostró superioridad frente al modelo estadístico tradicional (AUC 0,71), superó la capacidad predictiva de la evaluación clínica habitual (AUC 0,53)	El 47% de los participantes no se recuperó a los tres meses. Secundariamente, la inclusión de la evolución clínica de las primeras semanas no mejoró sustancialmente el rendimiento del modelo
Park et al., 2022 [36]	Estudio de casos y controles retrospectivo (n=281 reclutados, 176 analizados). 36,7 años (grupo App) y 38,3 años (grupo Control), 60,6% mujeres (grupo App) y 29,2% mujeres (grupo Control). Trabajadores de oficina con dolor lumbar	Deep Learning. MDMECA identifica 14 puntos corporales mediante la cámara del móvil para analizar el movimiento en tiempo real y proporcionar feedback visual y auditivo al paciente durante el ejercicio	Grupo que usó videos de ejercicios de YouTube	Tratamiento (Aplicación), ya que la IA se usa para analizar posturas y dar retroalimentación en tiempo real durante el ejercicio	Los pacientes tratados con la aplicación de IA mostraron mayores mejoras frente al control de video en el incremento de la calidad de vida (p<0,01) y lograron una mayor reducción en la intensidad del dolor lumbar (p<0,01) tras realizar el programa de ejercicios	El grupo con IA registró mayor adherencia y satisfacción (p<0,01). No hubo diferencias significativas entre ambos grupos respecto a la disposición económica a pagar por el programa (p=0,59)
Anan et al., 2021 [37]	ECA (n=121 aleatorizados, 94 analizados). ~42 años, 23,4% mujeres. Trabajadores de oficina con dolor de cuello/hombros y/o lumbar	Natural Language Processing. Secaide interactúa diariamente mediante una app móvil para enviar rutinas breves de ejercicios, adaptando automáticamente sus respuestas y mensajes de motivación según las opciones que selecciona el paciente	Rutina regular de ejercicios de la empresa	Variables de gestión (Adherencia), ya que la IA se usa para enviar mensajes diarios interactivos promoviendo la continuidad de los ejercicios	Los pacientes tratados con el programa asistido por IA mostraron mejoras significativas frente al control en la reducción del dolor lumbar, de cuello y hombros a las doce semanas. Además, lograron una tasa significativamente superior de mejora clínica subjetiva de sus síntomas (p<0,001)	Como hallazgo secundario, el grupo con IA logró una alta adherencia terapéutica al programa del 92%. Además, las sesiones de ejercicios diarios requirieron menos de un minuto para completarse

Fuente: Elaboración propia asistida por herramientas de IA generativa (NotebookLM)

*Estudio sobre dolor crónico complejo. Incluido por la relevancia metodológica de su algoritmo de fenotipado, aplicable al abordaje del dolor lumbar crónico

3.3 Evaluación Crítica

Se evaluó la calidad metodológica de los ensayos clínicos aleatorizados incluidos mediante la escala PEDro[17]. La puntuación media obtenida fue de 5,7/10 (rango 3-7) (véase el desglose detallado en la Tabla S2). La gran mayoría de los estudios (n=5) mostraron una calidad metodológica alta (≥6/10)[19,28,29,30,32], destacando por el estricto cumplimiento de la asignación aleatoria, el análisis por intención de tratar y un riguroso reporte estadístico. Por el contrario, dos ensayos presentaron un alto riesgo de sesgo (≤4/10)[27,37]. El análisis transversal revela que la principal debilidad metodológica de la literatura actual es el cegamiento de participantes y terapeutas, una dificultad inherente a las intervenciones que combinan ejercicio físico y dispositivos tecnológicos[19,27,37]. No obstante, este riesgo se mitigó parcialmente al implementar con éxito el cegamiento de los evaluadores independientes en más de la mitad de los ensayos. Por otro lado, la calidad metodológica de los 12 estudios observacionales (11 de cohortes y 1 de casos y controles) se evaluó mediante la escala de Newcastle-Ottawa (NOS)[18], presentando una puntuación media de 6,4/9 (rango 3-9) (véase el desglose detallado en la Tabla S3). Casi la mitad de la muestra (n=5) mostró una alta calidad (≥8/9)[23,24,25,26,34], exhibiendo una gran robustez en la representatividad de la cohorte y el ajuste estadístico multivariante. En un nivel intermedio, tres estudios mostraron un riesgo moderado de sesgo (6-7/9)[20,33,35]. No obstante, un tercio de las investigaciones (n=4) presentó un alto riesgo de sesgo (≤4/9)[21,22,31,36]. Los principales factores limitantes en estos estudios de menor calidad fueron la falta de comparabilidad (al no controlar factores de confusión clave como la edad o la severidad basal) y las deficiencias en el seguimiento, reportando pérdidas significativas que introducen un potencial sesgo de desgaste en la evidencia acumulada.

3.4 Síntesis de Resultados

Una característica distintiva de la literatura analizada es la multidimensionalidad de los datos procesados por las herramientas de Inteligencia Artificial. A diferencia de los modelos clínicos tradicionales, los algoritmos evaluados integraron de manera simultánea un amplio espectro de variables para sus predicciones y tratamientos. Estas se pueden clasificar en cinco grandes dominios: 1) sociodemográficos y antropométricos (ej. edad, sexo, índice de masa corporal, situación laboral y nivel de privación social); 2) clínicos (ej. intensidad y duración del dolor, uso de opioides y comorbilidades); 3) funcionales y biomecánicos (ej. índices ODI y RMDQ, fuerza muscular y rango de movimiento capturado por visión artificial); 4) psicosociales (ej. depresión, ansiedad, kinesiofobia, resiliencia y expectativas de recuperación); y 5) de interacción tecnológica, una categoría emergente que incluye la frecuencia de uso de las aplicaciones, la precisión en la ejecución de los ejercicios y el reporte de fatiga en tiempo real. Esta riqueza de variables es lo que permite a la IA realizar un fenotipado de pacientes mucho más profundo que el cribado estándar.

3.4.1 Descripción de los resultados obtenidos para cada variable

Los hallazgos se categorizaron conceptualmente en función de la utilidad clínica reportada: soporte a la decisión (pronóstico, fenotipado y triaje), tratamiento (diseño y aplicación) y variables de soporte y gestión (adherencia y escalabilidad).

Precisión Diagnóstica y Pronóstica: De los artículos revisados, 6 analizan específicamente el uso de algoritmos de aprendizaje automático para la estratificación y pronóstico de pacientes. A nivel metodológico, estos estudios se basan principalmente en diseños de cohorte (prospectivos y retrospectivos) y reanálisis de bases de datos clínicas para entrenar los algoritmos. La evidencia es mayoritariamente favorable: 5 de estos estudios sugieren que el aprendizaje automático ofrece una mayor capacidad predictiva para anticipar resultados clínicos[20,25,26,33,34]. Sin embargo, 1 de los estudios aporta un matiz importante al señalar que, si bien el aprendizaje automático supera a los cuestionarios de tamizaje estándar, su rendimiento puede ser equivalente (y no necesariamente superior) al de los modelos de regresión estadística tradicionales[35]. En términos generales, los algoritmos de aprendizaje automático mostraron, en la mayoría de los estudios revisados, una mayor capacidad que las herramientas convencionales para estratificar pacientes, aunque con variabilidad entre estudios (tales como el cuestionario validado *STarT Back*, la propia expectativa y juicio clínico del profesional):

- Algoritmos Random Forest identificaron cuartiles de mejora funcional con una precisión predictiva del 98% para los pacientes con mejores resultados, superando en rendimiento a los modelos clásicos de regresión lineal[20].
- El modelo predictivo XGBoost mostró un mejor rendimiento para anticipar el riesgo de cronificación (AUC 0,66; IC 95%: 0,56-0,70) en comparación con el uso del cuestionario *STarT Back* (AUC 0,53; IC 95%: 0,45-0,60) en pacientes evaluados en fase aguda[35].
- Redes neuronales y modelos LightGBM lograron identificar pacientes respondedores al tratamiento con valores de AUC de 0.71, permitiendo ajustes dinámicos en el plan de cuidados[25].

En resumen, el aprendizaje automático ofrece una capacidad predictiva superior o equivalente a las herramientas tradicionales para anticipar el riesgo de cronicidad

Efectividad Terapéutica (Dolor y Discapacidad):

De los artículos revisados, 8 evalúan directamente la eficacia clínica de estas intervenciones digitales frente a grupos control o tratamientos presenciales[19,21,28,30,31,32,36,37]. Cabe destacar el alto nivel de evidencia en este apartado, ya que 5 de estos estudios son ensayos clínicos aleatorizados[19,28,30,32,37], respaldados por 3 estudios observacionales (cohortes y casos-control)[21,31,36]. La evidencia refleja una tendencia mixta: 5 estudios reportan resultados favorables de la IA para reducir el dolor o la discapacidad[19,21,31,36,37], 1 estudio reporta una eficacia equivalente a la terapia tradicional[30], y 2 ensayos señalan la ausencia de beneficios adicionales globales, condicionando su éxito a la severidad inicial del paciente[28,32]. En conjunto, la eficacia clínica de las intervenciones digitales basadas en IA es variable e inconsistente, lo que subraya la necesidad de considerar el perfil clínico del paciente a la hora de seleccionar e implementar estas herramientas:

- Sistemas de visión artificial guiada reportaron reducciones del dolor estadísticamente significativas en comparación con los grupos control ($p<0,001$), con reducciones de hasta 3,00 puntos en la escala NRS[19].
- Las intervenciones digitales redujeron significativamente la discapacidad funcional lumbar, evidenciada por mejoras de hasta 3,8 puntos en el índice ODI, y aumentaron la calidad de vida de los pacientes[21,31]. Además, mostraron una eficacia clínica equivalente a la fisioterapia presencial ($p=0,44$), destacando como una alternativa significativamente más coste-efectiva[22,30].

- Sin embargo, algunos ensayos no encontraron beneficios adicionales significativos en la salud musculoesquelética global frente al cuidado habitual (reportando diferencias irrelevantes de apenas 0,62 puntos en el cuestionario MSK-HQ)[28], sugiriendo una eficacia dependiente del perfil de severidad del paciente. Por ejemplo, se ha observado que la IA logra mayores reducciones en la discapacidad cuando se aplica a pacientes con una intensidad de dolor basal alta (>5/10) en comparación con casos más leves[32]
- No obstante, en algunos estudios la magnitud de la reducción del dolor, aunque estadísticamente significativa, no alcanzó el umbral de diferencia mínima clínicamente importante (MCID) recomendado para el dolor lumbar crónico (estandarizado en una reducción de ≥2 puntos en la escala NRS). Por ejemplo, se han reportado intervenciones que logran reducciones absolutas de apenas 1,4 puntos, lo que limita el impacto práctico de estas herramientas en la mejora real percibida por el paciente[31].

En resumen, la eficacia clínica de las intervenciones basadas en IA es inconsistente; muestran un éxito variable que depende de la severidad del paciente y no siempre alcanzan la relevancia clínica

Pasando de los estudios orientados al tratamiento a los centrados en la viabilidad operativa del sistema, un tercer bloque de evidencia analiza específicamente la adherencia, la usabilidad y la escalabilidad de las herramientas digitales

Adherencia, Usabilidad y Escalabilidad: De los artículos revisados, al menos 5 evalúan de forma específica la viabilidad de la IA desde la perspectiva del comportamiento del paciente y la gestión del sistema de salud. A nivel metodológico, esta evidencia se sustenta en una combinación transversal de ensayos clínicos aleatorizados, estudios de cohorte prospectivos y el entrenamiento de modelos predictivos a partir de bases de datos clínicas. Los estudios sugieren una viabilidad prometedora de estas tecnologías en cuanto a seguridad y aceptabilidad. No obstante, se observan resultados heterogéneos en la utilización, la adherencia y el compromiso del paciente, que dependen de la herramienta concreta y del contexto de aplicación:

- Las intervenciones mediante *chatbots* lograron que el 92% de los pacientes asignados a esta tecnología (44 de 48 usuarios) completara el programa de ejercicios de 12 semanas[37].
- Los modelos de Inteligencia Artificial lograron predecir correctamente en el 83% de los casos evaluados (exactitud predictiva) qué pacientes cumplirían o no con su tratamiento de fisioterapia presencial (definiendo el cumplimiento como la asistencia a más del 70% de las sesiones de un protocolo supervisado de 3 semanas, a razón de 3 sesiones semanales), utilizando como predictores variables como el control postural, las fluctuaciones del dolor y el nivel previo de actividad física del paciente[23].
- En términos de seguridad y escalabilidad, las intervenciones digitales respaldadas por IA mostraron un perfil de riesgo bajo, con tasas de eventos adversos escasas y comparables al cuidado estándar (reportando incidencias de apenas un 0,58% en el grupo asistido por IA frente a un 0,69% en el grupo control; $p = 0,231$), sin evidenciar incrementos significativos de efectos indeseables asociados al uso de la tecnología[24].

En resumen, las herramientas digitales muestran una viabilidad prometedora y un buen perfil de seguridad, aunque con resultados heterogéneos en utilización, adherencia y compromiso según la herramienta y el contexto

3.4.2 Relación entre los resultados de las diferentes variables

El análisis cruzado revela asociaciones significativas entre las variables clínicas, la adherencia y la eficacia predictiva:

Intensidad y fluctuaciones del dolor: Algunos estudios sugieren que la eficacia del soporte digital podría depender de la severidad inicial. Pacientes con dolor alto (>5) lograron mayor reducción de discapacidad frente al control (interacción p=0,03), mientras que los casos leves no mostraron beneficios adicionales[32]. Además, las fluctuaciones constantes del dolor actuaron como el predictor principal de baja adherencia[23].

Sumado a esto, la relación entre la evolución del dolor y la adherencia no es lineal, sino que sigue una forma de "U" invertida. Los modelos predictivos sugieren que los pacientes tienden a abandonar la terapia digital tanto si experimentan aumentos marcados del dolor (por temor a que el ejercicio les esté perjudicando) como si experimentan disminuciones rápidas (por la falsa creencia de que ya están curados y no necesitan continuar)[23]

Adherencia y evolución clínica: Se halló una correlación negativa entre intensidad del dolor y frecuencia de ejercicios en la app (semana 8: p=0,004), así como entre la frecuencia de uso y la mejora del bienestar (p=0,01)[21]. Modelos predictivos (LightGBM) confirmaron que el cumplimiento y la motivación superan a las variables demográficas al predecir el alivio del dolor[25].

Factores psicosociales: Los análisis de agrupamiento sugieren que los perfiles con alta carga psiquiátrica predicen una mejora global significativamente menor a largo plazo[33]. Asimismo, una mayor carga sintomática inicial predijo una peor interferencia del dolor a un año de seguimiento (tamaño del efecto d=1,27)[26].

Complementariamente, los modelos de aprendizaje automático han revelado el peso de variables "protectoras". Se ha descubierto que una alta resiliencia y la expectativa de recuperación del paciente tienen un mayor potencial predictivo a los clásicos factores de distrés psicológico (como la depresión o la ansiedad) para identificar de forma temprana trayectorias de recuperación exitosas[35].

Feedback biomecánico y compromiso conductual: La evidencia disponible sugiere que el uso de visión artificial (Deep Learning) para guiar el movimiento no solo favorece una correcta ejecución técnica (mejorando el grosor de la musculatura central), sino que el feedback audiovisual en tiempo real actúa como un motor motivacional clave. Esta tecnología complementa la supervisión profesional logrando que los pacientes se sientan seguros y guiados, lo cual correlaciona directamente con mayores tasas de finalización de los programas en casa frente al uso de videos convencionales[19,36].

Escalabilidad del profesional y calidad percibida: Se ha comprobado que el uso de herramientas de soporte clínico basadas en IA permite aumentar drásticamente la capacidad de atención (pasando, por ejemplo, de un ratio de 1:57 a 1:129 pacientes por fisioterapeuta) sin comprometer los resultados clínicos en la reducción del dolor o la salud mental. Al contrario, la automatización del triaje y la planificación optimizan el tiempo del clínico, manteniendo niveles de adherencia y satisfacción del paciente idénticos o incluso superiores[24].

4. Discusión

El análisis de la literatura reciente (2020-2025) revela un salto cualitativo en la aplicación de la IA en fisioterapia. Si bien las revisiones previas situaban a la tecnología como un mero observador diagnóstico, la evidencia actual sugiere que los sistemas han integrado la valoración, el diagnóstico y el pronóstico para transitar hacia un rol de "agente terapéutico activo". Por un lado, los algoritmos de aprendizaje automático han perfeccionado su capacidad pronóstica, prediciendo la respuesta al tratamiento, el riesgo de cronicidad y la adherencia del paciente con gran exactitud[23,26,34,35]. Por otro lado, las herramientas basadas en visión artificial permiten una valoración cinemática objetiva en tiempo real, facilitando la clasificación en subgrupos clínicos y la emisión de diagnósticos funcionales[19,30]. Es precisamente sobre esta base clínica integral que los sistemas actuales logran intervenir directamente mediante la prescripción automatizada de ejercicio y la educación[19,21,30]. No obstante, se observa una paradoja relevante: aunque la precisión matemática de los modelos es alta, su traducción a resultados clínicos (dolor y función) es claramente heterogénea, con ensayos que muestran resultados más favorables de las intervenciones respaldadas por IA frente al tratamiento convencional[19,20,31,37], otros en los que únicamente se alcanza la no inferioridad frente al cuidado estándar[28] y estudios en los que las diferencias clínicas apenas superan, o incluso no alcanzan, la magnitud considerada mínimamente importante para el paciente[24,31].

Una inconsistencia crítica identificada en la literatura es la tensión entre la complejidad del algoritmo y su utilidad clínica. Se ha detectado que la "opacidad algorítmica" o fenómeno de "Caja Negra" constituye la principal barrera para la implementación[10,34]. La incapacidad de los modelos de Deep Learning para explicar el razonamiento y la lógica detrás de una recomendación clínica genera incertidumbre y erosiona la confianza del fisioterapeuta[14,24,34]. Esto sugiere que la adopción masiva no dependerá tanto de aumentar la potencia de cálculo, sino de transitar hacia una "IA Explicable" (XAI) que permita al profesional validar la lógica de la decisión[11,14,24]. A esto se suma la fricción tecnológica, ya que aunque la adherencia es alta en condiciones experimentales, en la práctica real existen barreras operativas (fallos técnicos, falta de interoperabilidad) que interrumpen el flujo de trabajo clínico[29].

Los hallazgos refuerzan la necesidad de un modelo de "Inteligencia Aumentada" frente a la automatización total. Aunque la IA muestra potencial en el procesamiento de datos y triaje (detección de banderas rojas para derivación médica y estratificación del paciente del riesgo de cronicidad para asignar el nivel de tratamiento adecuado)[3,20], carece de las capacidades afectivas necesarias para la alianza terapéutica[13,38]. Por tanto, la tecnología se perfila teóricamente no como un sustituto, sino como un soporte que libera al clínico de la carga cognitiva y burocrática, permitiéndole centrarse en los aspectos humanos de la rehabilitación[8,24].

Es necesario reconocer las limitaciones inherentes a esta revisión de alcance. La selección de estudios, aunque sistematizada, está sujeta a la interpretación cualitativa de los autores y no ofrece estimación cuantitativa agregada de un metaanálisis. Sin embargo, esta flexibilidad metodológica se justifica por la naturaleza emergente y heterogénea de las intervenciones tecnológicas evaluadas, que dificulta una síntesis estadística directa[1,2]. Asimismo, debe considerarse una limitación de validez externa en la evidencia primaria (generación limitada), ya que gran parte de los algoritmos han sido entrenados con datos de poblaciones muy específicas (como

seguros privados o alta alfabetización digital), lo que podría perpetuar sesgos de atención y dificultar la generalización de los resultados a sistemas de salud pública con recursos limitados o poblaciones geriátricas más vulnerables[19,20,25,27].

Otra limitación a considerar es la propia naturaleza de la tecnología evaluada. Dado que la Inteligencia Artificial es un campo en desarrollo exponencial, es posible que los modelos y algoritmos analizados evolucionen a un ritmo más rápido que los ciclos de publicación científica. Por tanto, existe una obsolescencia temporal inherente a este tipo de revisiones, ya que las herramientas actuales podrían verse superadas rápidamente por sistemas más avanzados; lo que refuerza la necesidad de establecer marcos conceptuales de integración clínica que perduren más allá de los algoritmos específicos.

Aunque la evidencia disponible sitúa a la IA como una herramienta prometedora para el triaje y el pronóstico, su impacto sobre resultados clínicos como el dolor y la discapacidad sigue siendo heterogéneo, y no todas las intervenciones alcanzan de forma consistente la diferencia mínima clínicamente importante recomendada para el dolor lumbar crónico[24,28]. Esta brecha sugiere que, para mejorar la efectividad, no basta con optimizar los algoritmos, sino que es necesario intervenir sobre varias barreras estructurales. En el plano profesional, destacan las barreras formativas, ya que muchos fisioterapeutas carecen de competencias en alfabetización digital y en interpretación crítica de modelos de IA, lo que limita su capacidad para utilizar estas herramientas sin delegar su criterio clínico[11,24]. A nivel de sistemas, persisten barreras tecnológicas como la falta de interoperabilidad con las historias clínicas electrónicas, la inestabilidad de algunas soluciones y la necesidad de infraestructura específica, que añaden fricción al flujo de trabajo y reducen la adopción fuera de los entornos experimentales[3,11]. Finalmente, las barreras éticas y legales derivadas de la opacidad algorítmica y de la incertidumbre sobre la responsabilidad profesional dificultan que los clínicos confíen en recomendaciones generadas por sistemas de "caja negra"[10,34]. Superar estas barreras tendría implicaciones económicas sustanciales para los sistemas sanitarios.

Desde la perspectiva del coste-efectividad, algunos de los estudios incluidos aportan datos relevantes: las intervenciones de fisioterapia basadas en visión artificial lograron resultados clínicos equivalentes a la terapia presencial convencional mostrando ser significativamente más rentables[30], y los programas de adherencia asistidos por chatbot alcanzaron tasas de finalización del 92% con sesiones de ejercicio que requerían menos de un minuto diario[37], lo que sugiere un potencial de escalabilidad asistencial sin un incremento proporcional de costes. Además, los modelos de aprendizaje automático predictivo lograron identificar correctamente en el 83% de los casos evaluados qué pacientes no cumplirían con el tratamiento presencial antes de que este fracasara[23], lo que permitiría orientar recursos hacia intervenciones digitales más eficientes en aquellos pacientes de menor adherencia. En el contexto español, donde el DLC genera un coste total estimado de casi 9.000 millones de euros anuales[4], incluso una reducción modesta de las visitas redundantes o una mejora de la adherencia mediante herramientas digitales de bajo coste representaría un ahorro asistencial potencialmente relevante. No obstante, estos beneficios exigen una inversión inicial en formación del profesional, infraestructura tecnológica interoperable y procesos de evaluación regulatoria rigurosa, evitando que la tecnología amplíe inequidades existentes o traslade el riesgo del fallo al clínico o al paciente[10,24].

Basándonos en los vacíos identificados, específicamente la necesidad de seguridad clínica y personalización escalable[13,24], se propone la viabilidad conceptual de sistemas tipo "Copiloto Clínico" bajo arquitectura Human-in-the-loop. A diferencia de las herramientas totalmente autónomas, este diseño teórico integra dos módulos secuenciales. Primero, un filtro estricto basado en reglas para la detección de Banderas Rojas (garantizando la derivación médica)[13,25,29]. Segundo, un motor de procesamiento de lenguaje natural (NLP, Natural Language Processing) que permite al sistema interpretar las descripciones en texto libre del paciente y transformarlas en variables clínicas estructuradas que el algoritmo puede procesar, eliminando la necesidad de que el usuario rellene formularios rígidos. Este componente se complementa con una arquitectura de generación aumentada por recuperación (RAG). Esta tecnología permite que el algoritmo consulte una base de datos cerrada de evidencia científica validada (guías clínicas) antes de responder, asegurando que la información sea veraz y evitando la generación de respuestas falsas o "alucinaciones" algorítmicas. Así, el sistema puede modular y personalizar la educación en dolor al perfil psicosocial del paciente (ej. reduciendo el lenguaje alarmista en caso de kinesiofobia)[13,14]. La arquitectura técnica de este sistema (Lumbar-AI), incluyendo el diseño de prompts y una simulación de caso clínico, se detalla en los Textos S1 y S2 del material suplementario. Esta propuesta responde directamente a la necesidad detectada en la revisión de equilibrar la automatización de procesos con la supervisión ética indispensable del fisioterapeuta[24,38].

5. Conclusiones

La revisión de la evidencia reciente sugiere que la Inteligencia Artificial presenta resultados prometedores como herramienta de soporte clínico, especialmente en tareas de triaje y pronóstico funcional, donde los algoritmos de aprendizaje automático muestran un potencial predictivo prometedor, comparable o superior al de las herramientas de valoración tradicionales[1,20,35]. Sin embargo, la eficacia terapéutica directa de estas soluciones es heterogénea, lo que posiciona al modelo híbrido (intervención presencial potenciada por soporte digital) como una estrategia pragmática y prometedora para el manejo del dolor lumbar crónico[19,22,27,28].

Desde una perspectiva conceptual, la opacidad algorítmica ("Caja Negra") constituye una de las barreras más relevantes para la integración clínica de estas herramientas, aunque no necesariamente la única ni la principal, ya que limita la confianza necesaria para la toma de decisiones críticas[10,34]. Por tanto, el avance del campo depende en gran medida de una transición hacia sistemas de IA Explicable (XAI) que permitan al profesional trazar la lógica clínica subyacente[10,24].

Como implicación práctica innovadora, este trabajo sugiere que el futuro de la asistencia digital no reside en la automatización independiente, sino en el diseño de "Copilotos Clínicos" bajo supervisión humana (Human-in-the-loop)[3,24]. Estos sistemas tienen la viabilidad teórica de escalar la personalización del tratamiento y filtrar banderas rojas, liberando al fisioterapeuta de la carga analítica para centrarse en la alianza terapéutica[13,24]. En definitiva, la profesión se dirige hacia un paradigma de "Inteligencia Aumentada", donde el fisioterapeuta no debe mantener un rol de usuario pasivo que deja el control y la toma de decisiones a la IA, sino que debe liderar el diseño ético y funcional de estas nuevas herramientas para convertirlas en instrumentos de soporte que agilicen y faciliten la práctica clínica[2,38]. (véanse arquitectura de prompts y simulación en Texto S1 y S2 del Material Suplementario).

Referencias

1. Seas A, Zachem TJ, Valan B, Goertz C, Nischal S, Chen SF, et al. Machine learning in the diagnosis, management, and care of patients with low back pain: a scoping review of the literature and future directions. Spine J. 2025; 25(1):18–31. doi: 10.1016/j.spinee.2024.09.010

2. Hasan F, Mudey A, Joshi A. Role of Internet of Things (IoT), Artificial Intelligence and Machine Learning in Musculoskeletal Pain: A Scoping Review. Cureus. 2023; 15(4):e37352. doi: 10.7759/cureus.37352

3. Toh ZA, Berg B, Han QYC, Hey HWD, Pikkarainen M, Grotle M, et al. Clinical Decision Support System Used in Spinal Disorders: Scoping Review. J Med Internet Res. 2024; 26:e53951. doi: 10.2196/53951

4. Alonso-García M, Sarría-Santamera A. The Economic and Social Burden of Low Back Pain in Spain: A National Assessment of the Economic and Social Impact of Low Back Pain in Spain. Spine (Phila Pa 1976). 2020; 45(16):E1026-E1032. doi: 10.1097/BRS.0000000000003476

5. Dal Prete A, Gandolla M, Andreoni G, Braghin F. Low Back Exoskeletons in Industry 5.0: From Machines to Perceiving Co-Pilots—A State-of-the-Art Review. Sensors. 2025; 25(7):1958. doi: 10.3390/s25071958

6. Falla D, Devecchi V, Jiménez-Grande D, Rügamer D, Liew BXW. Machine learning approaches applied in spinal pain research. J Electromyogr Kinesiol. 2021;61:102599. doi: 10.1016/j.jelekin.2021.102599

7. Rajan PV, Eisenberg J, Selverian S, Wawrose R, Yoon ST. The cost-effectiveness of physical therapy versus laminectomy for lumbar spinal stenosis: a Markov decision analysis. Spine J. 2025; 25(8):1608–17. doi: 10.1016/j.spinee.2025.01.005

8. Tropea P, Schlieter H, Sterpi I, Judica E, Gand K, Caprino M, et al. Rehabilitation, the Great Absentee of Virtual Coaching in Medical Care: Scoping Review. J Med Internet Res. 2019; 21(10):e12805. doi: 10.2196/12805

9. Chen K, Zhai X, Sun K, Wang H, Yang C, Li M. A narrative review of machine learning as promising revolution in clinical practice of scoliosis. Ann Transl Med. 2021; 9(1):67. doi: 10.21037/atm-20-5495

10. Ren G, Yu K, Xie Z, Wang P, Zhang W, Huang Y, et al. Current Applications of Machine Learning in Spine: From Clinical View. Global Spine J. 2022; 12(8):1827–40. doi: 10.1177/21925682211035363

11. Jawed AM, Zhang L, Zhang Z, Liu Q, Ahmed W, Wang H. Artificial intelligence and machine learning in spine care: Advancing precision diagnosis, treatment, and rehabilitation. World J Orthop. 2025; 16(8):107064. doi: 10.5312/wjo.v16.i8.107064

12. Kim HB, Kim HS, Kim SJ, Yoo JI. Spine muscle auto segmentation techniques in MRI imaging: a systematic review. BMC Musculoskelet Disord. 2024; 25(1):716. doi: 10.1186/s12891-024-07777-4

13. Luo M, Duan Z, Gao J, Sun Y, Chen L, Feng X. Evaluating the role of ChatGPT in rehabilitation medicine: a narrative review. Front Digit Health. 2025;7:1618510. doi: 10.3389/fdgth.2025.1618510

14. Khan MA, Koh RGL, Rashidiani S, Liu T, Tucci V, Kumbhare D, et al. Cracking the Chronic Pain code: A scoping review of Artificial Intelligence in Chronic Pain research. Artif Intell Med. 2024;151:102849. doi: 10.1016/j.artmed.2024.102849

15. Hill CJ, Banerjee A, Hill J, Stapleton C. Diagnostic clinical prediction rules for categorising low back pain: A systematic review. Musculoskeletal Care. 2023; 21(4):1482–96. doi: 10.1002/msc.1816

16. 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

17. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro Scale for Rating Quality of Randomized Controlled Trials. Phys Ther. 2003; 83(8):713–21. doi: 10.1093/ptj/83.8.713

18. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses [Internet]. Ottawa (ON): Ottawa Hospital Research Institute; [cited 2026 Feb 16]. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

19. Xiao C, Zhao Y, Li G, Zhang Z, Liu S, Fan W, et al. Clinical Efficacy of Multimodal Exercise Telerehabilitation Based on AI for Chronic Nonspecific Low Back Pain: Randomized Controlled Trial. JMIR Mhealth Uhealth. 2025; 13:e56176. doi: 10.2196/56176

20. Cheema C, Baldwin J, Rodeghero J, Werneke MW, Mioduski JE, Jeffries L, et al. Use of machine learning to identify prognostic variables for outcomes in chronic low back pain treatment: a retrospective analysis. J Man Manip Ther. 2025; 33(1):63–72. doi: 10.1080/10669817.2024.2424619

21. Griefahn A, Avermann F, Zalpour C, Marshall RP, Cordon Morillas I, Luedtke K. Exploring the Effect of an 8-Week AI-Composed Exercise Program on Pain Intensity and Well-Being in Patients With Spinal Pain: Retrospective Cohort Analysis. JMIR Form Res. 2025; 9:e57826. doi: 10.2196/57826

22. Alzouhayli K, Schilaty ND, Nagai T, Rigamonti L, McPherson AL, Holmes BD, et al. Artificial intelligence-guided therapy: Clinic versus home users. Long-term differences in patient-reported outcomes in patients with low back pain. J Orthop Rep. 2025; 4(2):100592. doi: 10.1016/j.jorep.2025.100592

23. Pfeifer AC, Schröder-Pfeifer P, Schiltenwolf M, Vogt L, Schneider C, Platen P, et al. Finding Predictive Factors of Stabilization Exercise Adherence in Randomized Controlled Trials on Low Back Pain: An Individual Data Reanalysis Using Machine Learning Techniques. Arch Phys Med Rehabil. 2025; 106(5):738–49. doi: 10.1016/j.apmr.2024.12.015

24. Areias AC, Janela D, Moulder RG, Molinos M, Bento V, Moreira C, et al. Applying AI to Safely and Effectively Scale Care to Address Chronic MSK Conditions. J Clin Med. 2024; 13(15):4366. doi: 10.3390/jcm13154366

25. Areias AC, Moulder RG, Molinos M, Janela D, Bento V, Moreira C, et al. Predicting Pain Response to a Remote Musculoskeletal Care Program for Low Back Pain Management: Development of a Prediction Tool. JMIR Med Inform. 2024; 12:e64806. doi: 10.2196/64806

26. Zhao X, Dannenberg K, Repsilber D, Gerdle B, Molander P, Hesser H. Prognostic subgroups of chronic pain patients using latent variable mixture modeling within a supervised machine learning framework. Sci Rep. 2024; 14(1):12579. doi: 10.1038/s41598-024-62542-w

27. Bates NA, Huffman A, Goodyear E, Nagai T, Rigamonti L, Breuer L, et al. Physical clinical care and artificial-intelligence-guided core resistance training improve endurance and patient-reported outcomes in subjects with lower back pain. Clin Biomech. 2023;103:105902. doi: 10.1016/j.clinbiomech.2023.105902

28. Marcuzzi A, Nordstoga AL, Bach K, Aasdahl L, Nilsen TIL, Bardal EM, et al. Effect of an Artificial Intelligence–Based Self-Management App on Musculoskeletal Health in Patients With Neck and/or Low Back Pain Referred to Specialist Care: A Randomized Clinical Trial. JAMA Netw Open. 2023; 6(6):e2320400. doi: 10.1001/jamanetworkopen.2023.20400

29. Tun Firzara AM, Teo CH, Teh S, Su JY, Zaini HSM, Suhaimi AB, et al. Evaluation of an electronic clinical decision support system (DeSSBack) to improve low back pain management: a pilot cluster randomized controlled trial. Fam Pract. 2023; 40(5–6):742–52. doi: 10.1093/fampra/cmad044

30. Park C, Yi C, Choi WJ, Lim HS, Yoon HU, You SJH. Long-term effects of deep-learning digital therapeutics on pain, movement control, and preliminary cost-effectiveness in low back pain: a randomized controlled trial. Digit Health. 2023;9:20552076231217816. doi: 10.1177/20552076231217816

31. Hartmann R, Avermann F, Zalpour C, Griefahn A. Impact of an AI app-based exercise program for people with low back pain compared to standard care: A longitudinal cohort-study. Health Sci Rep. 2023; 6(1):e1060. doi: 10.1002/hsr2.1060

32. Nordstoga AL, Aasdahl L, Sandal LF, Dalager T, Kongsvold A, Mork PJ, et al. The Role of Pain Duration and Pain Intensity on the Effectiveness of App-Delivered Self-Management for Low Back Pain (selfBACK): Secondary Analysis of a Randomized Controlled Trial. JMIR Mhealth Uhealth. 2023; 11:e40422. doi: 10.2196/40422

33. Strigo IA, Simmons AN, Giebler J, Schilling JM, Moeller-Bertram T. Unsupervised learning for prognostic validity in patients with chronic pain in transdisciplinary pain care. Sci Rep. 2023; 13(1):7603. doi: 10.1038/s41598-023-34611-z

34. Zmudzki F, Smeets RJEM. Machine learning clinical decision support for interdisciplinary multimodal chronic musculoskeletal pain treatment. Front Pain Res. 2023;4:1177070. doi: 10.3389/fpain.2023.1177070

35. Knoop J, van Lankveld W, Beijer L, Geerdink FJB, Heymans MW, Hoogeboom TJ, et al. Development and internal validation of a machine learning prediction model for low back pain non-recovery in patients with an acute episode consulting a physiotherapist in primary care. BMC Musculoskelet Disord. 2022; 23(1):834. doi: 10.1186/s12891-022-05718-7

36. Park J, Chung S, Park JH. Real-Time Exercise Feedback through a Convolutional Neural Network: A Machine Learning-Based Motion-Detecting Mobile Exercise Coaching Application. Yonsei Med J. 2022; 63 Suppl:S34–42. doi: 10.3349/ymj.2022.63.S34

37. Anan T, Kajiki S, Oka H, Fujii T, Kawamata K, Mori K, et al. Effects of an artificial intelligence-assisted health program on workers with neck/shoulder pain/stiffness and low back pain: Randomized controlled trial. JMIR Mhealth Uhealth. 2021; 9(9):e27535. doi: 10.2196/27535

38. Soer R. Artificial intelligence in musculoskeletal rehabilitation. J Back Musculoskelet Rehabil. 2023; 36(5):993–4. doi: 10.3233/BMR-235003

Información editorial

Correspondencia

869739@unizar.es

Fechas

Recibido: 08.03.2026
Aceptado: 02.06.2026
Publicado: 23.07.2026

Contribuciones de autoría

Conceptualización, M.C.L.; metodología, M.C.L.; investigación, M.C.L.; redacción—preparación del borrador original, M.C.L.; redacción—revisión y edición, M.C.L; supervisión, M.C.L. Todos los autores han leído y aceptado la versión publicada del manuscrito.

Declaración Ética

No aplicable.

Financiación

Esta investigación no recibió financiación externa.

Declaración de Disponibilidad de Datos

Los datos presentados en este estudio están disponibles en el cuerpo del artículo y en el material suplementario. No se generaron datos primarios con pacientes, ya que la investigación consiste en una revisión bibliográfica y una propuesta teórica.

Conflicto de intereses

El autor declara no tener conflictos de interés.

Cita sugerida

Campillos Lázaro M. Uso de la Inteligencia Artificial como Sistema de Apoyo a la Decisión Clínica en el Manejo del Dolor Lumbar Crónico. Una Revisión de Alcance con Componente Conceptual Aplicado. Invasive Physiother Musculoskelet Med. 2026;2:e25. doi: 10.63360/ipmm.v2.e25

Descargo de responsabilidad/Nota del Editor

Las declaraciones, opiniones y datos contenidos en todas las publicaciones son únicamente del o los autores individuales y contribuyentes, y no de IPMM y/o del o los editores. IPMM y/o el o los editores rechazan toda responsabilidad por cualquier daño a personas o propiedades que resulte de ideas, métodos, instrucciones o productos mencionados en el contenido.

Copyright

© 2026 Las/os Autoras/es.

Publicado bajo licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA) (https://creativecommons.org/licenses/by-nc-sa/4.0).

MATERIAL SUPLEMENTARIO

Fecha de consulta y Filtros

Fecha: 21-22 de noviembre de 2025.

Filtros: Publicaciones 2020-2025; Idioma Inglés o Español.

1. Estrategia de Búsqueda

Para evitar repeticiones, se definieron tres bloques temáticos combinados mediante operadores booleanos:

Bloque A (Tecnología): "Artificial Intelligence" OR "Decision Support Systems" OR "Machine Learning" OR "Deep Learning" OR "Neural Networks" OR "ChatGPT" OR "Large Language Model" OR "Generative AI" OR "CDSS" OR "Chatbot" OR "Virtual Assistant". (En PubMed se añadieron los términos MeSH correspondientes).

Bloque B (Intervención): "Physical Therapy" OR "Physiotherapy" OR "Rehabilitation" OR "Exercise Therapy".

Bloque C (Patología): "Low Back Pain" OR "Chronic Pain" OR "Sciatica" OR "LBP" OR "Lumbar" OR "Back Pain".

Tabla S1. Estrategia de búsqueda detallada y resultados por base de datos.

Base de Datos	Sintaxis y Campos Específicos	Resultados
PubMed (MEDLINE)	(Bloque A [MeSH/All Fields]) AND (Bloque B [MeSH/All Fields]) AND (Bloque C [MeSH/All Fields])	146
Scopus	TITLE-ABS-KEY(Bloque A) AND TITLE-ABS-KEY(Bloque B) AND TITLE-ABS-KEY(Bloque C)	264
Web of Science (WoS)	TS=(Bloque A AND Bloque B AND Bloque C). Campo: Topic.	170
Cochrane Library	Búsqueda Avanzada: Linea #1 (Bloque A) AND Linea #2 (Bloque B) AND Linea #3 (Bloque C)	55
PEDro	Búsqueda simple por término clave "Artificial Intelligence" (con cribado manual posterior).	26

Base de Datos	Sintaxis y Campos Específicos	Resultados
TOTAL BRUTO	Suma de registros identificados (antes de eliminar duplicados)	661

2. Evaluación de la Calidad Metodológica

Tabla S2. Desglose de puntuación de la escala PEDro para los ensayos clínicos incluidos.

Autor y Año	Total	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Park et al., 2023	7/10	✓	✓	✗	✓	✗	✗	✓	✓	✓	✓	✓
Xiao et al., 2025	7/10	✓	✓	✓	✓	✗	✗	✓	✓	✗	✓	✓
Tun Firzara et al., 2023	7/10	✓	✓	✗	✗	✓	✗	✓	✓	✓	✓	✓
Marcuzzi et al., 2023	6/10	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓
Nordstoga et al., 2023	6/10	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓	✓
Anan et al., 2021	4/10	✓	✓	✗	✓	✗	✗	✗	✗	✗	✓	✓
Bates et al., 2023	3/10	✓	✓	✗	✗	✗	✗	✓	✗	✗	✗	✓

Leyenda de Criterios PEDro:

C1: Criterios de elección especificados (no suma al total).

C2: Asignación aleatoria.

C3: Asignación oculta (Enmascaramiento).

C4: Similitud inicial entre grupos.

C5: Cegamiento de sujetos.

C6: Cegamiento de terapeutas.

C7: Cegamiento de evaluadores.

C8: Seguimiento adecuado (>85%).

C9: Análisis por intención de tratar.

C10: Comparación estadística entre grupos.

C11: Medidas puntuales y de variabilidad.

Evaluación de la Calidad Metodológica (Escala PEDro)

La calidad metodológica de los 7 ensayos clínicos aleatorizados (ECA) incluidos presentó una puntuación media de 5.7/10 (rango: 3-7). La gran mayoría de los estudios (n=5) demostraron una calidad metodológica alta (≥6/10), destacando por el estricto cumplimiento de la asignación aleatoria (C2) y un riguroso reporte estadístico (C10, C11), además de una alta adherencia al análisis por intención de tratar (C9, presente en 4 de estos 5 ensayos). Por el contrario, dos ensayos (Anan et al. y Bates et al.)[27,37] presentaron un alto riesgo de sesgo (≤4/10) debido a deficiencias significativas en la ocultación de la asignación y en las altas tasas de abandono.

El análisis transversal de los ítems revela que la principal debilidad metodológica de la literatura actual es el cegamiento. Como dificultad inherente a las intervenciones activas de ejercicio y tecnología digital, el cegamiento de los terapeutas (C6) fue nulo y el de los sujetos (C5) casi anecdótico (presente en un único estudio). No obstante, este riesgo de sesgo se mitigó parcialmente gracias a que el cegamiento de los evaluadores independientes (C7) sí se implementó con éxito en más de la mitad de los ensayos (n=4).

Tabla S3. Desglose de puntuación de la escala Newcastle-Ottawa (NOS) para los estudios de cohortes y casos-controles incluidos.

Autor y Año	Total	Diseño	S1	S2	S3	S4	C1	C2	R1	R2	R3
Areias et al. (Tool), 2024	9/9	Cohorte	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pfeifer et al., 2025	9/9	Cohorte	✓	✓	✓	✓	✓	✓	✓	✓	✓
Zmudzki & Smeets, 2023	8/9	Cohorte	✓	✓	✓	✓	✓	✓	✓	✓	✗
Zhao et al., 2024	8/9	Cohorte	✓	✓	✓	✗	✓	✓	✓	✓	✓
Areias et al. (Scale), 2024	8/9	Cohorte	✓	✓	✓	✓	✓	✓	✓	✓	✗
Cheema et al., 2024	7/9	Cohorte	✗	✓	✓	✓	✓	✓	✓	✓	✗
Knoop et al., 2022	7/9	Cohorte	✓	✓	✗	✓	✓	✓	✗	✓	✓
Strigo et al., 2023	6/9	Cohorte	✓	✓	✓	✓	✓	✗	✗	✓	✗
Alzouhayli et al., 2025	4/9	Cohorte	✓	✗	✓	✓	✗	✗	✗	✓	✗
Griefahn et al., 2025	4/9	Cohorte	✓	✗	✓	✓	✗	✗	✗	✓	✗
Hartmann et all., 2023	4/9	Cohorte	✗	✗	✓	✓	✗	✗	✗	✓	✓
Park et al., 2022	3/9	Casos y controles	✗	✓	✓	✓	✗	✗	✗	✗	✗

Leyenda de Criterios Newcastle-Ottawa (NOS): Cada ítem se puntúa con un máximo de 1 punto (✓), excepto la comparabilidad que permite hasta 2 puntos. Puntuación máxima: 9.

Selección (S1-S4):

S1: Representatividad de la cohorte expuesta.

S2: Selección de la cohorte no expuesta.

S3: Determinación de la exposición.

S4: Demostración de que el resultado no estaba presente al inicio.

Comparabilidad (C1-C2):

C1: Control del factor de confusión principal (ej. edad/severidad).

C2: Control de factores adicionales.

Resultados (R1-R3):

R1: Evaluación del resultado independiente/ciega.

R2: Seguimiento lo suficientemente largo.

R3: Adecuación del seguimiento (pérdidas <20%).

Evaluación de la Calidad Metodológica (Escala Newcastle-Ottawa)

La evaluación de los 12 estudios observacionales (11 cohortes y 1 casos-controles) mediante la Escala Newcastle-Ottawa (NOS) reveló una calidad metodológica heterogénea (rango: 3-9 sobre 9 puntos posibles). Casi la mitad de la muestra (n=5) demostró una alta calidad (≥8/9), destacando los trabajos de Areias et al. (Tool)[25] y Pfeifer et al.[23] con puntuaciones perfectas, los cuales exhibieron una gran robustez en la representatividad de la cohorte, el ajuste estadístico multivariante y la evaluación de resultados. En un nivel intermedio, tres estudios mostraron un riesgo moderado de sesgo (6-7/9), presentando deficiencias menores en la demostración de la ausencia del resultado al inicio del estudio o en la evaluación ciega.

Por el contrario, un tercio de las investigaciones (n=4) presentó un alto riesgo de sesgo (≤4/9). El análisis transversal de los dominios evidencia que la Comparabilidad fue el principal factor limitante: los estudios de menor calidad fallaron sistemáticamente al no controlar factores de confusión clave (como la edad o la severidad basal). Asimismo, dentro del dominio de Resultados, la adecuación del seguimiento se identificó como la principal debilidad global; más de la mitad de los estudios (n=7) reportaron pérdidas significativas o no detallaron adecuadamente las tasas de retención, introduciendo un potencial sesgo de desgaste en la evidencia acumulada.

3. Arquitectura Técnica del Sistema

Texto S1. Arquitectura de Prompts del Sistema Lumbar-AI.

Prompt Nivel 1: Módulo de Seguridad

ROL: Fisioterapeuta Experto en triaje (Primary Care).

TAREA: Analizar Banderas Rojas en el input del paciente.

RESTRICCIONES CRÍTICAS:

Si hay anestesia en silla de montar, pérdida de control de esfínteres, fiebre/pérdida de peso inexplicada o trauma severo -> ACCIÓN: "STOP_URGENCIA".

Si dolor constante, no mecánico y nocturno -> ACCIÓN: "STOP_DERIVACION_MEDICA".

Solo si NO hay banderas rojas -> ACCIÓN: "APTO" y clasificar fenotipo.

Fenotipos:

1: Nociceptivo mecánico simple (bajo riesgo psicosocial)

Perfil clínico: Dolor lumbar asociado al movimiento o a posturas (por ejemplo, intolerancia a flexión/extensión o deficiencia de estabilidad), sin banderas rojas. El pacien-

te presenta alta autoeficacia, resiliencia y ausencia de distrés emocional significativo. Impacto en Módulo de Intervención: Activa un protocolo directo de control motor y fortalecimiento, con énfasis en movilidad y ejercicio progresivo. La cápsula educativa se centra en automanejo básico y en fomentar la adherencia al ejercicio, dado el buen pronóstico esperado

2: Nociceptivo mecánico con kinesiofobia (miedo-evitación)

Perfil clínico: Dolor mecánico asociado a creencias desadaptativas de daño tisular. Puntuaciones elevadas en miedo al movimiento (por ejemplo, TSK-11) y conductas de evitación de la actividad física. Este fenotipo corresponde al caso de simulación descrito en el Texto S2. Impacto en Módulo de Intervención: Activa un protocolo de exposición gradual, seleccionando ejercicios de baja carga biomecánica inicial para reintroducir el movimiento. La cápsula educativa prioriza la reestructuración de la creencia “dolor = daño” y el uso de lenguaje no amenazante

3: Nociceptivo mecánico con distrés emocional

Perfil clínico: Dolor de intensidad leve-moderada que se acompaña de alta interferencia percibida debido a niveles elevados de ansiedad, bajo estado de ánimo o escaso apoyo social. Se corresponde con pacientes que, pese a no presentar el mayor dolor físico, muestran un impacto funcional alto por factores emocionales. Impacto en Módulo de Intervención: El algoritmo prioriza la inclusión de componentes de regulación del estrés (respiración, relajación, mindfulness) combinados con ejercicio. La cápsula educativa se centra en el manejo del estrés, la normalización del dolor crónico y la mejora del afrontamiento

4: Irritabilidad alta / severidad global (patrón mixto/nociplástico)

Perfil clínico: Dolor de alta intensidad, discapacidad relevante (por ejemplo, ODI elevado), posible extensión a múltiples áreas y altos niveles de catastrofismo y distrés. Aunque el paciente es “APTO” (sin banderas rojas), presenta indicios de sensibilización central y riesgo elevado de mala evolución. Impacto en Módulo de Intervención: Activa un protocolo conservador con ejercicios centrados en movilidad suave en rangos indoloros y educación intensiva en dolor. El sistema puede incluir un indicador interno para sugerir valoración multidisciplinar (fisioterapia + abordaje psicológico) si la respuesta clínica es insuficiente tras un periodo razonable

OUTPUT ESPERADO: JSON {"Estado": "APTO/STOP", "Razon_Seguridad": "…", "Fenotipo": "1/2/3/4" (si APTO), "Driver_Principal": "Mecanico / Kinesiofobia / Distres / Severidad"}.

Prompt Nivel 2: Módulo de Intervención

CONTEXTO: Paciente con Dolor Lumbar Crónico clasificado "APTO".

OBJETIVO: Diseñar sesión de ejercicio domiciliario (20 min).

INSTRUCCIONES DE PERSONALIZACIÓN:

Selección: Basada en control motor y exposición gradual.

Educación: Incluir SIEMPRE "Cápsula Educativa" (dolor ≠ daño).

Psicosocial: Si Kinesiofobia alta (>40 TSK-11), usar lenguaje no amenazante (evitar palabras como "daño", "rotura", "desgaste").

OUTPUT ESPERADO: Tabla con 3 ejercicios + 1 Consejo Educativo.

4. Prueba de Concepto

Texto S2. Simulación de Caso Clínico.

Para ilustrar el funcionamiento del sistema Lumbar-AI Assist en un entorno controlado, se presenta una simulación completa con un perfil de paciente estándar.

1. PERFIL DE ENTRADA (INPUT)

Paciente: Varón, 45 años. Administrativo.

Motivo de consulta: Dolor lumbar bajo de 3 meses de evolución, exacerbado al estar sentado.

Escalas:

EVA: 6/10.

Oswestry (ODI): 24% (Discapacidad moderada).

TSK-11 (Kinesiofobia): 48/68 (Alta).

Banderas Rojas: Negativas.

2. PROCESAMIENTO DEL ALGORITMO

Módulo 1 (Seguridad): Detecta ausencia de síntomas de alarma. Clasifica como APTO.

Fenotipo presuntivo: Nociceptivo mecánico con componente psicosocial (miedo-evitación) (Fenotipo 2 / Driver: Kinesiofobia)

Módulo 2 (Intervención): Activa el protocolo de "Exposición Gradual" debido a la alta kinesiofobia (TSK > 40). Selecciona ejercicios de baja carga biomecánica inicial.

3. SALIDA DEL SISTEMA (OUTPUT GENERADO)

Cápsula Educativa:

Mensaje generado: "Juan, tu espalda es robusta y segura. El dolor al sentarte es una 'alarma sensible', no daño real. El movimiento suave es el mejor analgésico".

Prescripción de Tareas:

- Báscula Pélvica en Supino (3x10): Para reintroducir movimiento sin miedo.
- Puente Glúteo Bipodal (3x8, solo en rango cómodo): Para fuerza de baja carga.
- Cat-Camel (2 min continuos, lento): Para control motor.

Criterio de Progresión: Si el dolor post-ejercicio es < 4/10, sugerir aumentar la carga en la siguiente sesión.