

INTELIGENCIA ARTIFICIAL Y FÚTBOL: USO Y EFICIENCIA DEL APRENDIZAJE AUTOMÁTICO EN LA PREDICCIÓN DE LESIONES

CIENCIAS DE LA ACTIVIDAD FÍSICA Y DEPORTE

TFG

**FACULTAD CIENCIAS DE LA ACTIVIDAD
FÍSICA Y EL DEPORTE**



Realizado por: Víctor Martínez y Andrés Varela

Año Académico: 2024-2025

Tutor/a: Iyán Iván Baragaño

Área: Revisión bibliográfica

Resumen Gráfico

Resumen

OBJETIVO

Analizar el uso de la inteligencia artificial como herramienta para predecir lesiones en futbolistas de élite.



METODOLOGÍA

Revisión sistemática de 22 estudios científicos.

MODELOS DE IA UTILIZADOS

Árboles de decisión (Random Forest, XGBoost, AdaBoost, Gradient Boosting): mayor eficacia.

SVM y redes neuronales: rendimiento variable, dependientes del balance de datos.



Regresiones penalizadas (LASSO, Elastic Net) y Subgroup Discovery: valor en perfiles individuales.

VARIABLES MÁS INFLUYENTES

Desaceleraciones, sprints, carga aguda-crónica.



Historial de lesiones, fuerza excéntrica, desequilibrios musculares.



PRINCIPALES LIMITACIONES

Desequilibrio entre clases (más datos de jugadores sanos).

Falta de validación en entornos reales.

CONCLUSIONES

La IA puede ser útil para prevenir lesiones si se aplica con rigor metodológico, se mejora el acceso a datos de calidad y se valida en contextos reales y diversos.



Resumen

El objetivo de esta revisión sistemática fue analizar la eficacia de la inteligencia artificial (IA) en la predicción de lesiones en el fútbol de élite, para ello fueron analizados 17 estudios que implementan modelos como XGBoost, Random Forest, SVM y redes neuronales. Se identificaron variables clave que aportaron significativamente a la capacidad predictiva de los modelos, entre las cuales sobresalen especialmente las desaceleraciones, asociadas con una alta capacidad predictiva (precisión del 92%, $F1 = 94,7\%$), el índice carga aguda-crónica (ACWR) ($AUC = 0,91$) y sprints repetidos (HSR) (precisión del 78%, $AUC = 0,85$). Modelos como XGBoost lograron sensibilidades superiores al 85% cuando integraron datos GPS y antecedentes lesionales. No obstante, en algunos estudios que alcanzaron precisiones del 90,38% se observaron limitaciones metodológicas significativas. Estos estudios estarían utilizando generación de datos sintéticos, con menor validez en contextos reales. Otras limitaciones metodológicas observadas fueron el desbalance de clases, que afectó la sensibilidad en varios modelos, y la falta de estandarización en métricas de evaluación (uso inconsistente de AUC, precisión, sensibilidad, MAE, entre otros). En contraposición, otros estudios demostraron alta aplicabilidad práctica al emplear datos longitudinales reales y variables fácilmente monitoreables en clubes profesionales. En conjunto, los hallazgos respaldan el potencial de la IA como herramienta complementaria en la prevención de lesiones, siempre que se aborden adecuadamente los desafíos metodológicos y se priorice la validación en contextos reales de competencia.

Palabras clave: Inteligencia artificial, fútbol, predicción de lesiones, riesgo de lesiones, modelo de aprendizaje

Graphic Abstract

Abstract

OBJECTIVE

To analyze the use of artificial intelligence as a tool to predict injuries in elite football players.



METHODOLOGY

Systematic review of 22 scientific studies.

AI MODELS USED

Decision tree models (Random Forest, XGBoost, AdaBoost, Gradient Boosting): highest effectiveness.

SVM and neural networks: variable performance, dependent on data balance.



Penalized regressions (LASSO, Elastic Net) and Subgroup Discovery: valuable for individualized profiles.

MOST INFLUENTIAL VARIABLES

Decelerations, sprints, acute-chronic workload ratio.



Injury history, eccentric strength, muscular imbalances.



MAIN LIMITATIONS

Class imbalance (more data from non-injured players).

Lack of validation in real-world environments.

CONCLUSIONS

AI can be a useful tool to prevent injuries if applied with methodological rigor, supported by high-quality data, and validated in real and diverse contexts.



Abstract

The objective of this systematic review was to analyze the effectiveness of artificial intelligence (AI) in injury prediction in elite football. To this end, 17 studies were analyzed that implemented models such as XGBoost, Random Forest, SVM, and neural networks. Key variables were identified that significantly contributed to the predictive capacity of the models, among which decelerations stood out in particular, being associated with high predictive performance (accuracy of 92%, $F1 = 94,7\%$), the acute:chronic workload ratio (ACWR) ($AUC = 0,91$), and repeated high-speed runs (HSR) (accuracy of 78%, $AUC = 0.85$). Models such as XGBoost achieved sensitivities above 85% when integrating GPS data and injury history. However, some studies that reported accuracies of up to 90,38% also presented significant methodological limitations, particularly those using synthetic data generation, which tend to have lower validity in real-world contexts. Other methodological issues included class imbalance, which negatively affected sensitivity in several models, and the lack of standardization in evaluation metrics (e.g., inconsistent use of AUC, accuracy, sensitivity, MAE, among others). In contrast, other studies showed high practical applicability by using real longitudinal data and easily monitored variables within professional club settings. Overall, the findings support the potential of AI as a complementary tool in injury prevention, provided that methodological challenges are properly addressed and validation in real competitive contexts is prioritized.

Keywords: Artificial intelligence, football, injury prediction, injury risk, machine learning

Índice:

1. Introducción.....	8
2. Objetivos.....	12
3. Metodología.....	12
3.1. Diseño.....	12
3.2. Estrategia de búsqueda.....	12
3.3. Criterios de selección.....	13
3.4. Diagrama de flujo.....	14
4. Discusión.....	15
4.1. Variables utilizadas en los modelos.....	16
4.2. Técnicas de Inteligencia Artificial utilizadas.....	19
4.3. Modelos de IA utilizados y su capacidad predictiva.....	19
4.4. Problemas metodológicos.....	20
5. Contribuciones a los Objetivos de Desarrollo Sostenible.....	22
6. Futuras líneas de investigación.....	22
7. Conclusiones.....	23
7.1. Objetivo General.....	23
7.2. Variables con mayor influencia.....	23
7.3. Identificación de las técnicas de IA con mejor rendimiento:.....	23
7.4. Limitaciones de la IA en la predicción de lesiones.....	23
8. Referencias.....	25
9. Anexos.....	31
9.1. Tabla 1. Cuadro de Artículos empleados.....	31

Índice de figuras:

Figura 1. Diagrama de flujo.....	13
---	----

Índice de tablas:

Tabla 1. Resumen de las variables analizadas en cada artículo y la eficacia predictiva de los modelos.....	12
---	----

1. Introducción

Una de las principales problemáticas que están presentes en el mundo del fútbol son las lesiones musculoesqueléticas. Desde una perspectiva tanto económica como de rendimiento, las lesiones suponen un impacto negativo significativo para los propios jugadores, los clubes de fútbol y las competiciones en las que participan. A partir del documento importado se podría afirmar que esto se da debido a un considerable aumento de volumen de exposición de los jugadores profesionales frente a cargas derivadas tanto de entrenamientos como partidos oficiales. En el caso de clubes que participan en competiciones europeas el calendario semanal tipo se extiende a dos encuentros oficiales y, al menos, tres sesiones de entrenamientos, además de los desplazamientos para la disputa de dichos eventos. Un estudio realizado por Howden Group (2023) revela que, tras el Mundial de Catar 2022, el número de lesiones en las principales ligas europeas aumentó un 24%, alcanzando las 3.967 lesiones. Según el mismo estudio, este incremento se traduce en un aumento del 40% en los costes asociados, alcanzando los 696 millones de euros. Las lesiones en la rodilla, especialmente graves, duplicaron la media de días de baja, pasando de 26 a 52 días en tres años. La congestión del calendario deportivo se intensifica cuando coincide con la celebración de competiciones de selecciones nacionales. La saturación del calendario no solo afecta la salud física de los jugadores, sino también su bienestar mental. El estrés derivado de la acumulación de partidos y viajes frecuentes puede llevar a la fatiga crónica y aumentar el riesgo de lesiones. FIFPRO ha destacado que la salud mental de los futbolistas profesionales se ve comprometida debido a la sobrecarga de competiciones y la falta de tiempo para la recuperación (FIFPRO, 2024). Un ejemplo reciente fue la campaña pasada 2023/2024 donde tuvo lugar la Eurocopa 2024 desde el 14 de junio hasta el 14 de julio o incluso la reciente incorporación de el nuevo formato del Mundial de Clubes que tendrá lugar el 14 de Junio hasta el 14 de Julio. A todo esto se le suma que la pretemporada este año 2024/25 inició el 31 de Julio, sólo 15 días después de terminar la temporada regular 2023/24. Además a esto se le suman las giras de carácter comercial y competitivo en diferentes países como la *Soccer Champions* realizada en Estados Unidos en la que participaron equipos como el Real Madrid,

Barcelona, Milan y Chelsea evidenciando la poca ventana de descanso y la continua exigencia física del fútbol moderno.

El impacto y la repercusión de estas lesiones no solo afectan a la salud y disponibilidad de los futbolistas sino que conllevan un peor rendimiento colectivo del equipo. Últimamente se ha popularizado el concepto de *Player availability* que define que mantener a los jugadores libres de lesiones y listos para participar en la competición es extremadamente importante en los escenarios actuales de deportes de equipo de élite, ya que está relacionado con el rendimiento del equipo. De hecho, la disponibilidad individual de los jugadores puede ser más importante que el hecho de que cada jugador esté en su nivel máximo de rendimiento individual (Calleja-González et al., 2023). Además un estudio elaborado por la UEFA señala que una menor carga de lesiones ($p=0.011$) y una mayor disponibilidad de jugadores para los partidos ($p=0.031$), en comparación con la temporada anterior, estuvieron asociadas con una mejor clasificación final en la liga (Häggglund et al., 2013). También se menciona que esto se ve reflejado en el ranking UEFA, es decir, en competiciones europeas. La consecuencia más evidente que puede acarrear para los clubes un peor rendimiento deportivo es un impacto económico negativo. En su estudio, López-Busto et al. (2016) señalan que en las instituciones de fútbol existe una correlación económica positiva con los partidos jugados en competiciones europeas y la posición final de los equipos en su liga doméstica. Además de esto, las lesiones tienen un impacto económico directo sobre los clubes o sus respectivas ligas. Por ejemplo, algunos estudios han calculado que cada lesión muscular supone unas pérdidas aproximadas a los clubes al mes de unos 365.812€ y al año alrededor de los 45 millones de libras por temporada debido al peor rendimiento de los jugadores después de una lesión (Eliakim et al., 2020; Torrejón et al., 2024). En La Liga, el coste total anual de las lesiones fue de 65,4 millones de euros para los equipos mejor clasificados Q1, y de 4,5 millones para los de menor clasificación Q4. Los equipos con mayor carga competitiva, como los que participan en múltiples torneos, sufren más lesiones debido a la congestión de partidos a corto plazo (Polvorinos, 2024). En adición a la información anterior observamos que la última tendencia en lesiones que más preocupa y que mayor repercusión negativa está

generando a los equipos es la rotura de ligamento cruzado anterior (LCA), lesión que requiere de un tiempo de recuperación muy amplio y que es más frecuente de lo esperado.

Durante el periodo de 2011/12 hasta 2017/18 en la primera división de la Liga Italiana se registraron un total de 84 lesiones de LCA, un ratio de incidencia de 0,062 por cada 1000 h (Grassi et al., 2019). Los equipos que mostraron una mayor incidencia de este tipo de lesiones fueron los que se localizaban en las primeras cuatro posiciones del campeonato. Es decir, aquellos que jugaron competiciones europeas y por lo tanto tenían un mayor volumen de partidos. Otro estudio relevante que abordó la incidencia de lesiones del LCA en el fútbol profesional es el de Grassi et al. (2020). Este estudio analizó las lesiones de LCA en la Serie A italiana durante siete temporadas consecutivas (2011-2012 a 2017-2018), encontrando un total de 84 lesiones. La tasa de incidencia general fue de 0,0618 por cada 1000 horas de juego, siendo significativamente mayor en partidos (0,4215/1000 h) que en entrenamientos (0,0305/1000 h), con un riesgo 14 veces superior en competición. Además, los equipos clasificados entre los cuatro primeros puestos presentaron una tasa de incidencia casi el doble (0,1256/1000 h) en comparación con los equipos del 5.º al 20.º lugar (0,0559/1000 h), lo que sugiere una relación entre el volumen de competición y el riesgo de lesión que reafirma la hipótesis de la alta carga de partidos y su relación directa con la probabilidad de lesión.

Esta problemática ha creado la necesidad de poder monitorear e intentar predecir cuando un jugador tiene mayor probabilidad de lesionarse y ha surgido gran interés en trabajar en modelos predictivos. La última tendencia en el ámbito deportivo y concretamente en el fútbol es el uso de inteligencia artificial (IA). En el caso de La Liga, se ha implementado el uso de la tecnología desarrollada por Microsoft a través de la plataforma Beyond Stats, la cual permite realizar predicciones en tiempo real sobre métricas avanzadas del juego como los goles esperados (xG) o la probabilidad de gol en una jugada. Cuanto mayor es la magnitud de la correlación en variables (xG, xGOT) más preciso es el resultado estimado (Ruiz-De-Alarcón-Quintero y De-La-Cruz-Torres, 2024). Gracias a estas métricas predictivas se observan las probabilidades de conversión de gol de un equipo basado en las ocasiones generadas por jugador y por equipo, valores que ayudan a los equipos a detectar su

efectividad de cara a gol basándose en los xG generados por partido (Narayanan y Pifer, 2023). Según estudios se puede observar como con algunos modelos de predicción como *Beyond xG* usando la recta de regresión de Ridge se obtuvo una mejora del rendimiento del modelo de un 12%. La mejora del modelo aumentó de un 83% a casi un 95% tal como indica la métrica *R-squared* (Malikov Kim, 2024). De igual forma que en fútbol profesional cada vez disponemos de mas datos relacionados con acciones técnicas, en los últimos años se ha observado que estos modelos que ya están implementados con eficiencia en el ecosistema del fútbol podrían, de manera escalable, ayudarnos a predecir con mayor exactitud la probabilidad de lesión de un jugador.

En consonancia con lo anterior, en los últimos años se ha observado una tendencia ascendente y cada vez una mayor popularidad el uso de dispositivos *wearables* como la tecnología de monitorización GPS. Estos dispositivos nos muestran datos de las variables del entrenamiento que nos ayudan a controlar la carga externa y la percepción de la carga interna durante un entrenamiento o partido. El objetivo de estos dispositivos es controlar estas cargas y ver la influencia en la misma de cada entrenamiento. El análisis de variables de carga externa como la distancia de carrera, aceleraciones, distancia de sprint y desaceleraciones son indicadores y carga interna como la frecuencia cardíaca y percepción de esfuerzo del entrenamiento nos reportan información para ver la influencia de la carga de trabajo de una sesión y nos generan información a los entrenadores sobre cómo controlar la carga de partido o de un entrenamiento (Grünbichler et al., 2019). Esta tecnología se está implementando en la preparación física para predecir riesgo lesional, para analizar demandas de tareas y percepción del esfuerzo. Para ello, se usan modelos de predicción de aprendizaje automático con una arquitectura profunda y en árbol donde se usan sistemas de validación de datos cruzados para comprobar la eficacia y rendimiento de los mismos. Algunos estudiosos modelos de aprendizaje profundo muestran un gran poder predictivo de la percepción del esfuerzo (RPE) en deportistas de élite con un margen de error de $1,08 \pm 0,07$ (Leppich et al., 2024). Otras vertientes en las que tiene peso el Aprendizaje automático es en el *scouting*. Mediante estos modelos de aprendizaje se analizan y predicen posibles comportamientos tácticos del rival. Un ejemplo de ello son los modelos que emplean

regresión en cascada para detectar puntos clave, los cuales, tras varios procesos de aprendizaje distribuidos en múltiples etapas de regresión, son capaces de generar el mapa de calor esperado (Xiaoliang et al., 2024).

2. Objetivos

Teniendo en cuenta todo lo anterior, el objetivo general de esta revisión es conocer cómo de eficiente es la IA a la hora de predecir lesiones en el fútbol de élite. De igual forma los objetivos específicos que se persiguen son los siguientes: i) Conocer las variables con mejor rendimiento a la hora de predecir lesiones, ii) Analizar los tipos de modelos utilizados y su eficacia predictiva y, iii) Descubrir las principales limitaciones de estos modelos predictivos

3. Metodología

3.1. Diseño

Para la realización de este estudio se llevó a cabo una revisión sistemática de la literatura publicada en las principales bases de datos. Para la realización de esta revisión se siguieron las indicaciones PRISMA (Page et al., 2021).

Cabe señalar que este objetivo general parte de la pregunta de investigación planteada inicialmente y estructurada de la siguiente forma: ¿Es eficiente la IA prediciendo lesiones en fútbol de élite masculino? A raíz de esta pregunta se hizo un análisis sobre las partes que la componían.

3.2. Estrategia de búsqueda

Una vez realizado este análisis se tradujo al lenguaje documental para poder realizar el proceso de búsqueda en las bases de datos de: MEDLINE Complete, Rehabilitation & Sports Medicine Source, SPORTDiscus with Full Text.

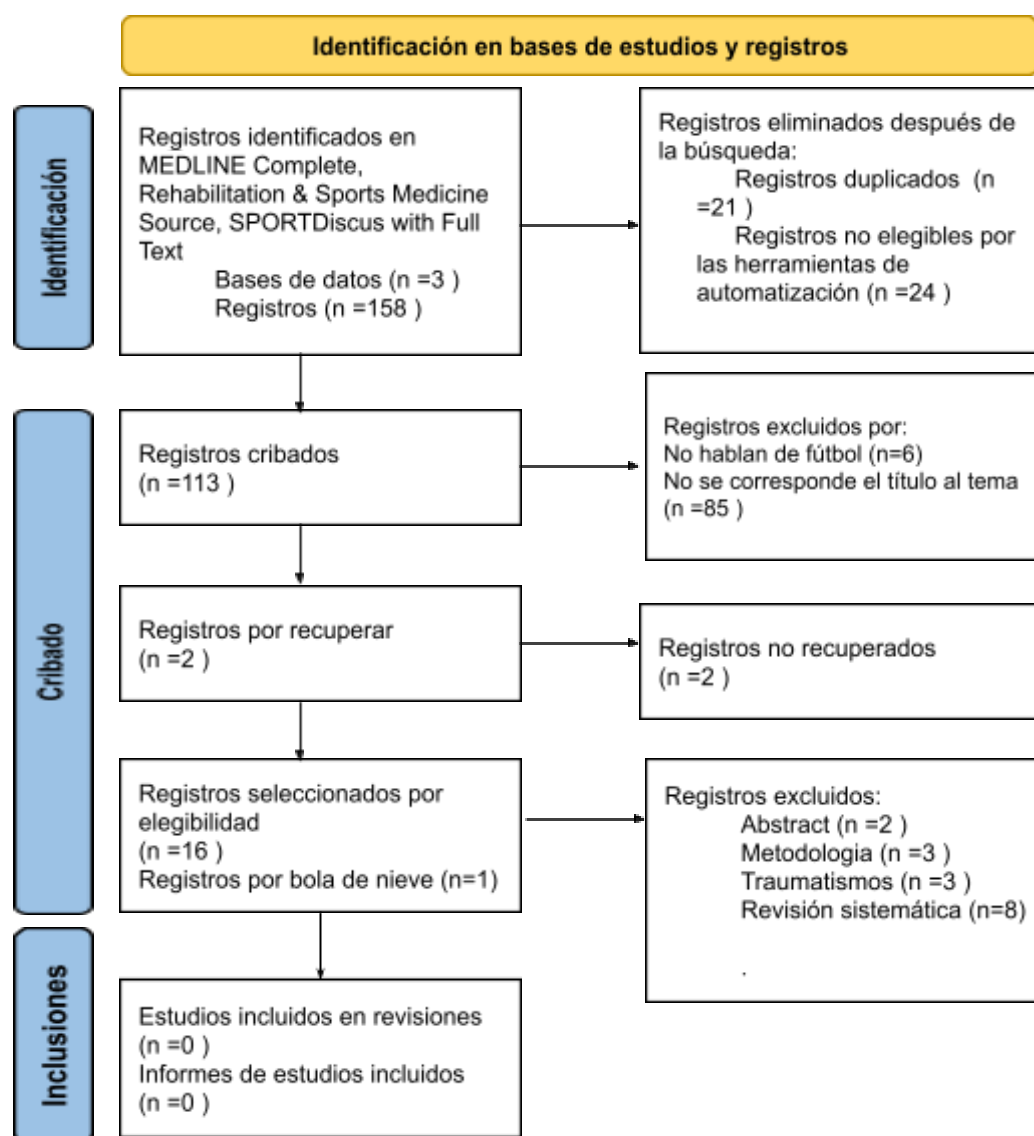
La ecuación de búsqueda utilizada fue: (Artificial intelligence OR machine learning OR deep learning) AND (football OR soccer) AND (injury OR injury prevention OR injury rate OR injury incidence OR prevention) AND (elite OR professional OR high performance). Con esta ecuación de búsqueda, el lenguaje documental y las bases de datos se llevó a cabo la búsqueda.

3.3. Criterios de selección

En aras de maximizar la representatividad de los artículos seleccionados, se establecieron los siguientes criterios de inclusión y exclusión. El único criterio de inclusión utilizado fue que fueran publicaciones académicas. Sin embargo como criterios de exclusión se añadieron: i) Publicaciones de Rugby, fútbol americano, fútbol australiano o NCAA, ii) Revisiones sistemáticas o meta-análisis y, iii) Impactos en la cabeza.

3.4. Diagrama de flujo

Figura 1.
Diagrama de flujo.



4. Discusión

Esta revisión sistemática tuvo como objetivo analizar la eficacia de la inteligencia artificial (IA) en la predicción de lesiones en el fútbol de élite así como profundizar en las metodologías empleadas para comprender cómo y porqué se generan los resultados. Un aspecto particularmente relevante de estos estudios es poder anticipar la aparición de lesiones en jugadores durante el próximo microciclo, lo que podría permitir optimizar la toma de decisiones antes de la competición especialmente en lo relativo a la gestión de cargas del entrenamiento. Para una adecuada interpretación de los hallazgos resulta esencial analizar detalladamente los procedimientos empleados en cada modelo. Comprender las etapas del proceso y la forma en que cada modelo construye sus predicciones constituye un punto clave para evaluar su validez, aplicabilidad y replicabilidad.

Los resultados obtenidos en esta revisión sistemática reflejan el considerable potencial de la inteligencia artificial (IA) como herramienta complementaria aportando valor especialmente en la detección temprana de riesgos y en la toma de decisiones basada en datos objetivos. en la prevención de lesiones en el fútbol de élite. No obstante, es importante señalar que la IA no sustituye el juicio clínico ni al análisis contextual que realizan los preparadores físicos y el equipo médico, sin embargo, podría posicionarse como una tecnología capaz de aportar valor en la detección temprana de riesgos y en la toma de decisiones basada en datos objetivos. Esta utilidad se ha visto reflejada especialmente en aquellos estudios que implementaron modelos complejos (Calderon et al., 2023; Freitas et al., 2025; Heksteden et al., 2023; Jauhianen et al. 2022; Pilka et al., 2023), entrenados con datos reales y bien estructurados, y que combinaron diferentes tipos de variables con un enfoque multidimensional. Podemos observar que varios de los estudios analizados en esta revisión predicen el riesgo de lesión con un nivel de precisión de más de un 80% (Abasi et al. 2025; Pilka et al. 2023; Roemmers et al. 2020; Saberisani et al. 2025). Observando estos porcentajes de precisión tenemos que atender al procedimiento llevado a cabo por dichos estudios. Por ejemplo, observamos como en el estudio de Abasi et al. (2025) se generaron casos aleatorios basados en la muestra real, lo cual permitió obtener valores tan elevados de

precisión como en este caso del 90,38%. Sin embargo, los resultados reportados con la muestra inicial no son positivos. Otros casos reflejan cómo los condicionantes que se tienen en cuenta para generar predicciones son menos específicos de lo deseado. Varios autores basan la predicción de lesión de su modelo en la probabilidad que tiene un jugador de sufrir una lesión a lo largo de una temporada sin especificar cuantos días antes es capaz de predecir el modelo (Heksteden et al., 2023; Jauhianen et al., 2022; Rommers et al., 2020). Información demasiado genérica y que no reporta datos concretos y de calidad para emplear y tomar decisiones en el fútbol de élite. En contraposición, hay estudios donde se reportan de manera específica esta probabilidad de lesión. Estudios indican la probabilidad de lesión diaria basada en datos de GPS (Freitas et al., 2025), probabilidad de lesiones con un rango de uno a cuatro semanas antes (Wagemans et al., 2023) y basado en la carga crónica de los últimos 28 días (Tsilimigkras et al., 2024).

4.1. Variables utilizadas en los modelos

Conocer las variables analizadas por cada autor y determinar cuales son las que más información aportan es fundamental para generar correlaciones con los modelos de predicción usados. En relación con las variables utilizadas en los modelos, no todas parecen tener el mismo peso o relevancia práctica. A simple vista, las desaceleraciones se muestran como una de las variables con más peso a la hora de anticipar una posible lesión (Pilka et al., 2023) con resultados de un 92% de precisión y estando presente en varios artículos. Vemos que los modelos con mayor rendimiento incluían repetidamente esta variable entre sus datos. Esto sugiere una clara relación entre los esfuerzos excéntricos y la probabilidad de sufrir una lesión, como por ejemplo muestra el estudio de Saberisani et al. (2025). Otra variable que también parecen aportar valor fueron los sprints repetidos y la distancia a alta intensidad, las cuales también muestran una fuerte asociación con el riesgo de lesión, especialmente cuando se considera su efecto acumulativo, como lo reporta Tsilimigkras et al. (2024), quien logró una precisión del 78% y un AUC de 0.85. Esto sugiere que no sólo la repetición de esfuerzos de alta intensidad es relevante así como su acumulación en el tiempo. Su impacto repetido durante una sesión se traduce en carga aguda y tras su acumulación en varias sesiones en carga crónica

lo que nos lleva a analizar la variable carga aguda-crónica (ACWR). La cual es de alta influencia, como indica Saberisani et al. (2025), donde alcanzó un AUC de 0.91, el más alto entre todas las variables analizadas. Este resultado refuerza la importancia del monitoreo de cargas de entrenamiento a lo largo de un macrociclo para prevenir picos de carga que puedan desencadenar lesiones. Junto a estas variables de carga externa, el historial de lesiones previas volvió a confirmarse como uno de los predictores más potentes. Su incorporación parece mejorar sustancialmente la capacidad de anticipar recaídas o lesiones similares (López-Valenciano et al., 2018) y en varios modelos la presencia de una lesión previa en una zona concreta aumentaba de forma significativa el riesgo lesional del jugador (Calderon-Diaz et al., 2023; Martins et al., 2022; Rossi et al., 2018).

Además, las variables neuromusculares y biomecánicas también mostraron tener gran potencial a la hora de detectar posibles perfiles de riesgo. Variables como la asimetría en fuerza isométrica, el control postural o los desequilibrios posturales fueron de gran interés para detectar perfiles funcionales de riesgo (Kolodziej et al., 2023; Wagemans et al., 2023).

Por último, algunos estudios también incorporaron variables contextuales y personales, como la edad, posición en el campo o el nivel competitivo del jugador. Aunque estas variables no destacaron de forma aislada, su combinación con datos de carga externa y el historial de lesiones mejoró la precisión del modelo, como se evidenció en el trabajo de Martins et al. (2022).

En conjunto, estos hallazgos muestran que, aunque las variables que mejores resultados muestran son los de carga externa (especialmente las desaceleraciones y los sprints), los modelos deben integrar múltiples dimensiones: variables de carga externa, carga interna, características biomecánicas, historial de lesiones además de variables personales y contextuales del juego.

Tabla 1.

Resumen de las variables analizadas en cada artículo y la eficacia predictiva de los modelos.

Variable	Modelos donde fue clave	Precisión / Impacto
Desaceleraciones	Pitka et al. (2023), Saberisani et al. (2025); Freitas et al. (2025); Tsilimigkras et al. (2024)	Pitka: Precisión 92%, F1 = 94.7% – mejor predictor en su estudio
Sprints y distancia en sprint	Tsilimigkras et al. (2024); Pitka et al. (2023); Freitas et al. (2025)	Tsilimigkras: Precisión 78%, AUC 0.85 – fuerte impacto acumulativo antes de lesión
Carga aguda-crónica (ACWR)	Tsilimigkras et al. (2024); Saberisani et al. (2025)	Saberisani: AUC 0.91 – variable con mayor influencia
Tiempo en zonas altas de frecuencia cardíaca	Tsilimigkras et al. (2024)	73% Sensibilidad – señal fisiológica de sobrecarga
Fuerza excéntrica de isquiotibiales	Calderón-Díaz et al. (2023); Wagemans et al. (2023)	XGBoost: Precisión 78%, riesgo elevado cuando hay desbalance o reducción
Desequilibrio de fuerza entre piernas	Wagemans et al. (2023)	Cuando se presenta 2 de 3 variables clave → riesgo de lesión sube al 50%
Posición del jugador	Freitas et al. (2025); Martins et al. (2022)	Freitas: SVM logró 74% precisión; posición influye en carga y tipo de lesión
Historial de lesiones previas	Rossi et al. (2018); Martins et al. (2022); Calderón-Díaz et al. (2023)	Rossi: AUC 0.76; historial relevante para identificar vulnerabilidad previa

4.2. Técnicas de Inteligencia Artificial utilizadas

Las técnicas de IA utilizadas en los diferentes estudios fueron diversas, si bien parece que las técnicas basadas en árboles de decisión presentaron la mayor eficacia global. Los algoritmos como XGBoost y Random Forest alcanzaron métricas de sensibilidad superiores al 85% además de altos valores de precisión y AUC. Esto puede ser debido a la capacidad de detectar relaciones entre variables y gestionar diferentes datos de múltiples dimensiones. Estos modelos han demostrado ser especialmente efectivos cuando incluían datos de carga externa de dispositivos GPS combinado con historial de lesiones, como en los estudios de Piłka et al. (2023) y Saberisani et al. (2025), esto sugiere una buena viabilidad de integración en los clubes y con las herramientas de las que actualmente disponen. Por otro lado, modelos como el Support Vector Machine (SVM) o las redes neuronales (FNN) han mostrado un rendimiento más variable. En casos observados como el estudio de Freitas et al. (2025), la sensibilidad fue aceptable pero la precisión se vio afectada por un elevado número de falsos positivos. En otros estudios como el de Haller et al. (2023), la sensibilidad fue muy baja (25%) en comparación con una alta exactitud global. En cuanto a los modelos de regresión penalizada (LASSO, Elastic Net, Ridge) su uso fue centrado en la predicción del número total de lesiones por jugador. Aunque su capacidad predictiva fue menor, pueden conseguir sentar las bases para descubrir factores de riesgo.

4.3. Modelos de IA utilizados y su capacidad predictiva

En los estudios revisados se hallaron una amplia variedad de modelos aplicados a la hora de predecir lesiones. En estos predominaban los basados en árboles de decisión, como Random Forest, XGBoost, Gradient Boosting y AdaBoost, por la capacidad que tienen de manejar grandes volúmenes de datos. En general, los modelos XGBoost alcanzaron un rendimiento, con precisiones superiores al 90%, sensibilidades de hasta 97,6% y áreas bajo la curva (AUC) cercanas o superiores a 0.90 (Piłka et al., 2023). Del mismo modo, el estudio de Saberisani et al. (2025) obtuvo una exactitud del 94,7% con variables únicamente de carga externa. Por otro lado, los modelos como Support Vector Machines (SVM) y las redes neuronales (FNN) ofrecieron resultados más variables, con sensibilidades entre el 25% y el 73%

(Freitas et al., 2025; Haller et al., 2023). En estos casos aunque la exactitud general fuera alta, la precisión real en la detección de lesiones fue baja, debido a un alto número de falsos negativos. Además, también se emplearon modelos de regresión penalizada (LASSO, Ridge y Elastic Net), particularmente en estimar el número total de lesiones esperadas por jugador. Estos ofrecieron buenos resultados en cuanto a error medio, con valores de RMSE inferiores a 0.6, como se reportó en Martins et al. (2022). Para concluir, algunos estudios optaron por explorar otros modelos como el Subgroup Discovery (Wagemans et al., 2023), este modelo no proporciona métricas tradicionales como sensibilidad o especificidad, pero sí permite identificar perfiles de riesgo específicos dentro de una población, lo que puede aportar valor clínico en la prevención de lesiones individualizadas.

4.4. Problemas metodológicos

Desde una perspectiva metodológica, uno de los principales problemas identificados fue el desequilibrio entre clases en los conjuntos de datos. Al tratarse de un fenómeno relativamente poco frecuente (la mayoría de los jugadores no se lesionan en cada microciclo o sesión), muchos modelos entrenan con una mayoría de ejemplos negativos, lo que distorsiona los resultados si no se aplican técnicas de sobremuestreo como SMOTE o undersampling. Esta situación genera una falsa sensación de precisión y perjudica especialmente la sensibilidad, haciendo que el modelo no detecte a tiempo a los jugadores con riesgo real. Es por ello que, además del rendimiento técnico del modelo, es necesario analizar la calidad del conjunto de datos y el tipo de métrica de validación utilizada. Esto indica que algunos modelos, si el conjunto de datos está desbalanceado (es decir, hay muchos más jugadores sanos que lesionados), tienden a acertar solo la clase mayoritaria. Este sesgo compromete la utilidad real de la IA en prevención, ya que los falsos negativos (jugadores que se lesionan pero no son identificados) son el error más crítico en este contexto. Otro factor limitante es la falta de estandarización en las métricas y protocolos. Algunos estudios reportan sensibilidad y especificidad; otros, precisión global o AUC; y otros utilizan errores medios como MAE o RMSE cuando el modelo es de regresión. Esta falta de uniformidad complica la comparación entre modelos y limita la posibilidad de realizar metaanálisis o evaluaciones conjuntas. Además, la

mayoría de los estudios se desarrollan en contextos retrospectivos, utilizando datos ya registrados, sin evaluar el impacto real del modelo en la práctica profesional. Los estudios que enfrentan problemas relacionados con el desequilibrio de clases y la falta de estandarización en las métricas han abordado el problema mediante la generación artificial de casos para aumentar la muestra de lesiones. López-Valenciano et al. (2017) aplicaron técnicas como SMOTE, ROS y RUS para equilibrar la proporción de ejemplos positivos y negativos, mejorando así la capacidad predictiva de los modelos. De forma similar, Jauhiainen et al. (2022) emplearon SMOTE junto con submuestreo aleatorio y ponderación de clases para compensar la baja incidencia de lesiones de LCA, aunque con resultados moderados (AUC entre 0.51 y 0.69). Por su parte, Rossi et al. (2018) utilizaron ADASYN para generar nuevos ejemplos sintéticos, lo que permitió incrementar la sensibilidad del modelo hasta un 80%. En una línea relacionada, Hecksteden et al. (2023) recurrieron al método ROSE para crear muestras balanceadas mediante interpolación de datos reales.

En cuanto a la aplicabilidad real pocos estudios han sido válidos en entornos de competición, lo que supone un reto importante para su implementación. Sin embargo, entre los estudios destacan por su alta aplicabilidad práctica sobresalen los trabajos de Piłka et al. (2023), Freitas et al. (2025) y Calderón-Díaz et al. (2023), ya que han sido desarrollados con datos reales de deportistas profesionales, incorporando variables fácilmente monitoreables en el día a día de los cuerpos técnicos, y validando sus modelos en contextos cercanos a la práctica profesional. Además, sus datos proceden de protocolos de screening ya integrados en muchos equipos, lo que facilita la adopción del modelo sin necesidad de implementar nuevos sistemas de medición. En conjunto, estos estudios representan propuestas sólidas y útiles para la práctica profesional, al ofrecer modelos predictivos basados en variables accesibles y validadas con datos longitudinales del deporte de élite. El resto de estudios han trabajado con modelos que se han construido con datos que no tienen disponibilidad permanente en los clubes o que requieren de equipamiento técnico y profesional para su recolección y análisis.

Pese a estas limitaciones observadas la inteligencia artificial ha demostrado que puede identificar patrones lesionales ocultos a simple vista y exprimir la información de las variables analizadas.

5. Contribuciones a los Objetivos de Desarrollo Sostenible

Este trabajo contribuye a los Objetivos de Desarrollo Sostenible (ODS), especialmente al ODS 3 promoviendo la salud y el bienestar de los futbolistas mediante la prevención de lesiones a través de la inteligencia artificial. Además, se alinea con el ODS 5, al poner de manifiesto la falta de representación del fútbol femenino en los modelos actuales y la necesidad de construir herramientas más inclusivas y equitativas. Por último, se relaciona con el ODS 9, ya que el desarrollo y aplicación de estas tecnologías muestra la brecha existente entre contextos de alto rendimiento y regiones del tercer mundo, donde la falta de acceso a infraestructuras tecnológicas limita la implementación de soluciones innovadoras en el deporte.

6. Futuras líneas de investigación

A pesar de los avances observados el campo de la predicción de lesiones aún presenta áreas críticas que deben ser exploradas. Un área fundamental a explorar sería la estandarización de los modelos y su replicabilidad y uso en contextos reales de entrenamiento y competición para permitir comparaciones rigurosas que evalúen su impacto real en la reducción de lesiones. Así mismo integrar esta tecnología en fútbol femenino y amateur sin necesidad de un coste elevado ni una gran cantidad de recursos contribuiría a generar una amplia base de datos multicéntrica y longitudinal que contribuyan a la preparación y entrenamiento de modelos más robustos y generalizables. Por último esta tecnología aún necesita mucha investigación y desarrollo por delante por lo que habría que incidir en cómo los modelos podrían mantener un aprendizaje constante y adaptarse en tiempo real a medida que recogen datos de un jugador mejorando su precisión a lo largo de la temporada y concretando el lapso de tiempo en el cual un jugador tiene alta probabilidad de lesión.

7. Conclusiones

En relación al objetivo general y a los objetivos específicos del estudio, las conclusiones obtenidas fueron:

7.1. Objetivo General

En conjunto, este trabajo permite responder al objetivo general de analizar la eficacia de la IA como herramienta para la prevención de lesiones en fútbol. Si bien hay resultados que muestran que estas técnicas podrían aportar valor, su efectividad depende de la calidad de los datos, tipos de variables y su diseño metodológico. Avanzar hacia modelos con mayor número de datos reales y con autonomía con respecto a técnicas de balanceo será clave para la mejora de la IA en la tarea de predecir lesiones.

7.2. Variables con mayor influencia

La discusión evidencia que las variables con mejores resultados son las desaceleraciones y el historial de lesiones. Las desaceleraciones particularmente están presentes en los modelos con mayor rendimiento, indicando su fuerte relación con las lesiones musculares. De igual forma, el historial médico mejora de manera significativa la capacidad del modelo de anticipar posibles recaídas.

7.3. Identificación de las técnicas de IA con mejor rendimiento:

Se concluye que los algoritmos basados en árboles de decisión, especialmente XGBoost y Random Forest, ofrecen los mejores resultados. Estos modelos destacan por su capacidad de interpretación de los datos siendo más eficaces que las redes neuronales o SVM. Además los modelos de regresión penalizada o técnicas como Subgroup Discovery han demostrado utilidad complementaria, pero no tan eficaz.

7.4. Limitaciones de la IA en la predicción de lesiones

Se identifican varios factores. El que más destaca es el desequilibrio entre clases dentro de los conjuntos de datos, ya que la mayoría de los jugadores no se lesionan con frecuencia. Esto indica que muchos modelos solo tienden a acertar cuando los

jugadores no se lesionan. Además se detecta una falta de estandarización de las métricas para poder comparar diversos estudios.

8. Referencias

- Abasi, A., Fatemi Aghda, S. A., Moezy, A., & Nazari, A. (2025). Predicting soccer injuries using machine learning models: A study on professional football players in Iran. *Journal of Sports Science & Medicine*, 25(3), 123-135. <https://doi.org/10.1186/s13040-025-00431-2>
- Calderón-Díaz, M., Aguirre, R. S., Váscquez, J. P., Yáñez, R., Roby, M., Querales, M., & Salas, R. (2023). Biomechanical analysis and machine learning models for identifying biomarkers of muscle injuries in soccer players. *Sports Biomechanics*, 22(1), 75-89. <https://doi.org/10.3390/s24010119>
- Calleja-González, J., Mallo, J., Cos, F., Sampaio, J., Jones, M. T., Marqués-Jiménez, D., Mielgo-Ayuso, J., Freitas, T. T., Alcaraz, P. E., Vilamitjana, J., Ibañez, S. J., Cuzzolin, F., Terrados, N., Bird, S. P., Zubillaga, A., Huyghe, T., Jukic, I., Lorenzo, A., Loturco, I., ... Lago-Peñas, C. (2023). A commentary of factors related to player availability and its influence on performance in elite team sports. *Frontiers In Sports And Active Living*, 4. <https://doi.org/10.3389/fspor.2022.1077934>
- Diniz, P., Abreu, M., Lacerda, D., Martins, A., Pereira, H., Ferreira, F. C., Kerkhoffs, G. MMJ., A, Fred. (2024). Exploring pre- and post-Achilles tendon rupture match data and machine learning models to predict future injuries. *Journal of Sports Medicine and Physical Fitness*, 64(2), 110-121. <https://doi.org/10.1007/s00167-022-07082-4>
- Eliakim, E., Morgulev, E., Lidor, R., & Meckel, Y. (2020). Estimation of injury costs: financial damage of English Premier League teams' underachievement due to injuries. *BMJ Open Sport & Exercise Medicine*, 6(1), e000675. <https://doi.org/10.1136/bmjsem-2019-000675>
- FIFPRO. (2024). *Informe anual de monitoreo de la carga de trabajo de los futbolistas 2023–2024*. FIFPRO. https://fifpro.org/media/zymmarhy/es-pwm_men-s-annual-report_digital_spanish.pdf

- Freitas, D. N., Mostafa, S. S., Caldeira, R., Santos, F., Fermé, E., Gouveia, É. R., & Morgado-Dias, F. (2025). Predicting noncontact injuries of professional football players using machine learning. *PLoS ONE*, 20(1), e0315481. <https://doi.org/10.1371/journal.pone.0315481>
- Grassi, A., Macchiarella, L., Filippini, M., Lucidi, G. A., Della Villa, F., & Zaffagnini, S. (2020). Epidemiology of anterior cruciate ligament injury in Italian first division soccer players. *Sports Health*, 12(3), 279–288. <https://doi.org/10.1177/1941738119885642>
- Grassi, A., Macchiarella, L., Filippini, M., Lucidi, G. A., Della Villa, F., & Zaffagnini, S. (2019b). Epidemiology of Anterior Cruciate Ligament Injury in Italian First Division Soccer Players. *Sports Health A Multidisciplinary Approach*, 12(3), 279-288. <https://doi.org/10.1177/1941738119885642>
- Grünbichler, J., Federolf, P., & Gatterer, H. (2019b). Workload efficiency as a new tool to describe external and internal competitive match load of a professional soccer team: A descriptive study on the relationship between pre-game training loads and relative match load. *European Journal Of Sport Science*, 20(8), 1034-1041. <https://doi.org/10.1080/17461391.2019.1697374>
- Häggglund, M., Waldén, M., Magnusson, H., Kristenson, K., Bengtsson, H., & Ekstrand, J. (2013). Injuries affect team performance negatively in professional football: an 11-year follow-up of the UEFA Champions League injury study. *British Journal Of Sports Medicine*, 47(12), 738-742. <https://doi.org/10.1136/bjsports-2013-092215>
- Haller, N., Kranzinger, S., Kranzinger, C., Blumkaitis, J. C., Strepp, T., Simon, P., Tomaskovic, A., O'Brien, J., Düring, M., & Stöggl, T. (2023). Predicting Injury and Illness with Machine Learning in Elite Youth Soccer: A Comprehensive Monitoring Approach over 3 Months. *Journal Of Sports Science And Medicine*, 476-487. <https://doi.org/10.52082/jssm.2023.476>.

- Hecksteden, A., Schmartz, G. P., Egyptien, Y., Fünten, K. A. D., Keller, A., & Meyer, T. (2022). Forecasting football injuries by combining screening, monitoring and machine learning. *Science And Medicine In Football*, 7(3), 214-228. <https://doi.org/10.1080/24733938.2022.2095006>
- Howden Group. (2023). El fútbol de élite enfrenta un calendario cada vez más exigente. Recuperado de <https://www.howdengroup.com/es-es/el-futbol-de-elite-enfrenta-un-calendario-cada-vez-mas-exigente>.
- Jauhiainen, S., Kauppi, J., Krosshaug, T., Bahr, R., Bartsch, J., & Äyrämö, S. (2022). Predicting ACL Injury Using Machine Learning on Data From an Extensive Screening Test Battery of 880 Female Elite Athletes. *The American Journal Of Sports Medicine*, 50(11), 2917-2924. <https://doi.org/10.1177/03635465221112095>
- Kolodziej, M., Groll, A., Nolte, K., Willwacher, S., Alt, T., Schmidt, M., & Jaitner, T. (2023). Predictive modeling of lower extremity injury risk in male elite youth soccer players using LASSO Regression. *Scandinavian Journal Of Medicine And Science In Sports*, 33(6), 1021-1033. <https://doi.org/10.1111/sms.14322>
- Leppich, R., Kunz, P., Bauer, A., Kounev, S., Sperlich, B., & Düking, P. (2024). Prediction of Perceived Exertion Ratings in National Level Soccer Players Using Wearable Sensor Data and Machine Learning Techniques. *Journal Of Sports Science And Medicine*, 744-753. <https://doi.org/10.52082/jssm.2024.744>
- López-Busto, A., García-Unanue, J., Gomez-Gonzalez, C., Barajas, Á., & Gallardo, L. (2016). Incidencia De Los Resultados Deportivos, Las Variables Económicas Y Administrativas En El Rendimiento Financiero De Los Clubes De Fútbol (Sports Scores, Financial and Administrative Variables on the Financial Performance of Football Clubs). *SSRN Electronic Journal*. https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2867326_code373446.pdf?abstractid=2867326&mirid=1&type=2

- López-Valenciano, A., Ayala, F., Puerta, J. M., De Ste Croix, M. B. A., Vera-Garcia, F. J., Hernández-Sánchez, S., Ruiz-Pérez, I., & Myer, G. D. (2017). A Preventive Model for Muscle Injuries. *Medicine & Science In Sports & Exercise*, 50(5), 915-927. <https://doi.org/10.1249/mss.0000000000001535>
- Malikov, D., & Kim, J. (2024). Beyond xG: A Dual Prediction Model for Analyzing Player Performance Through Expected and Actual Goals in European Soccer Leagues. *Applied Sciences*, 14(22), 10390. <https://doi.org/10.3390/app142210390>
- Mandorino, M., Clubb, J., & Lacombe, M. (2024). Predicting Soccer Players' Fitness Status Through a Machine-Learning Approach. *International Journal Of Sports Physiology And Performance*, 19(5), 443-453. <https://doi.org/10.1123/ijsp.2023-0444>
- Martins, F., Przednowek, K., França, C., Lopes, H., De Maio Nascimento, M., Sarmento, H., Marques, A., Ihle, A., Henriques, R., & Gouveia, É. R. (2022). Predictive Modeling of Injury Risk Based on Body Composition and Selected Physical Fitness Tests for Elite Football Players. *Journal Of Clinical Medicine*, 11(16), 4923. <https://doi.org/10.3390/jcm11164923>
- Narayanan, S., & Pifer, N. D. (2023). An xG of Their Own: Using Expected Goals to Explore the Analytical Shortcomings of Misapplied Gender Schemas in Football. *Journal Of Sport Management*, 38(2), 92-109. <https://doi.org/10.1123/jsm.2023-0022>
- Piłka, T., Grzelak, B., Sadurska, A., Górecki, T., & Dyczkowski, K. (2023). Predicting Injuries in Football Based on Data Collected from GPS-Based Wearable Sensors. *Sensors*, 23(3), 1227. <https://doi.org/10.3390/s23031227>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W.,

- Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, n71. <https://doi.org/10.1136/bmj.n71>
- Polvorinos, J. (2024). Lesiones y costes económicos: cuánto cuesta una lesión en LaLiga. Recuperado de <https://www.jorgepolvorinos.com/post/lesiones-y-costes-economicos-cuanto-cuesta-una-lesion-en-la-liga>
- Rommers, N., Rössler, R., Verhagen, E., Vandecasteele, F., Verstockt, S., Vaeyens, R., Lenoir, M., D'hondt, E., & Witvrouw, E. (2020). A Machine Learning Approach to Assess Injury Risk in Elite Youth Football Players. *Medicine & Science In Sports & Exercise*, 52(8), 1745-1751. <https://doi.org/10.1249/mss.0000000000002305>
- Rossi, A., Pappalardo, L., Cintia, P., Iaia, F. M., Fernández, J., & Medina, D. (2018). Effective injury forecasting in soccer with GPS training data and machine learning. *PLoS ONE*, 13(7), e0201264. <https://doi.org/10.1371/journal.pone.0201264>
- Ruiz-De-Alarcón-Quintero, A., & De-La-Cruz-Torres, B. (2024). An Expected Goals on Target (xGOT) Metric as a New Metric for Analyzing Elite Soccer Player Performance. *Data*, 9(9), 102. <https://doi.org/10.3390/data9090102>
- Saberisani, R., Barati, A. H., Zarei, M., Santos, P., Gorouhi, A., Ardigò, L. P., & Nobari, H. (2025). Predictive capacity of external load variables on injury risk in Iranian professional soccer players using decision trees. *Journal of Football Science*, 22(4), 134-145. <https://doi.org/10.3389/fspor.2025.1425180>
- Torrejón, L. N., Martínez-Serrano, A., Villalón, J. M., & Alcaraz, P. E. (2024). Economic impact of muscle injury rate and hamstring strain injuries in professional football clubs. Evidence from LaLiga. *PLoS ONE*, 19(6), e0301498. <https://doi.org/10.1371/journal.pone.0301498>

- Tsilimigkras, T., Kakkos, I., Matsopoulos, G. K., & Bogdanis, G. C. (2024). Enhancing Sports Injury Risk Assessment in Soccer Through Machine Learning and Training Load Analysis. *Journal Of Sports Science And Medicine*, 537-547. <https://doi.org/10.52082/jssm.2024.537>
- Wagemans, J., De Leeuw, A., Catteeuw, P., & Vissers, D. (2023). Development of an algorithm-based approach using neuromuscular test results to indicate an increased risk for non-contact lower limb injuries in elite football players. *BMJ Open Sport & Exercise Medicine*, 9(2), e001614. <https://doi.org/10.1136/bmjsem-2023-001614>
- Xiaoliang Ju, Xiyao Chen, Yu Zhang, Qishen Zhou, & Jie Zhao. (2024). Tactical Optimisation and Opponent Analysis for Football Teams with Big Data Mining and Machine Learning. *International Journal of Medicine & Science of Physical Activity & Sport / Revista Internacional de Medicina y Ciencias de La Actividad Física y Del Deporte*, 24(94), 371–392. <https://doi.org/10.15366/rimcafd2024.94.024>

9. Anexos

9.1. Tabla 1. Cuadro de Artículos empleados

Autor y año	Objetivos	Muestra	Variables	Materiales y Métodos	Conclusión
Abasi, A.; Fatemi Aghda, S. A.; Moezy, A.; Nazari, A. (2025).	Evaluar la eficacia de distintos modelos de aprendizaje automático para predecir lesiones en futbolistas profesionales iraníes.	256 futbolistas profesionales de 16 equipos nacionales y de primera división en Irán.	Datos fisiológicos derivados de CPET, incluyendo frecuencia cardíaca máxima (HRmax), frecuencia cardíaca a los 2 minutos de recuperación (HR2), consumo máximo de oxígeno (VO2max), umbral anaeróbico, historial de conmociones cerebrales, entre otras	modelos de aprendizaje automático como CatBoost, SVM, Random Forest y XGBoost. métricas como precisión, recall, F1-score, AUC y valores SHAP	CatBoost logró 89% de precisión, AUC = 0.92, F1-score = 0.87. Variables más influyentes: HR2, VO2max, y antecedentes de conmoción cerebral.
Calderón-Díaz, M., Aguirre, R. S., Váscquez, J. P., Yáñez, R., Roby, M., Querales, M., & Salas, R. (2023).	Identificar los biomarcadores de las lesiones musculares en jugadores a través de una análisis biomecánico y analizar mediante modelos de aprendizaje.	110 jugadores profesionales de futbol masculino. IMC<42	Nordic Hamstring : Fuerza excéntrica del cuádriceps: Puente de una pierna: Miotonometría: Prueba Bosco:	Modelo XGBoost SVM Árbol de decisión KNN Recta de regresión logística	PRECISIÓN Modelo XGBoost - 78% SVM - 73% Árbol de decisión - 73% KNN - 73% Recta de regresión logística - 70%

Diniz, P.; Abreu, M.; Lacerda, D.; Martins, A.; Pereira, H.; Ferreira, F.C.; Kerkhoffs, G. MMJ.; A, Fred. (2024).	Análisis de los datos exploratorios de partido antes y después de rotura de tendón de aquiles (ATR) y evaluación de modelo de aprendizaje automático para prever si un jugador se volverá a lesionar.	Jugadores que sufrieron ATR entre enero y marzo de 2021 en primera y segunda división en la Liga Alemana.	MPM 1 año antes de lesión MPM 1-2 años después de la lesión	Agrupamiento de K.medias Algoritmo de clasificación XGBoost Receptor AUROC Pérdida de puntuación de Brier (BSL)	Mediocentros y defensores incrementaron su participación en partidos durante el 1er año post lesión. Los porteros a los 2 años. Buen rendimiento de la herramienta de clasificación ML para predecir la participación en los partidos después de un ATR.
Freitas, D. N.; Mostafa, S. S.; Caldeira, R.; Santos, F.; Fermé, E.; Gouveia, É. R.; Morgado-Dias, F. (2025).	Desarrollar un sistema automatizado basado en <i>machine learning</i> para predecir el riesgo de lesiones sin contacto en jugadores de fútbol profesional, utilizando datos de GPS y variables descriptivas.	34 jugadores profesionales de un equipo de la Primera División de Portugal, analizados durante 36 semanas en la temporada 2020-2021. Se recopilaron datos de 217 entrenamientos y 38 partidos oficiales	Datos GPS: Velocidad, aceleración, carga de entrenamiento. Variables descriptivas: Posición del jugador, tipo de sesión (entrenamiento o partido), día de la semana, resultado de partido, ubicación del partido.	Se aplicaron modelos de <i>machine learning</i> , incluyendo Support Vector Machines (SVM), Feedforward Neural Networks (FNN) y Adaptive Boosting (AdaBoost). Se utilizó un método de selección de características (<i>Maximum Relevance – Minimum Redundancy, mRMR</i>) para optimizar las variables más relevantes. Se entrenaron	El modelo AdaBoost logró una sensibilidad del 78.57% y una especificidad del 65.02%. El modelo SVM obtuvo una sensibilidad del 71.43% y una especificidad del 74.19%, siendo el más equilibrado. Ambos modelos demostraron estabilidad y potencial para su aplicación práctica en la gestión del riesgo de lesiones en el fútbol profesional.

				modelos con aprendizaje sensible al costo (<i>cost-sensitive learning</i>) para mejorar la predicción de lesiones.	
Haller, N.; Kranzinger, S.; Kranzinger, C.; Blumkaitis, J. C.; Strepp, T.; Simon, Perikles.; Tomaskovic, A.; O'Brien, J.; Stöggl, T. (2023).	Evaluar la viabilidad de un enfoque de monitoreo integral y determinar la precisión de modelos de <i>machine learning</i> para predecir lesiones en futbolistas juveniles de élite.	25 jugadores de una academia juvenil de fútbol europeo (edad: 16.6 ± 0.9 años) durante 3 meses.	Carga de entrenamiento, datos de condición física, cuestionarios, biomarcadores sanguíneos, pruebas de rendimiento neuromuscular (salto CMJ), y fuerza de isquiotibiales y abductores/aductores de cadera.	Se utilizó un modelo de <i>Support Vector Machine</i> (SVM) para predecir lesiones basándose en los datos recopilados durante la temporada.	Aunque el modelo SVM mostró precisión en la detección de enfermedades, tuvo una baja capacidad predictiva para lesiones, con un alto número de falsos positivos. Se requieren más datos para mejorar la precisión de los modelos.
Hecksteden, A.; Schmartz, G.P.; Egyptian, Y.; Aus der Fünten, K.; Keller, A.; Meyer, T. (2023).	Medir la precisión de modelos de Machine Learning al combinar datos de screening (evaluaciones periódicas) y monitorización para predecir lesiones.	88 jugadores de 4 equipos de las tercera y cuarta divisiones alemanas. Se registraron 51 lesiones sin contacto.	Datos de screening: Edad, posición, lesiones, capacidad aeróbica, composición corporal, velocidad en sprint, calidad del movimiento. Datos de monitoreo diario: RPE, estrés y recuperación subjetiva, tipo de sesión.	Modelo Gradient Boosting con validación cruzada leave-one-out ROSE upsampling para equilibrar los datos y mejorar la detección de lesiones.	El modelo alcanzó un Área Bajo la Curva (AUC-ROC) de 0,61 en validación cruzada y un 0,62 en el test set La combinación de screening y monitoreo mejoró los resultados.

Jauhiainen, S.; Kauppi, J.-P.; Krosshaug, T.; Bahr, R.; Bartsch, J.; Åyrämö, S. (2022).	Evaluar la capacidad predictiva de diferentes métodos de machine learning para identificar el riesgo de lesión del ligamento cruzado anterior (LCA) en futbolistas y jugadoras de balonmano de élite Analizar el impacto de la variabilidad aleatoria en los resultados predictivos	880 atletas femeninas de élite (451 futbolistas y 429 jugadoras de balonmano) Datos recolectados entre 2007 y 2015 60 casos de lesión de LCA sin contacto	Factores biomecánicos: Análisis 3D del movimiento en pruebas de salto y cambios de dirección. Factores neuromusculares: Fuerza isocinética, estabilidad articular, test de equilibrio. Factores antropométricos: Altura, peso, índice de masa corporal (IMC). Historial de lesiones.	4 clasificadores de machine learning: Regresión logística regularizadora L2. Random Forest. Máquina de soporte vectorial (SVM) lineal. SVM con núcleo no lineal. Validación cruzada de 5-fold Técnicas como SMOTE, submuestreo aleatorio y ponderación de clases.	Mejor modelo: SVM lineal con un AUC-ROC promedio de 0.63 Rango de variación de AUC-ROC: 0.51 - 0.69 (variabilidad alta en resultados) Estrategias de balanceo de clases (SMOTE, submuestreo, ponderación) no mejoraron la capacidad predictiva.
Kolodziej, M., Groll, A., Nolte, K., Willwacher, S., Alt, T., Schmidt, M., & Jaitner, T. (2023).	Identificar los factores de riesgo neuromusculares y biomecánicos en jugadores de fútbol jóvenes y comprobar la capacidad de predicción del machine learning	6 equipos sub 16, sub 17 y sub 19 pertenecientes a 2 academias de fútbol profesional alemanas. Media de edad de 17,2 años, 179 cm altura +-8 cm y 70,4 kg +- 9,2 kg)	Análisis de movimiento 3D (SLDL y USCC) Control postural (DPSI) con 3 pruebas unilaterales Resistencia (ISomed2000) Flexión extensión y rotación isométrica	Modelo de regresión LASSO y Modelo LOO	Sensibilidad: 35% Especificidad: 79% AUC: 0.63

			de tronco. Abd y ad de cadera Extensión y flexión de rodilla		
López-Valenciano, A., Ayala, F., Puerta, J. M., De Ste Croix, M. B. A., Vera-Garcia, F. J., Hernández-Sánchez, S., Ruiz-Pérez, I., & Myer, G. D. (2017).	Análisis y comparativa del comportamiento de diferentes métodos de aprendizaje para la posterior selección del mejor modelo para identificar riesgo lesional muscular en MMII	132 jugadores profesionales de fútbol ligas nacionales en 1a(25) y 2a división B (73) española y 34 jugadores masculinos profesionales de balonmano en el 1er y 3er equipo de liga nacional española.	Factores de riesgo personales Factores de riesgo psicológico Factores de riesgo neuromuscular Control postural dinámico Fuerza isométrica de abd y ad de cadera Rango de movimiento de articulaciones de mmii Estabilidad del núcleo Fuerza isocinética de flexión y extensión de rodilla.	Análisis estadístico con SPSS 21.0 Algoritmo Weka de discretización no supervisada ADTree , RandomTree y Simplecart Sobremuestreo sintético SMOTE, aleatorio ROS y submuestreo aleatorio RUS	Sensibilidad 65.9% Especificidad 79.1% AUC 0.747
Mandorino, M.; Clubb, J.; Lancome, M. (2024).	Desarrollar una herramienta en machine learning para predecir el estado físico de los jugadores de fútbol	50 jugadores de un club profesional italiano. 3 categorías: Primer equipo (22), Equipo sub 19 (18) y Equipo	Carga externa: Distancia recorrida, velocidad máxima, nº de aceleraciones y desaceleraciones, ratio trabajo-descanso.	Varios modelos de machine learning (Random Forest Regression, Extreme Gradient Boosting, Linear Regression) Índice de Condición	Random Forest mejor desempeño (MAE = 3.8; RMSE = 5.11) Índice de Condición Física mostró una correlación muy alta ($r = 0.70$, $p < 0,01$)

	Evaluar la validez de la herramienta en comparación con una prueba de carrera submáxima (SMFT) Analizar el impacto de la carga de entrenamiento semanal sobre el índice y los resultados del SMFT.	sub 18 (10)	Carga interna: Frec. card	Física: Con valores de HR reales.	
Martins, F.; Przednowek, K.; França, C.; Lopes, H.; Nascimento, M. de M.; Sarmiento, H.; Marques, A.; Ihle, A.; Henriques, R.; Gouveia, É.R. (2022).	Analizar riesgo de lesión en jugadores de fútbol de élite utilizando modelos predictivos basados en composición corporal y pruebas de condición física. Evaluar cuáles son las variables más relevantes.	36 jugadores de fútbol profesional de la Primera Liga Portuguesa. Datos recopilados entre la temporada 2020/2021.	Información de los jugadores: Posición, edad, experiencia, historial de lesiones. Composición corporal: Altura, peso, masa grasa (BFM), agua corporal total (TBW), masa libre de grasa (FFM). Condición física: Flexibilidad (sit-and-reach), fuerza de agarre manual, número de flexiones en 1	Diferentes modelos de regresión: Regresión de mínimos cuadrados ordinarios (OLS) Regresión Ridge y LASSO Elastic Net Validación cruzada leave-one-out	Mejor modelo: Regresión Ridge con Elastic Net. Error cuadrático medio (RMSE): 0.592 (baja imprecisión en la predicción de lesiones) Predictores principales: Posición (Delanteros y defensores con mayor riesgo) Altura Corporal Sit-and-reach

			minuto, velocidad en 5m, 10m y 35m, VO₂máx estimado, test Yo-Yo. Variable dependiente: Número de lesiones por temporada.		
Piłka, T.; Grzelak, B.; Sadurska, A.; Górecki, T.; Dyczkowski, K. (2023).	Desarrollar un modelo de predicción de lesiones basado en datos recopilados mediante sensores GPS en jugadores de fútbol profesional.	36 jugadores de la PKO BP Ekstraklasa en las temporadas 2020/2021 y 2021/2022 Entrenamientos y partidos con GPS Catapult.	Carga de entrenamiento: Distancia total, sprint (> 25.2 km/h), aceleraciones y desaceleraciones. Historial de lesiones. Datos contextuales: Minutos jugados, número de partidos y microciclos de entrenamiento	Tres modelos diferentes: 1. Modelo basado en reglas 2. Modelo basado en lógica difusa 3. Modelo de machine learning (XGBoost)	Modelo Reglas: Precisión: 53% Sensibilidad: 58% Exactitud: 76% Modelo Difuso: Precisión 15.9% Sensibilidad: 22.6% Exactitud: 83.6% Modelo ML: Precisión: 92% Sensibilidad: 97.6% Exactitud: 90% F1-score: 94.7% El mejor predictor fueron las desaceleraciones y la distancia recorrida en sprint
Rommers, N.; Rössler, R.; Verhagen, E.; Vandecasteele, F.; Verstocket, S.; Vaeyens, R.; Lenoir, M.; D'Hondt,	Evaluar el riesgo de lesiones en futbolistas juveniles de élite mediante un	734 jugadores de fútbol juvenil (U10 a U15) de 7 academias de Bélgica.	Datos antropométricos: Altura, peso, longitud de piernas.	Se utilizó el algoritmo XGBoost. Se clasificaron en sobrecarga o agudas	El modelo predijo con una precisión 85%, sensibilidad 85% y exactitud del 85%. También logró clasificar

E.; Witvrouw, E. (2020).	modelo de machine learning, basado en medidas antropométricas, de coordinación motora y de rendimiento físico.	Se realizó un seguimiento durante una temporada completa (2017-2018)	Coord. Motora: Pruebas de salto, agilidad y dribbling. Rendimiento físico: Fuerza, flexibilidad, velocidad y resistencia.		lesiones como sobrecarga o agudas con precisión del 78%.
Rossi, A.; Pappalardo, L.; Cintia, P.; Iaia, F. M.; Fernández, J.; Medina, D. (2018).	Desarrollar un modelo predictivo basado en datos de GPS y machine learning para evaluar el riesgo de lesión.	26 jugadores de club profesional italiano Seguimiento de 23 semanas en la 2013/2014 931 sesiones de entrenamiento y 23 sesiones sin contacto	Carga de entrenamiento: Distancia total, distancia a alta velocidad, carga metabólica y número de aceleraciones/desaceleraciones. Historial de lesiones Datos personales: edad, imc, posición del campo, minutos jugados en partidos oficiales	Modelo predictivo usando árboles de decisión (Decision Tree, DT) Validación cruzada estratificada como Random Forest (RF) y regresión logística (LR) Se usó sobremuestreo ADASYN	Sensibilidad (Recall) del modelo DT: 80% Especificidad: No lo explica pero dice que el modelo redujo significativamente los falsos positivos. Precisión del modelo DT: 50% AUC-ROC: 0.76
Saberisani, R.; Barati, A. H.; Zarei, M.; Santos, P.; Gorouhi, A.; Ardigò, L. P.; Nobari, H. (2025).	Evaluar la capacidad predictiva de las variables de carga externa en el riesgo de lesión profesional iraníes basado en árboles	25 jugadores de un equipo de la Persian Gulf Pro League durante la temporada 2022 - 2023. Se analizaron 214	Carga de entrenamiento: Ratio carga aguda-crónica. Datos GPS: Distancia total, distancia recorrida a alta y media	Se utilizó árbol de decisión, XGBoost con validación del 80% de los datos para entrenamiento y 20% de datos para prueba.	Sensibilidad: 87,5% Precisión: 58,3% Exactitud: 94.7% AUC-ROC: 0,91 Menciona que la variable más importante fue el número de

	de decisión (DT)	sesiones de entrenamiento y 34 partidos oficiales.	velocidad, carga total de distancia, aceleraciones y desaceleraciones. Historial de lesiones		desaceleraciones.
Tsilimigkras, T.; Kakkos, I.; Matsopoulos, G. K.; Bogdanis, G. C. (2024).	Evaluar la eficacia del uso de modelos de <i>machine learning</i> en la predicción del riesgo de lesiones musculares sin contacto en futbolistas profesionales.	25 jugadores profesionales masculinos del Asteras Tripolis F.C. (Superliga de Grecia) con una lesión muscular sin contacto en un período de 4 años (2021-2024).	Se analizaron 9 métricas de carga externa (sprints, distancia total, aceleraciones, etc.) e interna (frecuencia cardíaca y tiempo en zonas de esfuerzo).	Se utilizó <i>machine learning</i> para evaluar los cambios en la carga de entrenamiento en los 28 días previos a la lesión y compararlos con períodos sin lesión. Se aplicaron modelos SVM (Máquinas de Soporte Vectorial) para la clasificación del riesgo de lesión con validación cruzada.	El modelo de predicción alcanzó una precisión del 78%, con una sensibilidad del 73% y una especificidad del 85%, lo que indica su potencial para evaluar el riesgo de lesiones musculares. Se destacó que aumentos repentinos en la distancia de carrera a alta velocidad, la cantidad de sprints y las métricas de sobrecarga fisiológica pueden ser predictores clave de lesiones.
Wagemans, J.; De Leeuw, A.-W.; Catteeuw, P.; Vissers, D. (2023)	Explorar la viabilidad de un enfoque de algoritmos de detección de subgrupos utilizando pruebas neuromusculares para identificar jugadores con riesgo de lesión sin contacto en	77 jugadores profesionales de un equipo de la Primera División de Bélgica. Temporada 2020-2021 hasta noviembre de 2022. Se registraron 92 lesiones.	Fuerza excéntrica de isquiotibiales. Fuerza isométrica de aductores y abductores de cadera. Prueba de salto con contramovimiento (CMJ).	Recolección de datos con dispositivos de VALD Performance. 4 momentos antes de la lesión: T4 (4 semanas), T3 (3 semanas), T2 (2 semanas) y T1 (1 semana antes). Algoritmo de	3 predictores clave de lesión: 1. Desbalance de fuerza en abducción $\geq 97\%$ del valor basal, tres semanas antes de la lesión. 2. Fuerza de aducción en pierna derecha ≤ 1.01 veces el valor basal, una semana antes de la lesión.

	miembros inferiores.		Desbalance de fuerza entre extremidades.	detección de subgrupos (Subgroup Discovery) Técnicas de imputación de datos faltantes	3. Reducción en la fuerza de aterrizaje en la pierna izquierda, cuatro semanas antes de la lesión. Cuando dos de estos estaban presentes el riesgo de lesión aumentaba al 50%, comprado con el 33% de la muestra global
--	----------------------	--	--	--	--