

Predictive model to prevent injuries in athletes using Machine Learning and sensitivity analysis

Christian Ovalle¹, Everardo Sánchez Puche², Waldyr Cerron Valverde³, Omar Santillán Achings⁴

¹Department of Engineering, Universidad Tecnológica del Perú, Lima, Perú, C19349@utp.edu.pe

²Department of Engineering, Universidad Popular del Cesar, Valledupar, Colombia, everardomsanchez@unicesar.edu.co

³Department of Engineering, Universidad Tecnológica del Perú, Lima, Perú, C21262@utp.edu.pe

⁴Department of Engineering, Universidad Tecnológica del Perú, Lima, Perú, C25611@utp.edu.pe

Abstract– This paper presents the development of a predictive system for the early prevention of muscle injuries in athletes, using Machine Learning and Deep Learning techniques applied to anthropometric and morphological variables. The methodological process was structured based on the CRISP-DM approach, including data cleaning, Min-Max normalization, ordinal encoding, class balancing, and stratified partitioning into training (80%) and test (20%) sets. Six predictive models were evaluated: Random Forest (RF), XGBoost, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and a Deep Learning model based on artificial neural networks implemented using Keras. Additionally, an Ensemble Voting scheme was developed to integrate individual predictions through majority voting, resulting in a more robust and stable output. Model performance was evaluated using metrics such as Accuracy, Precision, Recall, and F1-score, as well as Precision–Recall curves, allowing for an adequate assessment under class imbalance conditions. The system was subsequently implemented as an interactive web application that enables sports personnel to input athlete data and obtain real-time predictions, facilitating preventive decision-making and contributing to the reduction of muscle injury risk.

Keywords– Anthropometric analysis, Deep learning, Injury prediction, Machine learning, Muscle injuries

Modelo predictivo para prevenir lesiones en atletas utilizando machine learning y análisis de sensibilidad

Christian Ovalle¹, Everardo Sánchez Puche², Waldyr Cerron Valverde³, Omar Santillán Achings⁴

¹Department of Engineering, Universidad Tecnológica del Perú, Lima, Perú, C19349@utp.edu.pe

²Department of Engineering, Universidad Popular del Cesar, Lima, Perú, everardomsanchez@unicesar.edu.co

³Department of Engineering, Universidad Tecnológica del Perú, Lima, Perú, C21262@utp.edu.pe

⁴Department of Engineering, Universidad Tecnológica del Perú, Lima, Perú, C25611@utp.edu.pe

Resumen– El presente artículo presenta el desarrollo de un sistema predictivo para la prevención temprana de lesiones musculares en atletas, utilizando técnicas de aprendizaje automático y aprendizaje profundo aplicadas a variables antropométricas y morfológicas. El proceso metodológico se estructuró con base en el enfoque CRISP-DM, incluyendo limpieza de datos, normalización Min-Max, codificación ordinal, balanceo de clases y partición estratificada en conjuntos de entrenamiento (80%) y prueba (20%). Se evaluaron seis modelos predictivos: Random Forest (RF), XGBoost, Perceptrón Multicapa (MLP), Máquina de Vectores de Soporte (SVM), K-Vecinos Más Cercanos (KNN) y un modelo de aprendizaje profundo basado en redes neuronales artificiales implementado con Keras. Además, se desarrolló un esquema de votación de conjunto para integrar predicciones individuales mediante votación mayoritaria, lo que resultó en una salida más robusta y estable. El rendimiento del modelo se evaluó utilizando métricas como precisión, exhaustividad y puntuación F1, así como curvas de precisión-exhaustividad, lo que permitió una evaluación adecuada en condiciones de desequilibrio de clases. Posteriormente, el sistema se implementó como una aplicación web interactiva que permite al personal deportivo introducir datos de los atletas y obtener predicciones en tiempo real, facilitando así la toma de decisiones preventivas y contribuyendo a la reducción del riesgo de lesiones musculares.

Palabras clave–Análisis antropométrico, Aprendizaje profundo, Predicción de lesiones, Aprendizaje automático, Lesiones musculares.

I. INTRODUCCIÓN

Las lesiones deportivas constituyen una problemática recurrente en la mayoría de las disciplinas de alto rendimiento; sin embargo, en el fútbol profesional su impacto resulta particularmente crítico debido a la elevada exigencia física, la congestión del calendario competitivo y la intensidad de los entrenamientos. Diversos estudios reportan que entre el 45 % y el 65 % de los futbolistas profesionales sufren al menos una lesión durante cada temporada deportiva [1], [2]. Esta situación afecta tanto el rendimiento individual del atleta como el desempeño colectivo de los equipos y la planificación táctica. Asimismo, el impacto económico asociado a las lesiones es considerable; de acuerdo con el European Football Injury Index 2023–2024 elaborado por Howden, las pérdidas superan los 500 millones de euros por temporada, considerando costos médicos, ausencias prolongadas y depreciación del valor deportivo de los jugadores [3].

Dentro de este contexto, las lesiones musculares, en particular las que comprometen los isquiotibiales, representan aproximadamente el 30 % del total de lesiones reportadas,

evidenciando una tendencia persistente que continúa afectando el rendimiento competitivo a lo largo del tiempo [4], [5]. En este escenario, entrenadores, preparadores físicos y profesionales de la salud deportiva han intensificado la implementación de estrategias orientadas a la prevención de lesiones y a la detección temprana de factores de riesgo. El enfoque preventivo permite reducir la severidad de las lesiones, acortar los tiempos de recuperación y minimizar su recurrencia [6]. No obstante, las estrategias tradicionales, basadas en la observación clínica, la experiencia del cuerpo técnico y el seguimiento de variables fisiológicas básicas, presentan limitaciones frente a la alta variabilidad interindividual y la complejidad de las cargas físicas.

En los últimos años, el Machine Learning ha emergido como una herramienta relevante para abordar esta problemática, debido a su capacidad para analizar grandes volúmenes de datos y modelar relaciones no lineales entre múltiples variables asociadas al riesgo de lesión, tales como la carga de entrenamiento, la fatiga acumulada, la recuperación, las características morfológicas y los antecedentes lesivos [7], [8]. Estudios recientes sugieren que la incorporación de modelos predictivos basados en aprendizaje automático dentro de la planificación deportiva puede contribuir a reducir la incidencia de lesiones hasta en un 20 %, optimizando la toma de decisiones en la gestión de cargas y la individualización del entrenamiento [9]. A pesar de estos avances, la literatura evidencia importantes limitaciones que restringen la aplicabilidad de los modelos propuestos. Entre las principales debilidades se encuentran la escasez de conjuntos de datos representativos, la baja diversidad de las muestras analizadas y el entrenamiento de modelos en contextos altamente específicos, como clubes de élite con infraestructura tecnológica avanzada [10], [11].

En este sentido, la validación del desempeño de sistemas de monitoreo basados en sensores y algoritmos de Machine Learning en condiciones reales de entrenamiento y competición resulta imprescindible para su adopción operativa en el ámbito deportivo [12]. Las plataformas basadas en dispositivos wearables permiten capturar métricas de carga externa e interna, tales como datos GPS, acelerometría, esclerometría y frecuencia cardíaca, las cuales se han correlacionado con eventos de fatiga y riesgo de lesión [13]. Estudios experimentales reportan incrementos significativos en la sensibilidad de detección temprana cuando se emplean estos dispositivos junto con análisis automatizado [12], y trabajos de clasificación han alcanzado precisiones superiores

al 80 % en escenarios controlados [14]. Asimismo, la integración de datos objetivos provenientes de GPS con percepciones subjetivas de esfuerzo ha sido asociada con reducciones de la incidencia lesiva cercanas al 20 % en cohortes de futbolistas, evidenciando un impacto operativo potencial [15]. Adicionalmente, enfoques más recientes basados en Deep Learning, modelos gráficos y arquitecturas cloud-based han demostrado una elevada capacidad para modelar información espacio-temporal y operar en sistemas distribuidos, lo que abre nuevas posibilidades para el desarrollo de aplicaciones en tiempo real [16], [17]. El uso de sensores GPS de alta resolución ha permitido mejorar la estimación de cargas físicas y detectar microvariaciones asociadas a esfuerzos intensivos, reforzando el rol de la telemetría en el fútbol y otros deportes intermitentes [18]. De forma complementaria, técnicas basadas en imágenes térmicas automatizadas han mostrado potencial para identificar variaciones fisiológicas relacionadas con el riesgo de lesión, especialmente cuando se integran en flujos híbridos de análisis de datos [19].

No obstante, un análisis comparativo de estos enfoques revela brechas persistentes. La mayoría de los estudios se desarrolla en contextos controlados o en clubes con alto nivel de tecnificación, lo que limita la replicabilidad y escalabilidad de las soluciones propuestas en entornos locales o de recursos restringidos [20]. Asimismo, aunque los métodos ensemble como Random Forest (RF) y XGBoost, así como los modelos neuronales como Multi-Layer Perceptron (MLP) y redes neuronales convolucionales (CNN), han reportado altos valores de exactitud, AUC y sensibilidad en validaciones internas [21], [22], [23], muchos de estos modelos dependen de grandes volúmenes de datos longitudinales para evitar el sobreajuste y garantizar una adecuada capacidad de generalización [20], [24]. A ello se suma la falta de estandarización en los protocolos de captura de datos y en las métricas de evaluación, lo que dificulta la comparación directa entre estudios [15]. Otro aspecto crítico corresponde a la limitada interpretabilidad de los modelos complejos, lo que reduce su aceptación y uso por parte de entrenadores, fisioterapeutas y personal médico, quienes requieren comprender los factores que influyen en cada predicción para respaldar sus decisiones [21], [25-28]. Aunque técnicas explicativas como SHAP y LIME han contribuido a mejorar la comprensión de los modelos predictivos [29], [30], su integración en plataformas accesibles y de uso cotidiano sigue siendo limitada.

Actualmente, el desarrollo de modelos predictivos aplicables en escenarios reales continúa enfrentando barreras técnicas relevantes, particularmente en términos de interpretabilidad, requerimientos computacionales y complejidad de implementación en clubes con presupuestos limitados [16]. Si bien investigaciones recientes han incorporado validación cruzada avanzada, selección de

variables y arquitecturas profundas para mejorar la precisión en la predicción de lesiones [26], los sistemas resultantes suelen depender de infraestructura especializada y flujos de datos complejos provenientes de sensores avanzados [27], [28]. Esta situación evidencia la ausencia de soluciones ligeras que puedan ser utilizadas directamente por el personal deportivo sin necesidad de equipamiento adicional, y que operen exclusivamente con información fácilmente disponible, como los datos morfológicos del atleta.

En este contexto, la innovación central de la presente investigación consiste en el desarrollo de un sistema web predictivo, accesible y explicable, que estima el riesgo de lesión a partir de datos morfológicos ingresados manualmente por el usuario. El estudio compara de manera sistemática seis modelos: Random Forest (RF), XGBoost, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) y un modelo de Deep Learning basado en redes neuronales implementado mediante Keras, seleccionados por su diversidad metodológica y desempeño previo en aplicaciones deportivas [31], [32]. Mientras los modelos neuronales permiten capturar patrones complejos presentes en las características físicas del atleta, los métodos ensemble ofrecen robustez frente al sobreajuste, y algoritmos como SVM y KNN destacan por su eficiencia e interpretabilidad en escenarios con recursos limitados [33], [34]. La novedad del enfoque radica en identificar no solo el modelo con mayor precisión, sino aquel que presenta el mejor equilibrio entre rendimiento predictivo, interpretabilidad y viabilidad operativa dentro de una plataforma web. De este modo, la propuesta contribuye a cerrar brechas persistentes en la literatura y a facilitar la adopción de sistemas predictivos en contextos deportivos locales. El objetivo de este estudio es desarrollar y evaluar modelos predictivos basados en inteligencia artificial para apoyar la prevención de lesiones deportivas en atletas.

II. METODOLOGÍA

Se empleó la metodología CRISP-DM por su enfoque estructurado y adaptable a proyectos de minería de datos en el ámbito deportivo, respaldado por estudios que demuestran su vigencia y eficacia en diversos contextos analíticos [35], [36]. Este modelo permite gestionar de forma ordenada las etapas del proceso, desde la comprensión del problema hasta el despliegue del modelo predictivo, garantizando la trazabilidad y reproducibilidad del análisis [37]. Asimismo, proporciona un marco flexible que ha sido aplicado con éxito en modelos predictivos sobre datos estructurados y no estructurados, facilitando la comparación de múltiples algoritmos y el desarrollo de soluciones funcionales [38]–[40]. Por último, se presenta el diagrama de solución adaptado a la metodología CRISP-DM en la Figura 1.

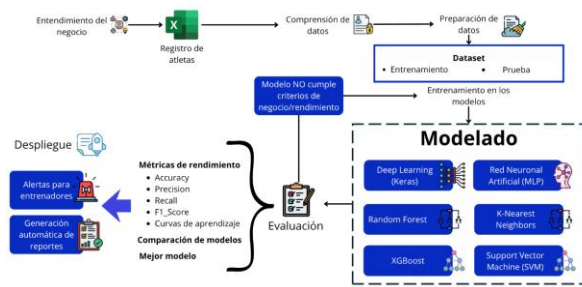


Figura 1. Diagrama de solución CRISP DM

A. Preparación de datos

Se aplicaron técnicas de preprocesamiento orientadas a garantizar la calidad y consistencia de los datos para el modelado. En primer lugar, se realizó la selección de variables relevantes mediante la identificación de columnas asociadas a evaluaciones, resultados e indicadores morfológicos, eliminando aquellas de carácter identificativo. Posteriormente, la variable objetivo fue estandarizada mediante la unificación de etiquetas semánticamente equivalentes en categorías consistentes. Las variables categóricas fueron codificadas utilizando One-Hot Encoding para aquellas con baja cardinalidad y Frequency Encoding para variables con alta cardinalidad. Por su parte, las variables numéricas fueron escaladas mediante el método StandardScaler, con el fin de normalizar su distribución y mejorar el desempeño de los algoritmos. Adicionalmente, se aplicó un proceso de selección de características utilizando el método SelectKBest basado en la prueba ANOVA F-test, conservando las variables con mayor relevancia estadística. Finalmente, el conjunto de datos fue dividido en subconjuntos de entrenamiento (80 %) y prueba (20 %), empleando un muestreo estratificado para preservar la proporción de clases. Asimismo, se aplicó validación cruzada de 10 pliegues (10-fold cross-validation) durante la fase de entrenamiento para garantizar la robustez y capacidad de generalización de los modelos. La Figura 2 presenta el mapa de calor de correlación entre variables numéricas, mientras que la Figura 3 muestra la matriz de dispersión para el análisis de relaciones bivariadas.



Figura 2. Mapa de calor de correlación

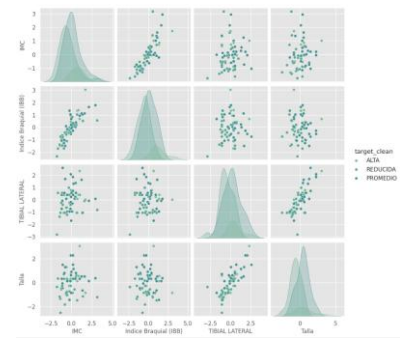


Figura 3. Matriz de dispersión (Scatterplot Matrix)

B. Entrenamiento de modelos

Se evaluaron seis modelos de aprendizaje automático: Random Forest (RF), XGBoost, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) y un modelo de Deep Learning basado en redes neuronales implementado mediante Keras. La selección de estos modelos se fundamenta en su desempeño reportado en estudios previos relacionados con la predicción de lesiones deportivas [41]–[45]. El entrenamiento de los modelos se realizó sobre el conjunto de datos de entrenamiento, utilizando validación cruzada de 10 pliegues para optimizar los hiperparámetros y reducir el riesgo de sobreajuste. El desempeño de los modelos fue evaluado mediante métricas estándar: accuracy (exactitud), precision (precisión), recall (sensibilidad) y F1-score. La Figura 4 presenta las curvas de aprendizaje (Learning Curves), las cuales permiten analizar el comportamiento de los modelos durante el entrenamiento e identificar posibles casos de sobreajuste o subajuste.

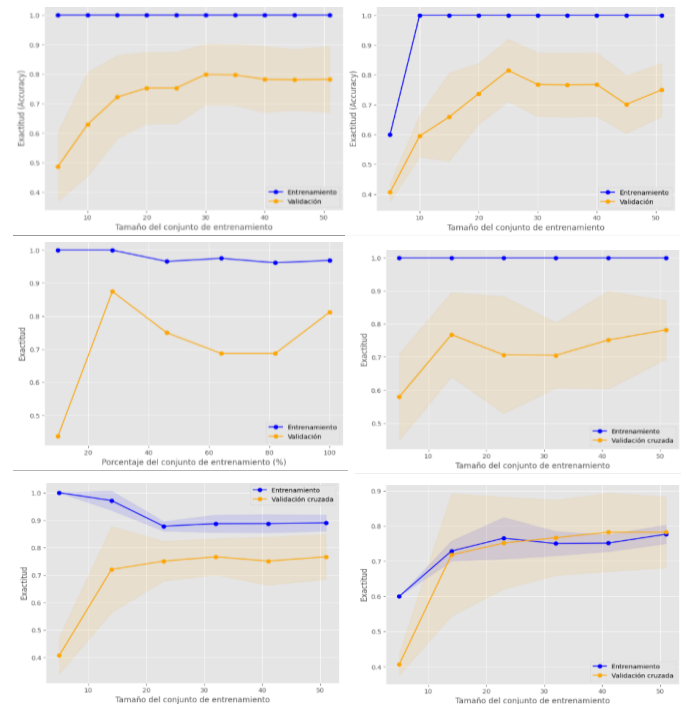


Figura 4. Curvas de aprendizaje

Asimismo, la Figura 5 muestra las matrices de confusión, que facilitan la interpretación del desempeño de cada modelo en términos de clasificación correcta e incorrecta

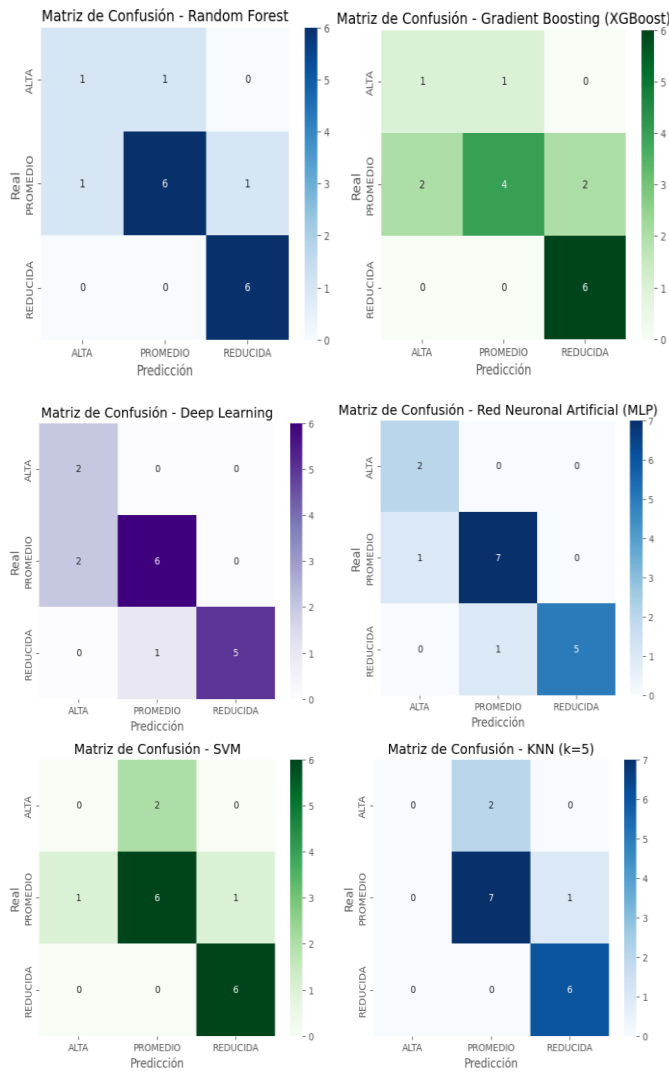


Figura 5. Matrices de confusión

C. Evaluación

Se realizó una comparación del desempeño de los modelos con base en las métricas accuracy, precision, recall y F1-score, permitiendo una evaluación integral de su capacidad predictiva. Este análisis facilita la identificación del modelo más adecuado para la predicción de lesiones deportivas en función de su equilibrio entre precisión y capacidad de generalización. Los resultados se presentan en la Tabla 1. El modelo Multi-Layer Perceptron (MLP) obtuvo el mejor desempeño global, con una exactitud de 0.875, seguido por Random Forest (RF) y el modelo de Deep Learning implementado con Keras, que también mostraron resultados competitivos.

TABLA I: COMPARACIÓN DE MÉTRICAS

Modelos	Accuracy	Precision	Recall	F1score
Random	0.8125	0.8125	0.8125	0.8087
Forest				
XGBoost	0.6880	0.7230	0.6880	0.6790
MLP	0.8750	0.8960	0.8750	0.8780
Keras	0.8120	0.8240	0.8120	0.8120
Support	0.7500	0.6960	0.7500	0.7210
Vector				
Machine				
K-Nearest	0.8120	0.7100	0.8120	0.7580
Neighbors				

D. Despliegue

Tras la evaluación de los modelos, se implementó una aplicación web interactiva que permite al personal deportivo realizar análisis predictivos en tiempo real. El sistema integra los modelos desarrollados y genera predicciones de riesgo de lesión a partir de los datos ingresados mediante un formulario, como se muestra en la Figura 6. Asimismo, la plataforma permite visualizar las probabilidades estimadas por cada modelo, facilitando la comparación de resultados y la toma de decisiones basada en evidencia. La Figura 7 presenta la interfaz de resultados del sistema predictivo.

Figure 6 shows a web form titled 'Predicción de Condición Muscular del Atleta'. It includes input fields for 'IMC', 'Índice Braquial (IBB)', 'Tibial Lateral', and 'Talla (cm)'. Below these, there are sections for 'EVALUACIÓN' and 'RESULTADO' for various metrics, including 'IMC', 'Metrocómicos Tronco Medio', 'BIA/BILIO', and 'ICP'. A button 'Analizar Resultados' is at the bottom.

Figura 6. Formulario para analizar el estado físico del atleta

Figure 7 shows the 'Resultados del Análisis' interface. It includes a table with columns 'Modelo', 'Predicción', and 'Probabilidades'. The table lists results for Random Forest, XGBoost, MLP, SVM, KNN (k=5), and Deep Learning (Keras). Below the table, it shows the 'Ensemble (voto mayoritario): PROMEDIO' and 'Confianza promedio (REDUCIDA): 0.1'. A button 'Volver al formulario' is at the bottom.

Modelo	Predicción	Probabilidades
Random Forest	PROMEDIO	[ALTA: 0.34, PROMEDIO: 0.59, REDUCIDA: 0.07]
XGBoost	ALTA	[ALTA: 0.77, PROMEDIO: 0.21, REDUCIDA: 0.02]
MLP	PROMEDIO	[ALTA: 0.08, PROMEDIO: 0.79, REDUCIDA: 0.13]
SVM	PROMEDIO	[ALTA: 0.24, PROMEDIO: 0.57, REDUCIDA: 0.18]
KNN (k=5)	PROMEDIO	[ALTA: 0.20, PROMEDIO: 0.80, REDUCIDA: 0.00]
Deep Learning (Keras)	ALTA	[ALTA: 0.58, PROMEDIO: 0.21, REDUCIDA: 0.21]

Ensemble (voto mayoritario): PROMEDIO
Confianza promedio (REDUCIDA): 0.1

Figura 7. Resultados del análisis predictivo de los modelos

III. RESULTADOS Y DISCUSIÓN

Esta sección presenta los resultados obtenidos tras aplicar el proceso completo de preprocesamiento al conjunto de datos. Dicho proceso incluyó la selección de variables relevantes mediante palabras clave, la eliminación de variables identificativas, la imputación de valores nulos, la normalización de la variable objetivo, la codificación de variables categóricas mediante One-Hot Encoding y Frequency Encoding, la estandarización de variables numéricas con StandardScaler y la selección de características mediante el método SelectKBest basado en la prueba ANOVA F-test.

Con el conjunto de datos procesado, se entrenaron seis modelos predictivos: Random Forest (RF), XGBoost, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) y un modelo de Deep Learning basado en redes neuronales implementado mediante Keras. El entrenamiento y evaluación se realizaron utilizando la partición previamente definida de entrenamiento (80 %) y prueba (20 %), incorporando muestreo estratificado y validación cruzada. El análisis de resultados se apoya en herramientas exploratorias como mapas de calor de correlación y matrices de dispersión, así como en curvas de aprendizaje, matrices de confusión y métricas cuantitativas calculadas para cada modelo.

A. Resultados del desempeño de los modelos

El desempeño de los modelos predictivos fue evaluado mediante métricas estándar de clasificación: accuracy (exactitud), precision (precisión), recall (sensibilidad) y F1-score. Entre estas, la accuracy se utilizó como métrica principal para el ranking, debido a su uso frecuente en estudios relacionados con la predicción de lesiones y clasificación fisiológica, lo que permite la comparación con investigaciones previas. Para identificar el modelo con mejor rendimiento, los seis modelos fueron ordenados en función de sus valores de accuracy obtenidos durante la evaluación. La Tabla 2 presenta el ranking correspondiente. La fórmula empleada se presenta a continuación:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

TABLA 2. CLASIFICACIÓN DE MODELOS SEGÚN SU PRECISIÓN

Ranking	Modelo	Accuracy
1	MLP	0.8750
2	Random Forest	0.8125
3	K-Nearest Neighbors	0.8120
4	Keras	0.8120
5	Support Vector Machine	0.7500
6	XGBoost	0.6880

B. Discusión

Los resultados obtenidos indican que el modelo Multi-Layer Perceptron (MLP) alcanzó el mejor desempeño, con una accuracy de 0.875, lo que evidencia su capacidad para modelar

relaciones no lineales entre variables fisiológicas y morfológicas. Las matrices de confusión muestran que la clase “Musculatura Promedio” fue la mejor clasificada por todos los modelos, mientras que clases como “Musculatura Alta” y “Musculatura Reducida” presentaron mayores niveles de confusión. Este comportamiento puede estar asociado al desbalance de clases y a la superposición en la distribución de características. La consistencia observada en las métricas de evaluación respalda la confiabilidad del sistema predictivo propuesto. Además, la integración del modelo en una aplicación web permite al personal deportivo obtener predicciones en tiempo real, facilitando la toma de decisiones orientadas a la prevención de lesiones. En conjunto, los resultados evidencian la efectividad de los enfoques basados en Machine Learning, especialmente de los modelos neuronales, para la identificación temprana del riesgo de lesiones a partir de variables morfológicas.

IV. CONCLUSIÓN

El presente estudio permitió desarrollar y evaluar un sistema predictivo para estimar el riesgo de lesión en atletas utilizando variables morfológicas y evaluaciones funcionales, siguiendo la metodología CRISP-DM desde la comprensión del problema hasta el despliegue del modelo. El proceso incluyó etapas de preprocesamiento tales como imputación de valores, codificación de variables categóricas, estandarización de variables numéricas y selección de características mediante el método ANOVA F-test. Se evaluaron seis modelos de aprendizaje automático: Random Forest (RF), XGBoost, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) y un modelo de Deep Learning basado en redes neuronales implementado mediante Keras. Los resultados evidenciaron que el modelo Multi-Layer Perceptron (MLP) alcanzó el mejor desempeño, con una accuracy de 0.8750, destacando su capacidad para modelar relaciones no lineales entre variables musculares y antropométricas.

Asimismo, se observó que las clases con mayor representación en el conjunto de datos fueron predichas con mayor estabilidad, mientras que las categorías minoritarias presentaron mayores niveles de confusión. Este comportamiento sugiere la necesidad de incorporar, en trabajos futuros, técnicas de balanceo de clases o ampliación del conjunto de datos para mejorar la capacidad de generalización de los modelos. La implementación del sistema en una aplicación web funcional permitió integrar los modelos desarrollados en un entorno accesible, capaz de generar predicciones en tiempo real. Esta característica facilita el monitoreo del estado físico del atleta y apoya la toma de decisiones orientadas a la prevención de lesiones.

En conjunto, los resultados confirman la viabilidad del uso de técnicas de Machine Learning como herramienta de apoyo en programas de prevención de lesiones deportivas. Asimismo, el estudio establece una base para futuras

investigaciones orientadas a la optimización de modelos, la incorporación de nuevas variables y la mejora de estrategias de intervención en contextos deportivos.

AGRADECIMIENTO/RECONOCIMIENTO

Agradecemos a la Universidad Tecnológica del Perú por su apoyo en la elaboración del presente artículo, lo cual fue de gran ayuda las herramientas brindadas para un resultado satisfactorio.

REFERENCIAS

- [1] L. Ren, Y. Wang, and K. Li, "Real-time sports injury monitoring system based on the deep learning algorithm," *BMC Med Imaging*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12880-024-01304-6.
- [2] Q. Jianjun, H. F. Islem, W. J. K. Almoghayer, and M. Khishe, "Predictive athlete performance modeling with machine learning and biometric data integration," *Scientific Reports* 2025 15:1, vol. 15, no. 1, pp. 16365-, May 2025, doi: 10.1038/s41598-025-01438-9.
- [3] "Howden's 2023/24 Men's European football injury index." Accessed: Nov. 18, 2025. [Online]. Available: <https://www.howdengroupholdings.com/reports/2023-24-mens-european-football-injury-index>
- [4] M. Kolodziej *et al.*, "Predictive modeling of lower extremity injury risk in male elite youth soccer players using least absolute shrinkage and selection operator regression," *Scand J Med Sci Sports*, vol. 33, no. 6, pp. 1021–1033, Jun. 2023, doi: 10.1111/SMS.14322.
- [5] F. J. Robles-Palazón *et al.*, "Predicting injury risk using machine learning in male youth soccer players," *Chaos Solitons Fractals*, vol. 167, p. 113079, Feb. 2023, doi: 10.1016/J.CHAOS.2022.113079.
- [6] X. Gai, "Application of flexible sensor multimodal data fusion system based on artificial synapse and machine learning in athletic injury prevention and health monitoring," *Discover Artificial Intelligence*, vol. 5, p. 31, 123AD, doi: 10.1007/s44163-025-00254-4.
- [7] A. Abasi, A. Nazari, A. Moezy, and S. A. Fatemi Aghda, "Machine learning models for reinjury risk prediction using cardiopulmonary exercise testing (CPET) data: optimizing athlete recovery," *BioData Mining* 2025 18:1, vol. 18, no. 1, pp. 16-, Feb. 2025, doi: 10.1186/S13040-025-00431-2.
- [8] G. Chellamuthu, S. Sahanand, S. Sundar, S. Bradley, A. Gudimetla, and D. V. Rajan, "Injury prediction model for lower-limb sports injuries: A novel machine learning-based approach," *J Clin Orthop Trauma*, vol. 69, Oct. 2025, doi: 10.1016/j.jcot.2025.103114.
- [9] J. B. Matas-Bustos, A. M. Mora-García, M. de Hoyo Lora, A. Nieto-Alarcón, and F. T. Gonzalez-Fernández, "Advanced feature engineering in Acute:Chronic Workload Ratio (ACWR) calculation for injury forecasting in elite soccer," *PLoS One*, vol. 20, no. 7, p. e0327960, Jul. 2025, doi: 10.1371/JOURNAL.PONE.0327960.
- [10] M. F. Burbano-Fernández, A. F. Villaquirán-Hurtado, V. M. Celis-Quinayas, J. A. Hoyos-Quisobony, and N. J. Molano-Tobar, "Ankle injury prevention in soccer using mMachine leArning: bibliometric AnAlysis," *Cultura, Ciencia y Deporte*, vol. 19, no. 61, pp. 27–43, Sep. 2024, doi: 10.12800/CCD.V19I61.2119.
- [11] T. Tsilimigkras, I. Kakkos, G. K. Matsopoulos, and G. C. Bogdanis, "Enhancing Sports Injury Risk Assessment in Soccer Through Machine Learning and Training Load Analysis," *J Sports Sci Med*, vol. 23, no. 3, pp. 537–547, Sep. 2024, doi: 10.52082/JSSM.2024.537.
- [12] M. Kovoor, M. Durairaj, M. S. Karyakarte, M. Zair Hussain, M. Ashraf, and L. P. Maguluri, "Sensor-enhanced wearables and automated analytics for injury prevention in sports," *Measurement: Sensors*, vol. 32, p. 101054, Apr. 2024, doi: 10.1016/J.MEASEN.2024.101054.
- [13] D. R. Seshadri *et al.*, "Wearable sensors for monitoring the internal and external workload of the athlete," *npj Digital Medicine* 2019 2:1, vol. 2, no. 1, pp. 71-, Jul. 2019, doi: 10.1038/s41746-019-0149-2.
- [14] R. E. D. Ayala, D. P. Granados, C. A. G. Gutiérrez, M. A. O. Ruiz, N. R. Espinosa, and E. C. Heredia, "Novel Study for the Early Identification of Injury Risks in Athletes Using Machine Learning Techniques," *Applied Sciences* 2024, Vol. 14, Page 570, vol. 14, no. 2, p. 570, Jan. 2024, doi: 10.3390/AP14020570.
- [15] R. Leppich, P. Kunz, A. Bauer, S. Kounev, B. Sperlich, and P. Düring, "Prediction of Perceived Exertion Ratings in National Level Soccer Players Using Wearable Sensor Data and Machine Learning Techniques," *J Sports Sci Med*, vol. 23, pp. 744–753, 2024, doi: 10.52082/jssm.2024.744.
- [16] J. Zhu, Z. Ye, M. Ren, and G. Ma, "Transformative skeletal motion analysis: optimization of exercise training and injury prevention through graph neural networks," *Front Neurosci*, vol. 18, p. 1353257, Mar. 2024, doi: 10.3389/FNINS.2024.1353257/BIBTEX.
- [17] X. Wu *et al.*, "Cloud-based deep learning-assisted system for diagnosis of sports injuries," *Journal of Cloud Computing* 2022 11:1, vol. 11, no. 1, pp. 82-, Nov. 2022, doi: 10.1186/S13677-022-00355-W.
- [18] M. Gamage Preethichandra *et al.*, "Predicting Injuries in Football Based on Data Collected from GPS-Based Wearable Sensors," *Sensors* 2023, Vol. 23, Page 1227, vol. 23, no. 3, p. 1227, Jan. 2023, doi: 10.3390/S23031227.
- [19] L. Li, Y. Wei, and S. Xiang, "Infrared thermal image monitoring based on artificial intelligence application in the prevention of sports injuries in aerobics: Computational thermal modeling," *Thermal Science and Engineering Progress*, vol. 57, p. 103126, Jan. 2025, doi: 10.1016/J.TSEP.2024.103126.
- [20] L. Zumeta-Olaskoaga, A. Bender, and D. J. Lee, "Flexible modelling of time-varying exposures and recurrent events to analyse training load effects in team sports injuries," *J R Stat Soc Ser C Appl Stat*, vol. 74, no. 2, pp. 391–405, Mar. 2025, doi: 10.1093/JRSSC/QLAE059.
- [21] M. Calderón-Díaz *et al.*, "Explainable Machine Learning Techniques to Predict Muscle Injuries in Professional Soccer Players through Biomechanical Analysis," *Sensors* 2024, Vol. 24, Page 119, vol. 24, no. 1, p. 119, Dec. 2023, doi: 10.3390/S24010119.
- [22] A. Abdusalomov *et al.*, "Lightweight Deep Learning Framework for Accurate Detection of Sports-Related Bone Fractures," *Diagnostics* 2025, Vol. 15, Page 271, vol. 15, no. 3, p. 271, Jan. 2025, doi: 10.3390/DIAGNOSTICS15030271.
- [23] E. Anbalagan and S. M. Anbhaazhagan, "Deep learning model using ensemble based approach for walking activity recognition and gait event prediction with grey level co-occurrence matrix," *Expert Syst Appl*, vol. 227, p. 120337, Oct. 2023, doi: 10.1016/J.ESWA.2023.120337.
- [24] G. Kakavas, N. Malliaropoulos, R. Pruna, and N. Maffulli, "Artificial intelligence: A tool for sports trauma prediction," *Injury*, vol. 51, pp. S63–S65, Aug. 2020, doi: 10.1016/j.injury.2019.08.033.
- [25] J. Briand, S. Deguire, S. Gaudet, and F. Bieuzien, "Monitoring Variables Influence on Random Forest Models to Forecast Injuries in Short-Track Speed Skating," *Front Sports Act Living*, vol. 4, p. 896828, Jul. 2022, doi: 10.3389/FSPOR.2022.896828/BIBTEX.
- [26] D. Xiao, F. Zhu, J. Jiang, and X. Niu, "Leveraging natural cognitive systems in conjunction with ResNet50-BiGRU model and attention mechanism for enhanced medical image analysis and sports injury prediction," *Front Neurosci*, vol. 17, p. 1273931, Sep. 2023, doi: 10.3389/FNINS.2023.1273931/BIBTEX.
- [27] X. Ye, Y. Huang, Z. Bai, and Y. Wang, "A novel approach for sports injury risk prediction: based on time-series image encoding and deep learning," *Front Physiol*, vol. 14, p. 1174525, Dec. 2023, doi: 10.3389/FPHYS.2023.1174525/BIBTEX.
- [28] Z. G. Zhu, "Design and implementation of an intelligent sports management system (ISMS) using wireless sensor networks," *PeerJ Comput Sci*, vol. 11, p. e2637, Jan. 2025, doi: 10.7717/PEERJ-CS.2637/SUPP-1.
- [29] K. Sandamal, S. Arachchi, V. O. Erkudiv, K. U. Rozumbetov, and U. Rathnayake, "Explainable artificial intelligence for fitness prediction of young athletes living in unfavorable environmental conditions," *Results in Engineering*, vol. 23, p. 102592, Sep. 2024, doi: 10.1016/J.RINENG.2024.102592.
- [30] A. Li, Y. Wang, and H. Chen, "AI driven cardiovascular risk prediction using NLP and Large Language Models for personalized medicine in athletes," *SLAS Technol*, vol. 32, p. 100286, Jun. 2025, doi: 10.1016/j.slast.2025.100286.
- [31] Y. Hu, Y. Li, B. Cui, H. Su, and P. Zhu, "Internet of things enabled deep learning monitoring system for realtime performance metrics and athlete feedback in college sports," *Scientific Reports* 2025 15:1, vol. 15, no. 1, pp. 28405-, Aug. 2025, doi: 10.1038/s41598-025-13949-6.
- [32] P. Diniz *et al.*, "Pre-injury performance is most important for predicting the level of match participation after Achilles tendon ruptures in elite soccer players: a study using a machine learning classifier," *Knee Surgery, Sports*

Traumatology, Arthroscopy 2022 30:12, vol. 30, no. 12, pp. 4225–4237, Aug. 2022, doi: 10.1007/S00167-022-07082-4.

[33] F. Martins *et al.*, “Predictive Modeling of Injury Risk Based on Body Composition and Selected Physical Fitness Tests for Elite Football Players,” *Journal of Clinical Medicine* 2022, Vol. 11, Page 4923, vol. 11, no. 16, p. 4923, Aug. 2022, doi: 10.3390/JCM11164923.

[34] D. N. Freitas *et al.*, “Predicting noncontact injuries of professional football players using machine learning,” *PLoS One*, vol. 20, no. 1, p. e0315481, Jan. 2025, doi: 10.1371/JOURNAL.PONE.0315481.

[35] C. Schröder, F. Kruse, and J. M. Gómez, “A Systematic Literature Review on Applying CRISP-DM Process Model,” *Procedia Comput Sci*, vol. 181, pp. 526–534, Jan. 2021, doi: 10.1016/J.PROCS.2021.01.199.

[36] R. A. Casonatto, T. D. P. G. Souza, and A. M. Mariano, “Quality and Risk Management in Data Mining: A CRISP-DM Perspective,” *Procedia Comput Sci*, vol. 242, pp. 161–168, Jan. 2024, doi: 10.1016/J.PROCS.2024.08.257.

[37] M. Elkabalawy, A. Al-Sakkaf, E. Mohammed Abdelkader, and G. Alfalah, “CRISP-DM-Based Data-Driven Approach for Building Energy Prediction Utilizing Indoor and Environmental Factors,” *Sustainability* 2024, Vol. 16, Page 7249, vol. 16, no. 17, p. 7249, Aug. 2024, doi: 10.3390/SU16177249.

[38] W. Y. Ayele, “Adapting CRISP-DM for Idea Mining: A Data Mining Process for Generating Ideas Using a Textual Dataset,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 6, May 2021, doi: 10.14569/issn.2156-5570.

[39] M. C. Vale Tavares, P. Alencar, and D. Cowan, “A Variability-Aware Design Approach to the Data Analysis Modeling Process,” *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, pp. 2818–2827, Dec. 2018, doi: 10.1109/BigData.2018.8622036.

[40] V. Camomilla *et al.*, “Machine Learning to Predict Lower Extremity Musculoskeletal Injury Risk in Student Athletes,” *Front Sports Act Living*, vol. 2, p. 576655, Nov. 2020, doi: 10.3389/FSPOR.2020.576655.

[41] M. Calderón-Díaz *et al.*, “Explainable Machine Learning Techniques to Predict Muscle Injuries in Professional Soccer Players through Biomechanical Analysis,” *Sensors* 2024, Vol. 24, Page 119, vol. 24, no. 1, p. 119, Dec. 2023, doi: 10.3390/S24010119.

[42] H. Van Eetvelde, L. D. Mendonça, C. Ley, R. Seil, and T. Tischer, “Machine learning methods in sport injury prediction and prevention: a systematic review,” *J Exp Orthop*, vol. 8, no. 1, p. 27, Dec. 2021, doi: 10.1186/S40634-021-00346-X;SUBPAGE:STRING:FULL.

[43] A. Majumdar, R. Bakirov, D. Hodges, S. Scott, and T. Rees, “Machine Learning for Understanding and Predicting Injuries in Football,” *Sports Medicine - Open* 2022 8:1, vol. 8, no. 1, pp. 73–, Jun. 2022, doi: 10.1186/S40798-022-00465-4.

[44] G. Zhang, Z. Zhao, and F. Xie, “Risk prediction modeling in the context of machine learning for injury prevention in sports,” *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Jan. 2024, doi: 10.2478/AMNS-2024-3067.

[45] K. He, “Prediction Model of Juvenile Football Players’ Sports Injury Based on Text Classification Technology of Machine Learning,” *Mobile Information Systems*, vol. 2021, 2021, doi: 10.1155/2021/2955215.