

## Article

# Novel Study for the Early Identification of Injury Risks in Athletes Using Machine Learning Techniques

Rocío Elizabeth Duarte Ayala <sup>1,\*</sup>, David Pérez Granados <sup>2</sup>, Carlos Alberto González Gutiérrez <sup>3,\*</sup>,  
Mauricio Alberto Ortega Ruíz <sup>2,4</sup>, Natalia Rojas Espinosa <sup>5</sup> and Emanuel Canto Heredia <sup>6</sup>

- <sup>1</sup> School of Health Sciences, Campus Lomas Verdes, Universidad del Valle de México, Lomas Verdes 53220, Mexico
- <sup>2</sup> Department of Engineering, CIIDETEC—Coyoacán, Universidad del Valle de México, Coyoacán 04910, Mexico; a340348289@my.uvm.edu.mx (D.P.G.); mauricio.ortegarua@uvmnet.edu (M.A.O.R.)
- <sup>3</sup> Department of Engineering, CIIDETEC—Querétaro, Universidad del Valle de México, Querétaro 76230, Mexico
- <sup>4</sup> School of Science and Technology, University of London, London EC1V 0HB, UK
- <sup>5</sup> School of Health Sciences, Campus Coyoacán, Universidad del Valle de México, Coyoacán 04910, Mexico; a340298426@my.uvm.edu.mx
- <sup>6</sup> School of Health Sciences, Campus Chihuahua, Universidad del Valle de México, Chihuahua 31625, Mexico; emanuel.canto@uvmnet.edu
- \* Correspondence: rocio.duarte@uvmnet.edu (R.E.D.A.); calberto.gonzalez@uvmnet.edu (C.A.G.G.)

**Abstract:** This innovative study addresses the prevalent issue of sports injuries, particularly focusing on ankle injuries, utilizing advanced analytical tools such as artificial intelligence (AI) and machine learning (ML). Employing a logistic regression model, the research achieves a remarkable accuracy of 90.0%, providing a robust predictive tool for identifying and classifying athletes with injuries. The comprehensive evaluation of performance metrics, including recall, precision, and F1-Score, emphasizes the model's reliability. Key determinants like practicing sports with injury risk and kinesiophobia reveal significant associations, offering vital insights for early risk detection and personalized preventive strategies. The study's contribution extends beyond predictive modeling, incorporating a predictive factors analysis that sheds light on the nuanced relationships between various predictors and the occurrence of injuries. In essence, this research not only advances our understanding of sports injuries but also presents a potent tool with practical implications for injury prevention in athletes, bridging the gap between data-driven insights and actionable strategies.

**Keywords:** injury prediction; athlete health; machine learning; sports risk factors; preventive strategies; kinesiophobia



**Citation:** Ayala, R.E.D.; Granados, D.P.; Gutiérrez, C.A.G.; Ruíz, M.A.O.; Espinosa, N.R.; Heredia, E.C. Novel Study for the Early Identification of Injury Risks in Athletes Using Machine Learning Techniques. *Appl. Sci.* **2024**, *14*, 570. <https://doi.org/10.3390/app14020570>

Academic Editors: Dimitris Mourtzis and Arkady Voloshin

Received: 5 December 2023

Revised: 28 December 2023

Accepted: 5 January 2024

Published: 9 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Sports injuries, such as ankle injuries, are a common and recurring problem for many athletes, representing the first worldwide disability in terms of sports, typically manifesting 31% of the injuries in football players and 45% in basketball players [1]. These injuries can be debilitating and require a significant recovery time, taking the athletes out of the game for 3 or more weeks [2]. One of the most common ankle injuries are the sprains that cause chronic symptoms in 40% of individuals, such as pain, swelling, and instability, that provoke bad motor control of the ankle joint and can lead to functional disability [3].

Age is a factor that can impact an athlete given that the athlete's risk of injury increases with age. This is because the body's tissues begin the aging process after the age of 35, leading to a loss of elasticity and slower tissue repair [4].

Gender has a minimal impact on the risk of injury as the human body has an equal likelihood of getting injured regardless of gender [5].

The number of hours and days of training has a low impact on the incidence of injuries since the time dedicated to training contributes to muscle strengthening and memory [6].

Previous injuries are a variable that may yield a low result, as athletes with injuries undergo active recovery treatment, play with reduced intensity, and manage play time to prevent fatigue [7].

Corrective treatment is a variant with a small possibility of influencing injuries, as athletes either have undergone or undergo specific treatment for their injuries. Athletes who are injured tend to play with less intensity and in positions with lower physical demands. However, athletes with preventive treatment may have muscles and tissues that have been prepared for physical demands, potentially reducing the incidence of injuries [8,9].

Hydration is also an important factor in the possibility of getting injured, because it is involved in the metabolism, nutrient transport, blood circulation, and body temperature regulation [10]. In terms of sports, hypohydration plays an important part in the weight loss of athletes, because plasma, blood flow, blood volume, cardiovascular functions, and thermoregulatory capacity are mechanisms that are affected [10,11]. During training and games, athletes lose different body liquids, mainly by sweating, and these liquids depend on the exercise intensity and the weight of the individual.

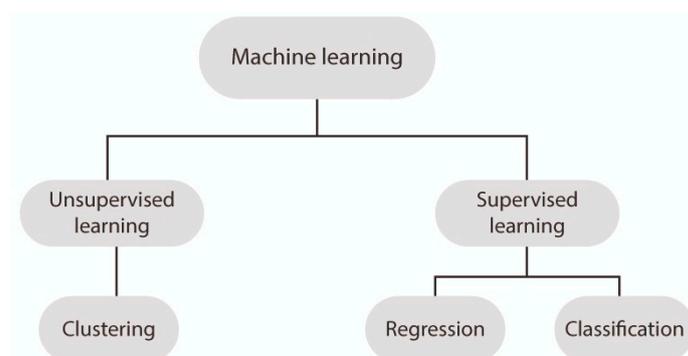
Kinesiophobia, known as the fear of performing a movement or an activity, plays a role in functional ankle instability. This fear induces negative thoughts, leading to mechanical alterations that impede proper joint functioning. Consequently, the repercussions of the ankle's behavior are a loss of strength, more postural balance, and an alteration to the proprioception of musculoskeletal disorders [3,12–16].

Fatigue in athletes can have a significant impact on their performance and increase the risk of injuries [17]. Fatigue can be the result of a variety of factors, including physical exertion, lack of sleep, poor nutrition, and mental stress [7,18]. Athletes who experience fatigue may have an increased risk of injuries due to several factors: alterations in coordination, reduction in performance capacity [19,20], and increased vulnerability to infection, also known as chronic fatigue [21,22].

Hydration is a factor that can have a high impact because muscles not adequately hydrated are prone to fatigue and injuries [23].

Artificial intelligence (AI) and machine learning (ML) are advancing in numerous fields, including medicine [24–27]. One of the most promising applications of AI in this field is the prediction and prevention of sports injuries.

ML is a subset of AI and is divided into supervised and unsupervised learning. In supervised learning a model is trained on a labeled dataset, and the input data are associated with a correct output [28–30]. The model learns from these data and then is ready to predict the output of new data. In health sciences, supervised learning can be useful for predicting diseases based on certain symptoms or risk factors [31,32]. See Figure 1.



**Figure 1.** Flow chart of the different machine learning techniques used.

### 1.1. Regression

Regression is a statistical technique used to identify the mathematical behavior of an unknown model. Its fundamental purpose lies in the identification of a mathematical formula that allows clarifying the existing correlation between these variables and projecting the value of the dependent variable based on the specific values assumed by the independent variables [33,34]. This procedure seeks to provide a deeper and more accurate understanding of the underlying relationship, thus allowing for more accurate predictions about the dependent variable based on the observations of the independent variables [35].

### 1.2. Logistic Regression Model

Logistic regression is a statistical method used to predict the likelihood of a dependent variable being in a certain category, based on one or more independent variables. For example, logistic regression can be used to estimate the probability of a patient having a disease, based on their symptoms, age, sex, etc. Logistic regression is based on the logistic function, which transforms input values into a range between 0 and 1, which is interpreted as the probability of belonging to the positive category. Logistic regression is applied to binary classification problems (where the dependent variable only has two possible values) or the multiclass (where the dependent variable has more than two possible values) [36–39].

The University of Valle de Mexico is home to accomplished Olympic athletes, professionals, and high-performance individuals. Annually, a massive event known as “Interlincas” takes place, featuring various sports such as soccer, tennis, American football, touch football, basketball, swimming, animation, taekwondo, gymnastics, and volleyball, involving a total of 3500 athletes in the year 2022 [40]. In the present work, a logistic regression model played a key role in the analysis and prediction of injuries in athletes within the framework of this research. We surveyed 500 athletes and took into consideration the different parameters of the individual, as presented in Section 2.

The significance of these parameters has an impact on injury prediction, whether the individuals have or have not been injured.

In this article, our contributions are:

- **A Cutting-Edge Predictive Model:** We developed an accurate logistic regression model with an accuracy of 90.0%, standing out as a leading tool in predicting sports injuries.
- **Identification of Determining Factors:** We revealed significant associations, such as practicing sports with a risk of injury and kinesiphobia, providing crucial insights for early risk detection and personalized preventive strategies.
- **A Comprehensive Performance Evaluation:** We conducted a thorough analysis of various machine learning models, highlighting the versatility of the logistic regression model and supporting its practical utility and reliability in medical and sports environments.
- **Detailed Performance Metrics:** Beyond high accuracy, we provided a detailed analysis with metrics such as recall and precision, offering a comprehensive evaluation of the model’s performance in crucial situations of accurate injury detection in athletes.

## 2. Materials and Methods

In this research, a sample of 400 athletes who participated in the comprehensive survey is presented in Table 1. It is stated that 50 independent data points were employed.

**Table 1.** Ankle injury status by sport.

Sport	Ankle Injured	Not Injured	Total
Soccer	167	22	189
Volleyball	90	10	100
Basketball	93	10	103
Tennis	7	1	8

In the survey database, we collected information related to various aspects such as age, gender, the number of hours dedicated to daily training, the frequency of weekly workouts, the history of previous ankle injuries, the medical treatment received after an injury, participation in sports practice despite injuries, the levels of stress experienced during sports activity, the presence of kinesiophobia, the fatigue experienced, the daily hydration average, and the amount of hydration during specific sports events.

### 2.1. Data Collection and Analysis Tools

The data obtained from the survey were refined into a database, providing a comprehensive view of each individual's health and sports practices while ensuring each individual's privacy. The outcome variable was classified into two categories: 400 injured and uninjured athletes. MATLAB R2023a was used as the main tool to carry out the analysis. MATLAB was chosen for its robust capability for efficient data manipulation and the application of machine learning techniques such as Fine Tree, Linear Discriminant, Binary GLM Logistic Regression, Gaussian Naive Bayes, Linear SVM, Fine KNN, SVM Kernel, Boosted Trees, and Logistic Regression.

### 2.2. Collection Dataset

The system was purposefully designed for data acquisition and comprehension, facilitating the examination of values, patterns, and trends that could contribute to ankle injuries in athletes. This functionality enhances the ability to predict and evaluate outcomes. A detailed description of the dataset is presented in Table 2.

**Table 2.** Predictor variables.

Variable	Value	Type
Age	A numerical value representing the age of the athlete.	Numeric
Gender	Categorical variable indicating the gender of the athlete.	Categorical
Hours of training	A numerical value representing the number of training hours.	Numeric
Days of training	A numerical value indicating the weekly frequency of training.	Numeric
Previous injuries	Categorical variable indicating if the athlete has had previous injuries.	Categorical
Corrective treatment	Categorical variable indicating if the athlete has received corrective treatment.	Categorical
Sport with injury	Categorical variable indicating if the athlete practices a sport with a risk of injuries.	Categorical
Preventive treatment	Categorical variable indicating if the athlete has received preventive treatment.	Categorical
Stress	Numerical value related to psychological factors.	Numeric
Kinesiophobia	Numerical value related to psychological factors.	Numeric
Fatigue	Numerical value related to physical factors.	Numeric
Previous warmup	Categorical variable indicating if the athlete performs a warmup before training.	Categorical
Average hydration	Numerical value related to average hydration.	Numeric
Hydration on event day	Numerical value related to hydration on the day of the event.	Numeric

The data set related to the injuries consists of 357 uninjured individuals and 43 injured individuals (see Table 3). It is stated that 50 independent data points were employed.

**Table 3.** Response variable.

Variable	Value	Type
Outcome	Binary variable indicating whether the athlete has experienced an injury.	Binary

### 2.3. Data Preprocessing

Data preprocessing is a crucial phase in data analysis and modeling and plays a key role in the quality and effectiveness of the results obtained. In the framework of this research on injuries to athletes, the preprocessing will address various tasks, ensuring that the data are accurate, reliable, and ready for analysis, as part of the preparation for the application of machine learning algorithms.

To address the challenge posed by class imbalance in our dataset, specific techniques were implemented during the model training process. Class weighting strategies were employed to assign greater importance to the minority class, and experimentation was conducted with subsampling methods, including the application of the synthetic minority over-sampling technique (SMOTE) [41,42]. This approach was implemented to alleviate the impact of class imbalance on logistic regression. Furthermore, performance metrics such as precision, recall, and the F1-Score were assessed to comprehensively capture the model's effectiveness in detecting injuries in athletes.

The initial database comprises 500 surveyed athletes. The database underwent a debugging process to ensure the integrity and consistency of the data. A total of 50 outliers, duplicates, and inconsistent records were identified and addressed. Data cleaning is essential to avoid biases and errors in the subsequent analysis.

### 3. Results of Data Training and Discussion

The architecture of the injury prediction model includes four different modules:

- Data set collection
- Data preprocessing
- Logistic regression modeling
- Evaluation.

#### 3.1. Logistic Regression Modeling

An exhaustive evaluation of various models was performed, and the accuracy achieved by each model is presented in Table 4, with features including age, gender, hours of training, days of training, previous injuries, corrective treatment, sport with injury, preventive treatment, stress, kinesiphobia, fatigue, previous warmup, average hydration on event day and outcome. In particular, the logistic regression model demonstrated exceptional accuracy, reaching an impressive accuracy rate of 90.0%. This means an outstanding ability to accurately classify athletes based on the presence or absence of injuries. Interestingly, the SVM Kernel, Linear Discriminant, and Binary GLM Logistic Regression models also showed high levels of accuracy, reaching 89.2%, 89.0%, and 89.0%, respectively. In contrast, the models with the lowest accuracy were the Fine Tree and the Gaussian Naive Bayes, with 83.2% and 84.8%, respectively. The application of logistic regression, known for its reliability, is crucial to achieving the desired results, and these findings have significant implications for identifying and preventing injuries in athletes.

**Table 4.** Response variable.

Model	Accuracy
Fine Tree	83.2%
Linear Discriminant	89.0%
Binary GLM Logistic Regression	89.0%
Gaussian Naive Bayes	84.8%
Linear SVM	89.2%
Fine KNN	86.0%
SVM Kernel	89.2%
Boosted Trees	87.8%
Logistic Regression	90.0%

The confusion matrix for the logistic regression model is illustrated in Table 5. It reveals that the model correctly predicted 6 cases without injuries and 354 cases with injuries. However, there were 37 cases where the model incorrectly predicted that there were no injuries when there were, and 3 cases where it incorrectly predicted injuries when there were none.

**Table 5.** Logistic regression confusion matrix.

	No Injury (Predicted)	Injury (Predicted)
No injury (Actual)	6	37
Injury (Actual)	3	354

### 3.1.1. Accuracy

The accuracy metric is essential to be able to evaluate the overall performance of the model. It is calculated using Equation (1).

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

where:

- TP (True positives) = 354;
- TN (True negatives) = 6;
- FP (False positives) = 37;
- FN (False negatives) = 3.

In our case:

$$\text{accuracy} = \frac{354 + 6}{354 + 6 + 37 + 3} = \frac{391}{400} \approx 0.90 \quad (2)$$

The accuracy achieved in Equation (2) is 0.90, which indicates that the model managed to correctly classify 90.0% of the cases, both negative and positive. This result is of great relevance as it suggests a robust ability of the model to discriminate between athletes with and without injuries.

### 3.1.2. Recall

Recall focuses on the ability of the model to correctly identify the positive cases. It is calculated using Equation (3):

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

In our case:

$$\text{recall} = \frac{354}{354 + 3} = \frac{354}{357} \approx 0.9916 \quad (4)$$

### 3.1.3. Precision

The precision in machine learning measures the proportion of correct positive predictions, highlighting the ability of the model to avoid false positives. It is calculated using Equation (5):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

In our case:

$$\text{Precision} = \frac{354}{354 + 37} = \frac{354}{391} \approx 0.9054 \quad (6)$$

### 3.1.4. F1-Score

F1-Score is a metric that combines accuracy and recall into a single measure. It is especially useful in situations where both metrics are important and a balance between them is sought. It is calculated using Equation (7).

$$\text{F1 - Score} = 2 \times \frac{\text{Accuracy} \times \text{Recall}}{\text{Accuracy} + \text{Recall}} \quad (7)$$

In our case:

$$\text{F1 - Score} = 2 \times \frac{0.90 \times 0.9916}{0.90 + 0.9916} \approx 0.9447 \quad (8)$$

As shown in Table 6, the high accuracy suggests that the model generally performs well in classifying cases. However, upon examining the recall, we observe that while it is high, there are some instances of false negatives (FNs). This means there are situations where the model did not correctly identify the presence of injuries, which is within the range in the context of athletes' health. The recall is approximately 0.9916, meaning the model correctly identified 99.16% of injury cases among all actual injury cases in athletes. This metric is crucial in contexts where identifying positive cases is of particular importance, such as in preventing injuries to athletes. The precision, with a value of approximately 0.9054, indicates that the model has a fairly high ability to correctly classify positive cases (athletes with injuries). In other words, when the model predicts that an athlete has an injury, there is a 90.54% chance that they have an injury. The F1-Score is approximately 0.9447. This score is relatively high and suggests a reasonable balance between the model's ability to correctly predict positive and negative cases.

**Table 6.** Model performance metrics.

Metric	Value
Accuracy	0.90
Recall	0.9916
Precision	0.9054
F1-Score	0.9447

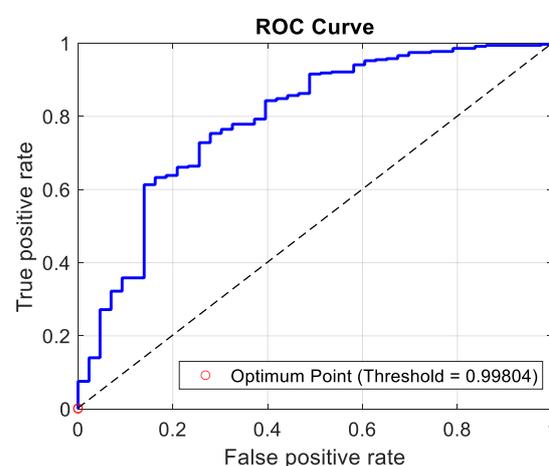
It is important to note that although the accuracy is high, a detailed analysis of other metrics like recall and precision provides a more complete view of the model's performance, especially in situations where identifying positive cases is critical. That is why it is essential to consider these metrics together to obtain a comprehensive evaluation of the model's performance. In medical and sports applications, where accurate identification of injuries is crucial, this detailed analysis allows for informed decisions about the practical utility of the model in the specific research domain.

It is worth mentioning that 50 independent athletes were considered as a training sample. By using this, we emphasize that the model works correctly for this study.

### 3.2. Receiver Operating Characteristics (ROC) and Area under the Curve (AUC)

The receiver operating characteristic (ROC curve) is a graphical representation that shows the relationship between the true positive rate (TPR or recall) and the false positive rate (FPR) for different classification thresholds. The area under the ROC curve (AUC) quantifies the model's ability to distinguish between classes. An AUC of 79.15% indicates reasonable performance.

As shown in Figure 2, the decision-making threshold selected is 0.99804. This threshold influences how the model classifies instances, with an emphasis on precision.



**Figure 2.** Receiver Operating Characteristics.

### 3.3. Predictive Factors Analysis

In this analysis, a generalized linear regression model with binomial distribution was applied to evaluate the relationship between an outcome variable and 14 potential predictors. Table 7 shows the estimated coefficients, providing key information about the strength and direction of these associations.

**Table 7.** Factors affecting sports injuries.

Variable	Estimate	Standard Error	tStat	p Value
(Intercept)	−1.6796	1.6205	−1.0365	0.29998
Age	−0.0014065	0.041393	−0.03398	0.97289
Gender	0.7634	0.37307	2.0463	0.040728
Training Hours	0.17989	0.17137	1.0497	0.29385
Training Days	0.28663	0.15229	1.8821	0.05982
Previous Injuries	0.33075	0.57787	0.57235	0.56708
Corrective Treatment	0.35246	0.55783	0.63185	0.52748
Sport with Injury	−1.0194	0.47542	−2.1442	0.03202
Preventive Treatment	−0.2304	0.37792	−0.60966	0.54209
Stress	0.019871	0.007484	2.6552	0.007927
Kinesiophobia	0.58079	0.16835	3.4498	0.00056105
Fatigue	0.04619	0.12905	0.35792	0.7204
Previous Warmup	−0.37725	0.22096	−1.7073	0.087765
Average Hydration	0.3888	0.21644	1.7963	0.072445
Event Day Hydration	−0.14376	0.090605	−1.5867	0.11259

Table 7 of estimated coefficients provides key information about the strength and direction of these associations, while additional measures such as the  $p$ -Value,  $\chi^2$ -statistic, and dispersion offer insights into the model's goodness of fit.

Athletes who practice their sport with an active injury have a low probability of getting injured because coaches decide to allow them to play for a shorter period or in positions with lower physical demands. During the development of academy soccer players (ASPs), specific skills or physical qualities can lead to players being selected for certain playing positions due to variations in the tactical and physiological requirements of those positions. In professional soccer, goalkeepers occupy the majority of low intensity actions, unlike outfield players, who exhibit more running, ball possession, and high-intensity activity. However, the distance covered and the frequency of game actions within the match among outfield positions may contribute to the different physical demands experienced by field ASPs [43]. Likewise, the variable "Sport with injury" presents a coefficient of  $-1.0194$  and a  $p$ -value of  $0.03202$ , indicating a significant negative association. Those involved in "Sports with injury" may have a lower probability of achieving the desired outcome.

On the other hand, "Kinesiophobia" shows a significant positive association (coefficient =  $0.58079$ ,  $p$ -value =  $0.00056105$ ), suggesting that kinesiophobia is positively related to the outcome variable.

In addition, observation of action interventions and game techniques can be effective in improving the rehabilitation outcomes of lower limb injuries. Therefore, their application should be considered along with standard treatment protocols. This allows us to employ specific strengthening of the injured muscles, as well as correcting the game technique, resulting in athletes reducing their probability of injury [44].

#### 3.3.1. Non-Significant Variables

Several variables, such as "Age", "Training Hours", "Previous Injuries", and others, show no significant association as their  $p$ -values are greater than  $0.05$ . These results indicate that these variables may not be determining factors in the outcome.

### 3.3.2. Overall Model Evaluation

The model as a whole is evaluated by the  $\chi^2$   $p$ -value of  $6.15 \times 10^{-6}$ , indicating that at least one of the predictor variables has a significant effect on the outcome variable. A dispersion of 1 suggests that the model fits the binomial distribution adequately.

## 4. Conclusions

This novel study with original data provides an effective tool for predicting injuries in athletes, and the importance of considering detailed metrics for a comprehensive evaluation of the model's performance. We, the authors, consider this work to be of great importance as it offers key perspectives that could revolutionize injury prevention in athletes, contributing to their health and optimal performance.

The culmination of this study, which addressed the prediction of injuries in athletes using a logistic regression model, represents a significant advance in understanding and preventing sports risks. The robustness of the model, backed by an accuracy of 90.0%, underscores its effectiveness in classifying athletes based on the presence or absence of injuries.

In the detailed analysis of performance metrics, a high recall of 99.16% was observed, indicating the model's ability to correctly identify athletes with injuries. This metric is essential in contexts where accurate detection of positive cases is crucial, such as in injury prevention in sports.

The precision of 90.54% reinforces confidence in the model's ability to correctly classify positive cases, underlining its practical utility. The F1-Score, which combines precision and recall, showed a reasonable balance of 94.47%, consolidating the overall effectiveness of the model in situations where both metrics are fundamental.

It is relevant to highlight the importance of key variables, such as practicing sports with injury and kinesiophobia, which demonstrated significant associations with sports injuries. These findings offer valuable information that can be fundamental in the early identification of risk factors and the implementation of personalized preventive strategies.

## 5. Future Work

Shortly, a second survey will be conducted with a larger sample of athletes, aiming to expand the database and refine the acceptance percentages of the injury prediction algorithm in addition to including the sport that athletes practice. This initiative aims not only to validate and improve the robustness of the model but also to provide a more complete and representative view of various sports conditions and practices. In addition, the development of a dedicated mobile application that will allow real-time injury prediction is being considered. This application will be an invaluable tool at sports events, offering the ability to anticipate possible injuries and providing timely preventive measures, thus reinforcing the attention to and comprehensive care of athletes' health in high-performance situations. This innovative approach integrates technology and research to advance prevention and care of sports injuries. In addition, a detailed investigation will be carried out regarding the age range most affected by kinesiophobia.

**Author Contributions:** Conceptualization, R.E.D.A., D.P.G. and E.C.H. methodology, R.E.D.A., D.P.G. and E.C.H.; validation, D.P.G., M.A.O.R. and C.A.G.G.; formal analysis, D.P.G., M.A.O.R. and C.A.G.G.; investigation, R.E.D.A., D.P.G. and N.R.E.; data curation, D.P.G., M.A.O.R. and C.A.G.G.; writing—original draft preparation, D.P.G. and N.R.E. writing—review and editing, R.E.D.A., M.A.O.R., C.A.G.G., E.C.H., N.R.E. and D.P.G.; visualization, D.P.G. and R.E.D.A. supervision, R.E.D.A.; project administration, R.E.D.A. and D.P.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** The analyzed data were anonymized by athletes. Informed consent was obtained from all subjects during the data acquisition.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Crowell, M.S.; Mason, J.S.; Morris, J.B.; Dummar, M.K.; Kuwik, P.A. Diagnostic Imaging for Distal Extremity Injuries in Direct Access Physical Therapy: An Observational Study. *Int. J. Sports Phys. Ther.* **2023**, *18*, 431–438. [[CrossRef](#)] [[PubMed](#)]
2. Ruiz-Sánchez, F.J.; Ruiz-Muñoz, M.; Martín-Martín, J.; Coheña-Jimenez, M.; Perez-Belloso, A.J.; Pilar Romero-Galisteo, R.; González-Sánchez, M. Management and Treatment of Ankle Sprain According to Clinical Practice Guidelines: A PRISMA Systematic Review. *Medicine* **2022**, *101*, E31087. [[CrossRef](#)] [[PubMed](#)]
3. Alshahrani, M.S.; Reddy, R.S. Relationship between Kinesiophobia and Ankle Joint Position Sense and Postural Control in Individuals with Chronic Ankle Instability-A Cross-Sectional Study. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2792. [[CrossRef](#)] [[PubMed](#)]
4. Maas, A.I.R.; Menon, D.K.; Manley, G.T.; Abrams, M.; Åkerlund, C.; Andelic, N.; Aries, M.; Bashford, T.; Bell, M.J.; Bodien, Y.G.; et al. Traumatic Brain Injury: Progress and Challenges in Prevention, Clinical Care, and Research. *Lancet Neurol.* **2022**, *21*, 1004–1060. [[CrossRef](#)] [[PubMed](#)]
5. Daetz, C.D.; Toro, F.R.; Mendoza, V.T. Lesiones Deportivas En Deportistas Universitarios Chilenos (Sports Injuries in Chilean University Athletes). *Retos* **2020**, *38*, 490–496. [[CrossRef](#)]
6. Page, R.M.; Field, A.; Langley, B.; Harper, L.D.; Julian, R. The Effects of Fixture Congestion on Injury in Professional Male Soccer: A Systematic Review. *Sports Med.* **2023**, *53*, 667–685. [[CrossRef](#)]
7. Taylor, A.H.; Dorn, L. Stress, Fatigue, Health, and Risk of Road Traffic Accidents Among Professional Drivers: The Contribution of Physical Inactivity. *Annu. Rev. Public Health* **2006**, *27*, 371–391. [[CrossRef](#)]
8. Soto, M.d.V.; Marqueta, P.M.; Tarrero, L.T.; González, B.M.; Heredia, Á.G.d.I.R.; Bonafonte, L.F.; Galván, C.d.T.; Ansón, J.P.; Aurrekoetxea, T.G.; Díaz, J.F.J.; et al. Lesiones Deportivas “versus” Accidentes Deportivos. Documento de Consenso. Grupo de Prevención En El Deporte de La Sociedad Española de Medicina Del Deporte (SEMED-FEMEDE). *Arch. Med. Deporte Rev. Fed. Esp. Med. Deporte Confed. Iberoamericana Med. Deporte* **2018**, *35*, 6–16.
9. Perez-De-Arrilucea-Le-Floc’h, U.A.; Dote-Montero, M.; Carle-Calo, A.; Sánchez-Delgado, G.; Ruiz, J.R.; Amaro-Gahete, F.J. Acute Effects of Whole-Body Electromyostimulation on Energy Expenditure at Resting and during Uphill Walking in Healthy Young Men. *Metabolites* **2022**, *12*, 781. [[CrossRef](#)]
10. Barley, O.R.; Chapman, D.W.; Abbiss, C.R. Reviewing the Current Methods of Assessing Hydration in Athletes. *J. Int. Soc. Sports Nutr.* **2020**, *17*, 52. [[CrossRef](#)]
11. Belval, L.N.; Hosokawa, Y.; Casa, D.J.; Adams, W.M.; Armstrong, L.E.; Baker, L.B.; Burke, L.; Chevront, S.; Chiampas, G.; González-Alonso, J.; et al. Practical Hydration Solutions for Sports. *Nutrients* **2019**, *11*, 1150. [[CrossRef](#)] [[PubMed](#)]
12. Gómez-Pérez, L.; López-Martínez, A.E.; Ruiz-Párraga, G.T. Psychometric Properties of the Spanish Version of the Tampa Scale for Kinesiophobia (TSK). *J. Pain* **2011**, *12*, 425–435. [[CrossRef](#)]
13. Romero, E.A.S.; Lim, T.; Villafañe, J.H.; Boutin, G.; Aguado, V.R.; Martín Pintado-Zugasti, A.; Luis, J.; Pérez, A.; Fernández Carnero, J.; Romero, S.; et al. The Influence of Verbal Suggestion on Post-Needling Soreness and Pain Processing after Dry Needling Treatment: An Experimental Study. *Int. J. Environ. Res. Public Health* **2021**, *18*, 4206. [[CrossRef](#)] [[PubMed](#)]
14. Fernández-Carnero, J.; Beltrán-Alacreu, H.; Arribas-Romano, A.; Cerezo-Téllez, E.; Cuenca-Zaldivar, J.N.; Sánchez-Romero, E.A.; Lerma Lara, S.; Villafañe, J.H. Prediction of Patient Satisfaction after Treatment of Chronic Neck Pain with Mulligan’s Mobilization. *Life* **2023**, *13*, 48. [[CrossRef](#)] [[PubMed](#)]
15. Bordeleau, M.; Vincenot, M.; Lefevre, S.; Duport, A.; Seggio, L.; Breton, T.; Lelard, T.; Serra, E.; Roussel, N.; Das Neves, J.F.; et al. Treatments for Kinesiophobia in People with Chronic Pain: A Scoping Review. *Front Behav. Neurosci.* **2022**, *16*, 933483. [[CrossRef](#)] [[PubMed](#)]
16. Milá, Z.S.; Muñoz, T.V.; Sánchez, M.d.R.F.; Llanes, R.F.; Casas, J.M.B.; Sanz, D.R.; Saornil, J.V. Therapeutic Exercise Parameters, Considerations and Recommendations for the Treatment of Non-Specific Low Back Pain: International DELPHI Study. *J. Pers. Med.* **2023**, *13*, 1510. [[CrossRef](#)] [[PubMed](#)]
17. Slobounov, S. Fatigue-Related Injuries in Athletes. In *Injuries in Athletics: Causes and Consequences*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 77–95. [[CrossRef](#)]
18. Mihajlovic, M.; Cabarkapa, D.; Cabarkapa, D.V.; Philipp, N.M.; Fry, A.C. Recovery Methods in Basketball: A Systematic Review. *Sports* **2023**, *11*, 230. [[CrossRef](#)]
19. Cooper, C.N.; Dabbs, N.C.; Davis, J.; Sauls, N.M. Effects of Lower-Body Muscular Fatigue on Vertical Jump and Balance Performance. *J. Strength Cond. Res.* **2020**, *34*, 2903–2910. [[CrossRef](#)]
20. Bellenger, C.R.; Arnold, J.B.; Buckley, J.D.; Thewlis, D.; Fuller, J.T. Detrended Fluctuation Analysis Detects Altered Coordination of Running Gait in Athletes Following a Heavy Period of Training. *J. Sci. Med. Sport* **2019**, *22*, 294–299. [[CrossRef](#)]

21. Simpson, R.; Campbell, J.; Gleeson, M.; Krüger, K.; Nieman, D.; Pyne, D.; Turner, J.; Walsh, N. Can Exercise Affect Immune Function to Increase Susceptibility to Infection? *Exerc. Immunol. Rev.* **2020**, *26*, 8–22.
22. Reid, V.L.; Gleeson, M.; Williams, N.; Clancy, R.L. Clinical Investigation of Athletes with Persistent Fatigue and/or Recurrent Infections. *Br. J. Sports Med.* **2004**, *38*, 42–45. [[CrossRef](#)] [[PubMed](#)]
23. Rowlands, D.S.; Kopetschny, B.H.; Badenhorst, C.E. The Hydrating Effects of Hypertonic, Isotonic and Hypotonic Sports Drinks and Waters on Central Hydration During Continuous Exercise: A Systematic Meta-Analysis and Perspective. *Sports Med.* **2022**, *52*, 349–375. [[CrossRef](#)] [[PubMed](#)]
24. Skoki, A.; Napravnik, M.; Polonijo, M.; Štajduhar, I.; Lerga, J. Revolutionizing Soccer Injury Management: Predicting Muscle Injury Recovery Time Using ML. *Appl. Sci.* **2023**, *13*, 6222. [[CrossRef](#)]
25. González-Alday, R.; García-Cuesta, E.; Kulikowski, C.A.; Maojo, V. A Scoping Review on the Progress, Applicability, and Future of Explainable Artificial Intelligence in Medicine. *Appl. Sci.* **2023**, *13*, 10778. [[CrossRef](#)]
26. Antoniadi, A.M.; Du, Y.; Guendouz, Y.; Wei, L.; Mazo, C.; Becker, B.A.; Mooney, C. Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review. *Appl. Sci.* **2021**, *11*, 5088. [[CrossRef](#)]
27. Nardi, C. Special Issue on Artificial Intelligence in Medical Imaging: The Beginning of a New Era. *Appl. Sci.* **2023**, *13*, 11562. [[CrossRef](#)]
28. Nasteski, V. An Overview of the Supervised Machine Learning Methods. *Horizons. B* **2017**, *4*, 51–62. [[CrossRef](#)]
29. Cunningham, P.; Cord, M.; Delany, S.J. Supervised Learning. In *Cognitive Technologies*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 21–49. [[CrossRef](#)]
30. Albalawi, F.; Alamoud, K.A. Trends and Application of Artificial Intelligence Technology in Orthodontic Diagnosis and Treatment Planning—A Review. *Appl. Sci.* **2022**, *12*, 11864. [[CrossRef](#)]
31. Uddin, S.; Khan, A.; Hossain, M.E.; Moni, M.A. Comparing Different Supervised Machine Learning Algorithms for Disease Prediction. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 281. [[CrossRef](#)]
32. Bratić, B.; Kurbalija, V.; Ivanović, M.; Oder, I.; Bosnić, Z. Machine Learning for Predicting Cognitive Diseases: Methods, Data Sources and Risk Factors. *J. Med. Syst.* **2018**, *42*, 243. [[CrossRef](#)]
33. Black, I.M.; Richmond, M.; Kolios, A. Condition Monitoring Systems: A Systematic Literature Review on Machine-Learning Methods Improving Offshore-Wind Turbine Operational Management. *Int. J. Sustain. Energy* **2021**, *40*, 923–946. [[CrossRef](#)]
34. Wang, H.; Barone, G.; Smith, A. Current and Future Role of Data Fusion and Machine Learning in Infrastructure Health Monitoring. *Struct. Infrastruct. Eng.* **2023**. [[CrossRef](#)]
35. Drogkoula, M.; Kokkinos, K.; Samaras, N. A Comprehensive Survey of Machine Learning Methodologies with Emphasis in Water Resources Management. *Appl. Sci.* **2023**, *13*, 12147. [[CrossRef](#)]
36. Nick, T.G.; Campbell, K.M. Logistic Regression. *Methods Mol. Biol.* **2007**, *404*, 273–301. [[CrossRef](#)]
37. Boateng, E.Y.; Abaye, D.A.; Boateng, E.Y.; Abaye, D.A. A Review of the Logistic Regression Model with Emphasis on Medical Research. *J. Data Anal. Inf. Process.* **2019**, *7*, 190–207. [[CrossRef](#)]
38. Stoltzfus, J.C. Logistic Regression: A Brief Primer. *Acad. Emerg. Med.* **2011**, *18*, 1099–1104. [[CrossRef](#)] [[PubMed](#)]
39. Zhu, B.; Shi, Y.; Hao, J.; Fu, G. Prediction of Coal Mine Pressure Hazard Based on Logistic Regression and Adagrad Algorithm—A Case Study of C Coal Mine. *Appl. Sci.* **2023**, *13*, 12227. [[CrossRef](#)]
40. Prensa UVM. La UVM da el Banderazo de Inicio a la XV Edición de sus Juegos Deportivos Interlincos—Sala de Prensa UVM. Available online: <https://laureate-comunicacion.com/prensa/la-uvm-da-el-banderazo-de-inicio-a-la-xv-edicion-de-sus-juegos-deportivos-interlincos/> (accessed on 27 November 2023).
41. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-Sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
42. Elreedy, D.; Atiya, A.F.; Kamalov, F. A Theoretical Distribution Analysis of Synthetic Minority Oversampling Technique (SMOTE) for Imbalanced Learning. *Mach. Learn.* **2023**. [[CrossRef](#)]
43. Hall, E.C.R.; Larruskain, J.; Gil, S.M.; Lekue, J.A.; Baumert, P.; Rienzi, E.; Moreno, S.; Tannure, M.; Murtagh, C.F.; Ade, J.D.; et al. Playing Position and the Injury Incidence Rate in Male Academy Soccer Players. *J. Athl. Train.* **2022**, *57*, 696–703. [[CrossRef](#)]
44. Nanbancha, A.; Mawhinney, C.; Sinsurin, K. The Effect of Motor Imagery and Action Observation in the Rehabilitation of Lower Limb Injuries: A Scoping Review. *Clin. Rehabil.* **2023**, *37*, 145–161. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.