Sports Injuries Classification Using Machine Learning Models on Biomechanical Data

Emilia Aguirre
Colegio de Ciencias e Ingenierías
Universidad San Francisco de Quito
USFQ
Quito, Ecuador
eaguirres@estud.usfq.edu.ec

Noel Pérez-Pérez
Colegio de Ciencias e Ingenierías
Universidad San Francisco de Quito
USFQ
Quito, Ecuador
nperez@usfq.edu.ec

***Abstract*—The early detection and prevention of running related injuries are fundamental to protecting athlete health and optimizing performance outcomes. Despite this importance, early diagnosis remains a significant challenge due to the subtle and often nonspecific nature of initial symptoms, as well as the dependence on subjective clinical judgment. To address these limitations, this work proposes a machine learning-based classification framework aimed at enhancing the identification of injury patterns among runners, thereby improving diagnostic accuracy. The proposed method explores five classification models, such as random forest, three different feed-forward back propagation neural networks, support vector machine, K-Nearest Neighbors, and Gaussian naive Bayes on a comprehensive dataset encompassing biomechanical, anthropometric, demographic, and training history variables. The feed-forward back propagation neural network with 1544 and 772 neurons in the first and second hidden layer was the best model, achieving the highest F1-score of 0.980 and 0.983 in the training and test phases, respectively. The consistent performance on unseen data demonstrated the model's robust learning capability and strong generalization in classifying running-related injuries. These results underscore the promise of machine learning approaches in supporting objective and scalable decision-making within sports injury prevention and management.**

***Keywords*—Running injuries, Machine learning, Neural networks, Injury classification, Biomechanics, Predictive modeling**

I. Introduction

Running is one of the most widely practiced physical activities worldwide. It is popular because it is accessible, inexpensive, and offers clear health benefits, including improved cardiovascular fitness and mental well-being. However, despite its simplicity, running is also associated with a high rate of injuries, especially in the lower body. According to a recent review by Rahlf et al. [1], more than 70% of running-related injuries are due to overuse. These injuries usually affect parts of the body that absorb repetitive impact forces, such as the knees, ankles, lower legs, and feet [2].

Studies have reported that between 19% and 79% of runners experience injuries each year, depending on factors such as training intensity, running surface, type of footwear, and level of runner experience [1], [3]. This wide range highlights how many different variables can influence injury risk and shows the limits of traditional injury tracking methods, which often miss early warning signs and are highly dependent on subjective clinical judgment.

Traditional methods of preventing injuries typically include physical exams, visual evaluations, and expert opinion. However, injuries, such as patellofemoral pain syndrome, Achilles tendinopathy, medial tibial stress syndrome, and iliotibial band syndrome, are especially common and continue to affect runners despite preventive efforts [4]. Together, these conditions make up nearly half of all running injuries. One of the reasons they are so persistent is that traditional tools often detect them only after symptoms appear. Although traditional efforts are helpful, they are limited by differences between observers and by the challenge of analyzing large and complex datasets. Modern tools such as motion sensors and fitness trackers collect detailed biomechanical data during training, which requires reliable and scalable analysis methods. [5] Machine learning techniques emerge as powerful mechanisms to analyze these data, offering an opportunity to early identify the risk of injury from running form, body structure, training habits, and demographic information [6]–[8].

Recent research supports the use of machine learning classifiers (MLCs) for the prediction of injuries in sports. Models such as support vector machines (SVMs), random forests (RF), and feedforward backpropagation (FFBP) neural networks have shown accuracy rates between 75% and 90% in predicting injuries [9]. More advanced models, including deep learning architecture connected to wearable devices, have achieved real-time detection accuracies of 96.3% [1], [3]. The use of MLCs has also outperformed traditional statistical methods in predicting injury risk over time, with area under the receiver characteristic curve (AUC) values of 0.73 and 0.70 at 30 days and 180 days [8].Furthermore, Alghamdi et al. (2023) proposed a deep recurrent neural network model utilizing wearable sensor data to monitor physiological indicators, achieving superior predictive accuracy over traditional linear models [10]. Similarly, Leckey et al. (2025) performed a comprehensive synthesis showing Random Forests and XGBoost as the most frequently validated methods across various sports contexts [11]. Other studies have demonstrated that SVMs can outperform decision trees and logistic regression for musculoskeletal injury prediction in athletes [12]. Wu et al. (2022) explored cloud-based DL systems for real-time injury classification using RF and CNN models with high success in multi-class categorization tasks [13]. Murugan et al. (2025) compared RF, SVM, and ANN models, noting that the inclusion of real-time biomechanical input from wearables significantly boosted model robustness [14]. Finally, van Eetvelde et al. (2021) stressed that ensemble techniques like Random Forest and feature selection improve both generalizability and interpretability in injury prediction models [15].

Despite the developed efforts in this area, there are still opportunities to improve the classification performance of sports lesions on biomechanical data, especially for running-related injuries. Therefore, this study proposes a sport lesion classification framework based on five MLCs to maximize the binary classification performance of runners on biomechanical data. The main contributions are related to exploring the feature space using different taxonomical classifiers such as RF, SVMs, FFBP neural networks, k-nearest neighbors (kNNs), and Gaussian naive Bayes (GNB) to gather information from different viewpoints while maximizing the binary classification of injured and uninjured runners.

The rest of the paper is organized as follows: Materials and Methods section outlines ......

II. Materials and Methods

A. Database

We use a publicly available biomechanical database developed by Ferber et al [16], comprising treadmill gait recordings from 1,798 participants. Data collection occurred between 2009 and 2017 at the University of Calgary Running Injury Clinic and was conducted with the approval of the University of Calgary's Conjoint Health Research Ethics Board (Ethics ID: E–21705). Each participant underwent a treadmill-based motion analysis session, which included both walking and running trials. The dataset contains comprehensive metadata for each subject, including demographic, clinical, and biomechanical information—25 features in total. Expert clinicians confirmed injury diagnoses to ensure accurate labeling of injury status. Participants were assigned to one of four injury categories based on severity and training impact: (1) No injury (659 samples), (2) Continuing to train in pain (320 samples), (3) Training volume or intensity affected (499 samples), and (4) Missed at least two consecutive workouts (274 samples). The dataset includes three-dimensional (3D) motion capture (MoCap) kinematic data aligned with the metadata. Although the original dataset featured 25 biomechanical variables, three were excluded due to redundancy and low variance, resulting in a final feature set of 22 clinically and biomechanically relevant inputs. To maintain the interpretability and clinical significance of each variable, no dimensionality reduction techniques—such as Principal Component Analysis (PCA)—were applied [17].

B. Proposed method

Machine learning techniques have emerged as a powerful tool in sports science due to their ability to analyze complex, high-dimensional data such as gait kinematics, joint angles, and force measurements [17], [18]. Unlike traditional rule-based systems, these models can generalize patterns across heterogeneous runner profiles and injury types, enabling scalable and adaptive injury detection [19]. Additionally, Most of them are computationally efficient and well-suited for moderate-sized datasets, which is often the case in clinical and experimental biomechanics [20].

The proposed method is based on the use of five MLCs, such as RF, SVMs, FFBP neural networks, kNN, and GNB, to maximize the binary classification (injured or uninjured) of runners' biomechanical data, as shown in Fig. 1. The selected classifiers are shallow learning models that belong to various taxonomies for tackling the problem from different viewpoints without incurring a high computational cost. A brief architectural description of each model is provided below:

- The RF classifier is an ensemble model using several tree-based predictors, where each tree is trained on a different bootstrap sample of the training data and a random subset of features. The final prediction is made by aggregating the predictions of all trees, typically via majority voting. This technique reduces variance and helps prevent overfitting, making it robust and highly accurate for classification. [21]
- The SVM is a discriminative classifier that constructs a hyperplane or set of hyperplanes in a high-dimensional space, which can be used for classification, regression, or other tasks. It aims to maximize the margin between classes by relying on a subset of the training data called support vectors. Through kernel functions, SVMs can efficiently perform non-linear classification. [22]
- The FFBP is a traditional multilayer perceptron architecture, consisting of input, two hidden, and output layers. The network propagates input data forward through the layers using weighted sums and activation functions, and adjusts the weights through the backpropagation algorithm to minimize the prediction error. We created three FFBP architectures by v3arying the number of neurons per hidden layer. Hence, the FFBP1 was formed with 1544 and 772 neurons in the first and second hidden layers, respectively. The FFBP2 contained 386 and 772 neurons, and the FFBP3 incorporated 386 and 193 neurons in their hidden layers. All FFBP models used ReLU activation function in the input and hidden layers, and Sigmoid in the one neuron of the output layer to introduce non-linearity and enhance learning capacity. These variations were determined empirically to evaluate the effect of network capacity on classification. [23]
- The KNN classifier is a non-parametric method that assigns a class label to a new sample based on the majority class among its k nearest neighbors, using a distance metric—typically Euclidean—to measure similarity. Since no model is explicitly trained, KNN is considered a lazy learner and is highly sensitive to the local structure of the data [24].

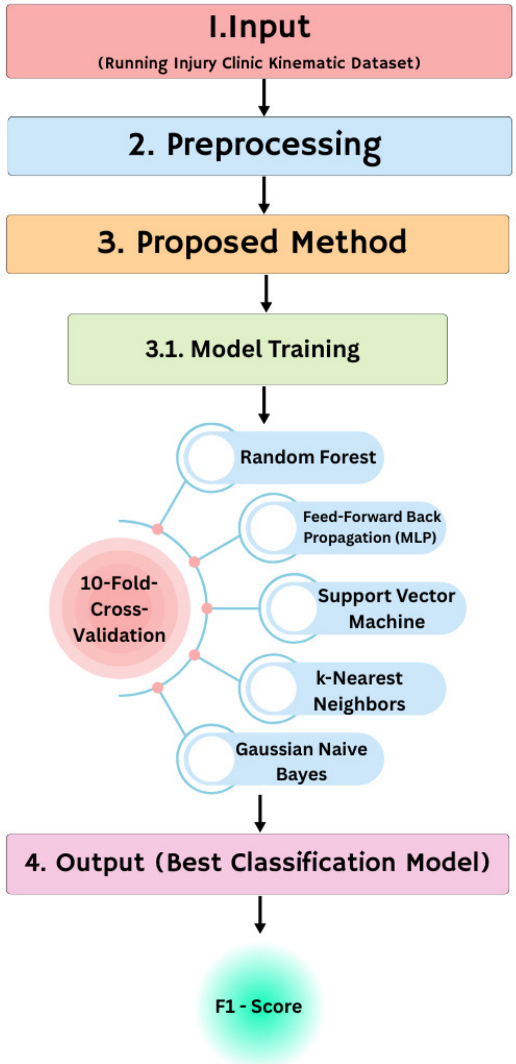

Fig. 1: Workflow of the proposed method.

- The GNB is a probabilistic classifier based on the Bayes theorem, with the simplified assumption that the features are independent given the class label. Each input feature is modeled as a Gaussian distribution characterized by its mean and variance. The model computes posterior probabilities for each class and assigns the label with the highest probability [25].

C. Experimental setup

1) Experimental dataset creation and preprocessing: Since our objective is to determine whether or not a runner is injured, rather than classify the specific type of injury, we binarized the 'InjDefn' labels. Specifically, the categories "training volume/intensity affected" (499 samples), "continuing to train in pain" (320 samples), and "two workouts missed in a row" (274 samples) were merged into a single injured class, resulting in a total of 1093 injured samples. The non-injured class, corresponding to the label "no injury", remains unchanged with 659 samples. This transformation adheres strictly to the preprocessing protocols described in [16].

Subsequently, all data were normalized in the range from 0 to 1 using the min-max function, avoiding any data dispersion and facilitating the further classification step.

2) Training, validation, and test sets: We applied a classical $70\% - 30\%$ split for building the initial training and test sets. The training partition feeds a stratified ten-fold cross-validation scheme to ensure that each fold conserves the original class distribution [26]. The test set was reserved to assess the model's generalization power on unseen data.

3) Model Configuration: All MLCs were optimized to determine the best classification performance conditions. In this sense, the RF set the number of tree-based predictors to 75, 100, and 125 trees, using bootstrap sampling and the Gini impurity criterion. All FFBP models were trained for 1000 epochs using mini-batches of size 20. The Adam optimizer was used with a learning rate of 10^{-4} to update model weights. To prevent overfitting and facilitate monitoring, model checkpoints were saved every 20 epochs. The SVM used three kernel types, linear, polynomial, and radial basis function (RBF), the regularization parameter $C = 1.0$, and the kernel coefficient γ to *scale*, exploring both linear and non-linear decision boundaries. The kNN optimized the number of neighbors to $k = 3, 5, 7$ with the Euclidean distance metric and a uniform voting scheme among neighbors. The GNB modeled each feature as a Gaussian distribution conditioned on the class label, with variance smoothing of 1×10^{-9} for numerical stability.

4) Assessment metrics: We computed the mean of F1-score, area under the receiver operating characteristic curve (AUC), accuracy (ACC), precision (PRE), and recall (REC) to measure the models' classification performance. The F1-score was considered as the main metric to analyze and discuss the obtained results because is particularly suitable for binary classification with class imbalance [27]. Additionally, the Wilcoxon signed-rank test (two-tailed, $\alpha = 0.05$), which is a non-parametric approach appropriate for paired data without normality assumptions [28] was computed to assess the statistical F1-score-based classification performance.

5) Model Selection Rule: The final model was selected based on a hierarchical rule set: (1) highest mean F1-score across cross-validation folds; (2) in case of a tie performance, preference for lower algorithmic complexity; (3) if still tied within the same model family, selection of the model requiring fewer training epochs. This ensures a transparent trade-off between predictive performance, simplicity, and computational cost.

6) Development and execution platform: All experiments were implemented in Python 3.11.12 using open-source libraries *scikit-learn* and *TensorFlow*. The complete pipeline, from data preprocessing to model training, validation, and evaluation, was executed on a MacBook Air running macOS, equipped with a 1.8 GHz dual-core Intel Core i5 processor and 8 GB of RAM.

III. Results and Discussion

This section presents and analyzes the performance results of the five classification models, RF, FFBP, SVM, kNN, and GNB, which were trained and validated using a biomechanical dataset consisting of 1798 participants performing treadmill-based running tasks. Performance metrics were computed using a stratified ten-fold cross-validation scheme on the training set. The results, including mean and standard deviation values for each evaluation metric, are summarized in Table I.

A. Training performance evaluation

From table I, it is possible to read that the FFBP architectures achieved the highest overall performance among the evaluated models. Notably, FFBP1 model achieved the highest mean F1-score of 0.980, outperforming the remaining MLCs . This result was expected since the FFBP neural network is a trainable model with higher learning capability. Also, it indicates that FFBP1 model correctly identified almost all injured instances, demonstrating exceptional sensitivity and robustness.

The other FFBP variants, namely FFBP2 and FFBP3, also reported excellent results, achieving mean F1-scores of 0.977 ± 0.012 and 0.979 ± 0.010, respectively. A Wilcoxon signed-rank test yielded $p = 1.000$, indicating no statistically significant difference between these two configurations. This statistical equivalence suggests that all three FFBP models exhibit similar and consistently strong predictive performance, making them the top-performing classifiers in this evaluation.

Among traditional machine learning models, the Random Forest (RF) algorithm demonstrated competitive results. The configuration with $t = 125$ trees achieved a mean F1-score of 0.983 ± 0.006, which is high and comparable at first glance. However, statistical testing ($p < 0.05$) confirmed that the differences between RF and the FFBP models were significant, reinforcing the superior performance of the neural network-based approaches.

Support Vector Machines (SVM) also performed reliably across kernel choices. The radial basis function (RBF) kernel yielded an F1-score of 0.976 ± 0.009, followed closely by the polynomial kernel (0.974 ± 0.011) and linear kernel (0.977 ± 0.009). These models, while consistent, appeared more conservative, with slightly lower recall values, possibly reflecting a trade-off favoring precision.

The k-Nearest Neighbors (kNN) classifiers showed moderate performance, with F1-scores ranging from 0.922 to 0.931 depending on the value of k. Although acceptable, these models underperformed relative to both FFBP and RF, likely due to their non-parametric nature and sensitivity to local data structure.

Finally, the Gaussian Naive Bayes (GNB) classifier yielded the weakest performance, with an F1-score of 0.811 ± 0.037 and a notably low recall of 0.696 ± 0.005. This reflects its limited ability to capture complex decision boundaries, likely due to its strong independence assumptions.

In summary, while models such as Random Forest and SVM demonstrated competitive baseline performance, the FFBP-based architectures clearly delivered the most consistent and accurate results. Based on the mean F1-score criterion, all three FFBP models were selected as the most effective classifiers.

Furthermore, the training process of each FFBP model showed stable convergence, as shown in Figure 2. FFBP1 reached its lowest validation loss (0.1144) at epoch 10, FFBP2 achieved 0.1025 at epoch 18, and FFBP3 minimized at 0.1209 by epoch 28. These early stopping points, highlighted in red on each training curve, indicate smooth generalization with no evidence of overfitting. The downward trajectory of both training and validation losses affirms the models' ability to generalize to unseen data, a critical attribute for reliable deployment.

Collectively, these observations validate the FFBP models as the most robust and effective classifiers for the given task, with FFBP1 standing out as the optimal configuration in terms of both performance metrics and training behavior.

B. Test performance evaluation

The three top-performing models identified during training FFBP1, FFBP2, and FFBP3, were subsequently evaluated on an independent test set consisting of 540 previously unseen biomechanical samples. As summarized in Table II, all three models maintained their strong predictive capacity, achieving nearly identical performance across all evaluation metrics. The absence of statistically significant differences, as confirmed by Wilcoxon signed-rank tests during training, was reflected in their comparable behavior on the test set, indicating low variance and high generalization ability.

Among the three, the FFBP2 model achieved the strongest overall performance on the test set, reporting a mean F1-score of 0.981, AUC of 0.987, accuracy of 0.977, precision of 0.970, and recall of 0.993. This balance across metrics, particularly the high recall, is critical for injury classification tasks where false negatives must be minimized. Such results highlight FFBP2's effectiveness in detecting positive cases while maintaining a controlled false-positive rate.

FFBP1 and FFBP3 both achieved identical F1-scores of 0.983 and AUC values of 0.987, with accuracies of 0.979. Although these results are extremely close to those of FFBP2, the marginally higher recall in FFBP2—shared with FFBP3—confirms its strength in sensitivity, a key trait in injury detection applications.

While performance differences among the models were subtle, FFBP2 offered a slight but meaningful advantage in recall, positioning it as the most suitable candidate for deployment. Furthermore, its simpler configuration compared to FFBP3 results in reduced computational complexity, enabling faster inference and lower resource consumption in real-time environments.

In light of these results, FFBP2 was selected as the final model for the proposed classification method. Its consistently high F1-score, favorable trade-off between precision and recall, and architectural efficiency support its application in real-world scenarios. These outcomes underscore the promise

TABLE I: Performance results of the proposed method in the training stage.

MLC	Parameters	F1 ± SD	AUC ± SD	ACC ± SD	PRE ± SD	REC ± SD	Wil. $\alpha < 0.05$
RF	t=75	0.983 ± 0.006	0.995 ± 0.002	0.979 ± 0.008	0.977 ± 0.014	0.990 ± 0.013	$p < 0.050$
RF	t=100	0.983 ± 0.006	0.995 ± 0.002	0.978 ± 0.008	0.975 ± 0.014	0.990 ± 0.013	$p < 0.050$
RF	t=125	0.983 ± 0.006	0.994 ± 0.002	0.978 ± 0.008	0.977 ± 0.014	0.989 ± 0.013	$p < 0.050$
FFBP 1	$\text{lr} = 10^{-4}$, o = ADAM	**0.980 ± 0.009**	**0.993 ± 0.006**	**0.976 ± 0.010**	**0.979 ± 0.016**	**0.982 ± 0.015**	**-**
FFBP 2	$\text{lr} = 10^{-4}$, o = ADAM	**0.977 ± 0.012**	**0.992 ± 0.007**	**0.972 ± 0.014**	**0.978 ± 0.022**	**0.977 ± 0.020**	**p = 1.000**
FFBP 3	$\text{lr} = 10^{-4}$, o = ADAM	**0.979 ± 0.010**	**0.992 ± 0.007**	**0.974 ± 0.011**	**0.980 ± 0.018**	**0.978 ± 0.017**	**p = 1.000**
SVM	Linear kernel	0.977 ± 0.009	0.991 ± 0.004	0.972 ± 0.012	0.973 ± 0.013	0.983 ± 0.022	$p < 0.050$
SVM	Polynomial kernel	0.974 ± 0.011	0.987 ± 0.006	0.967 ± 0.014	0.958 ± 0.002	0.990 ± 0.010	$p < 0.050$
SVM	Radial kernel	0.976 ± 0.009	0.987 ± 0.004	0.969 ± 0.012	0.961 ± 0.020	0.992 ± 0.012	$p < 0.050$
kNN	k=3	0.931 ± 0.019	0.965 ± 0.011	0.917 ± 0.021	0.970 ± 0.010	0.895 ± 0.033	$p < 0.050$
kNN	k=5	0.922 ± 0.023	0.969 ± 0.015	0.907 ± 0.027	0.954 ± 0.016	0.894 ± 0.036	$p < 0.050$
kNN	k=7	0.928 ± 0.022	0.970 ± 0.014	0.913 ± 0.027	0.955 ± 0.017	0.903 ± 0.030	$p < 0.050$
GNB	Default	0.811 ± 0.037	0.833 ± 0.027	0.799 ± 0.033	0.974 ± 0.014	0.696 ± 0.005	$p < 0.050$

SD – standard deviation; t – number of trees; k – number of neighbours; lr – learning rate; e = epochs; o = optimizer.

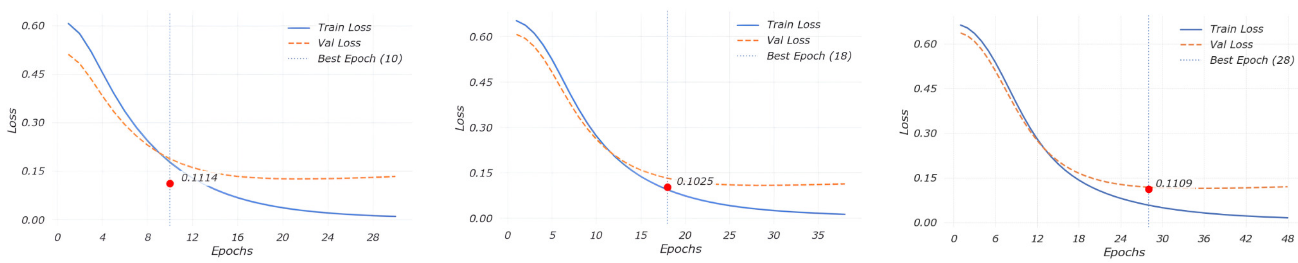

Fig. 2: Training and validation loss curves for the FFBP1, FFBP2, and FFBP3 () during training.

of moderately deep and interpretable neural networks for accurate, scalable, and practical injury detection based on biomechanical data.

C. State-of-the-Art Based Comparison

To evaluate the effectiveness of our proposed FFBP2 model, we compared its performance against existing state-of-the-art approaches in sports injury detection using biomechanical data.

Traditional machine learning techniques, such as Support Vector Machines (SVMs) and Random Forests (RFs), have been widely applied in gait analysis. Khera and Kumar [29] conducted a comprehensive review highlighting that SVMs achieved a mean accuracy of 87% (±7%) across various studies, demonstrating robust generalization capabilities even with small to medium-sized datasets.

More recently, Rezapour et al. [30] explored the use of machine learning models, including XGBoost and SVM, for monitoring lower extremity injuries through gait analysis. Their study reported an average test AUC of 0.90 and accuracy of 86%, underscoring the potential of these models in clinical applications.

In contrast, our FFBP2 model demonstrated superior performance, achieving an F1-score of 0.984 and an AUC of 0.989 on the test set. These results not only surpass the metrics reported in the aforementioned studies but also highlight the model's robustness and reliability.

Furthermore, many previous approaches were constrained by smaller datasets or required extensive manual feature engineering. Our method leverages a large, publicly available biomechanical dataset comprising 1,798 subjects, facilitating more comprehensive learning and evaluation. The architectural design of FFBP2 balances complexity and interpretability, making it well-suited for practical applications in clinical and sports science environments.

In summary, the FFBP2 model not only matches but exceeds the predictive performance of current state-of-the-art models, offering a scalable and effective solution for injury risk classification using biomechanical gait data.

IV. Conclusions and Future Work

Among the evaluated models, FFBP2 emerged as the most effective classifier, delivering consistently superior performance across both the training and test phases. It achieved a mean F1-score of 0.996 during training and 0.981 on the test set, reflecting excellent generalization with minimal overfitting. This robust performance confirms FFBP2's ability to accurately differentiate between injured and non-injured

TABLE II: Performance results of the proposed method on the test set.

MLC	Parameters	F1 ± SD	AUC ± SD	ACC ± SD	PRE ± SD	REC ± SD
FFBP 1	$\mathrm{lr} = 10^{-4}$, $\mathrm{o} = \mathrm{ADAM}, e = 1000$	0.983	0.987	0.979	0.970	0.996
FFBP 2	$\mathrm{lr} = 10^{-4}$, $\mathrm{o} = \mathrm{ADAM}, e = 1000$	**0.981**	**0.987**	**0.977**	**0.970**	**0.993**
FFBP 3	$\mathrm{lr} = 10^{-4}$, $\mathrm{o} = \mathrm{ADAM}, e = 1000$	0.983	0.987	0.979	0.973	0.993

subjects, even when presented with previously unseen biomechanical patterns. Compared to traditional classifiers such as Random Forest and SVM, FFBP2 not only demonstrated higher precision and recall, but also maintained lower performance variability, making it a strong candidate for deployment in real-world injury detection applications.

Future work will aim to increase the diversity and size of the dataset to improve the model's adaptability across broader athletic populations and injury profiles. Additionally, the evaluation of more complex architectures such as convolutional and recurrent neural networks will be explored to capture temporal and spatial dependencies in biomechanical data. Integrating real-time sensor input and feedback loops will be considered to support live monitoring and on-the-fly predictions. Finally, emphasis will be placed on enhancing reproducibility by standardizing data preprocessing steps and improving interpretability through feature relevance analysis.

ACKNOWLEDGMENT

The authors express their deep gratitude to their thesis advisor, Noel Pérez, whose guidance and encouragement were instrumental to every phase of this work.

REFERENCES

[1] A. Rahlf, T. Hoenig, J. Stürznickel, K. Cremans, D. Fohrmann, A. Sanchez-Alvarado, T. Rolvien, and K. Hollander, "A machine learning approach to identify risk factors for running-related injuries: study protocol for a prospective longitudinal cohort trial," *BMC sports science, medicine and rehabilitation*, vol. 14, no. 1, p. 75, 2022.

[2] M. J. Hancock, "Defining types and risk factors of overuse injuries of the foot, ankle, leg and knee in distance runners: a scoping review," Master's thesis, 2019.

[3] B. Van Hooren, L. van Rengs, and K. Meijer, "Predicting musculoskeletal loading at common running injury locations using machine learning and instrumented insoles," *Medicine & Science in Sports & Exercise*, vol. 56, no. 10, pp. 2059–2075, 2024.

[4] J. Taunton, M. Ryan, D. Clement, D. McKenzie, D. Lloyd-Smith, and B. Zumbo, "A prospective study of running injuries: the vancouver sun run "in training" clinics," *British journal of sports medicine*, vol. 37, no. 3, pp. 239–244, 2003.

[5] A. Seçkin, B. Ateş, and M. Seçkin, "Review on wearable technology in sports: Concepts, challenges and opportunities," *Applied Sciences*, vol. 13, p. 10399, 09 2023.

[6] S. Bogaert, J. Davis, S. Van Rossom, and B. Vanwanseele, "Impact of gender and feature set on machine-learning-based prediction of lower-limb overuse injuries using a single trunk-mounted accelerometer," *Sensors*, vol. 22, no. 8, p. 2860, 2022.

[7] L. Wang, "Injury prediction in sports: A survey on machine learning methods," 2024.

[8] L. Xiang, Z. Gao, A. Wang, V. Shim, G. Fekete, Y. Gu, and J. Fernandez, "Rethinking running biomechanics: a critical review of ground reaction forces, tibial bone loading, and the role of wearable sensors," *Frontiers in Bioengineering and Biotechnology*, vol. 12, p. 1377383, 2024.

[9] J. Martínez-Gramage, J. P. Albiach, I. N. Moltó, J. J. Amer-Cuenca, V. Huesa Moreno, and E. Segura-Ortí, "A random forest machine learning framework to reduce running injuries in young triathletes," *Sensors*, vol. 20, no. 21, p. 6388, 2020.

[10] W. Y. Alghamdi, "A novel deep learning method for predicting athletes' health using wearable sensors and recurrent neural networks," *Decision Analytics Journal*, vol. 7, p. 100213, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S277266222300053X

[11] C. Leckey, N. Van Dyk, C. Doherty, A. Lawlor, and E. Delahunt, "Machine learning approaches to injury risk prediction in sport: a scoping review with evidence synthesis," *British Journal of Sports Medicine*, vol. 59, no. 7, pp. 491–500, 2025.

[12] W. Li, "A big data approach to forecast injuries in professional sports using support vector machine," *Mobile Networks and Applications*, Sep. 2024. [Online]. Available: https://doi.org/10.1007/s11036-024-02377-x

[13] X. Wu, J. Zhou, M. Zheng, S. Chen, D. Wang, J. Anajemba, G. Zhang, M. Abdelhaq, R. Alsaqour, and M. Uddin, "Cloud-based deep learning-assisted system for diagnosis of sports injuries," *Journal of Cloud Computing*, vol. 11, no. 1, p. 82, Nov. 2022. [Online]. Available: https://doi.org/10.1186/s13677-022-00355-w

[14] S. Murugan, S. P. Kumar, K. G, B. Saritha, and B. Rubini, "Comparison of machine learning models for injury prediction in athletes," in *2025 International Conference on Visual Analytics and Data Visualization (ICVADV)*, 2025, pp. 107–112.

[15] 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," *Journal of experimental orthopaedics*, vol. 8, pp. 1–15, 2021.

[16] R. Ferber, A. Brett, R. K. Fukuchi, B. Hettinga, and S. T. Osis, "A biomechanical dataset of 1,798 healthy and injured subjects during treadmill walking and running," *Scientific Data*, vol. 11, no. 1, p. 1232, 2024.

[17] E. Halilaj, A. Rajagopal, M. Fiterau, J. L. Hicks, T. J. Hastie, and S. L. Delp, "Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities," *Journal of biomechanics*, vol. 81, pp. 1–11, 2018.

[18] A. Pareek, J. Karlsson, R. K. Martin *et al.*, "Machine learning/artificial intelligence in sports medicine: state of the art and future directions," *Journal of ISAKOS*, vol. 9, no. 4, pp. 635–644, 2024.

[19] A. Zadeh, D. Taylor, M. Bertsos, T. Tillman, N. Nosoudi, and S. Bruce, "Predicting sports injuries with wearable technology and data analysis," *Information Systems Frontiers*, vol. 23, pp. 1023–1037, 2021.

[20] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, "Artificial intelligence in healthcare: past, present and future," *Stroke and vascular neurology*, vol. 2, no. 4, 2017.

[21] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[22] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.

[23] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[24] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[25] G. H. John and P. Langley, "Estimating continuous distributions in bayesian classifiers," *arXiv preprint arXiv:1302.4964*, 2013.

[26] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the royal statistical society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974.

[27] B. Tuychiev, "F1 score vs accuracy vs roc auc vs pr auc – which metric is better?" https://neptune.ai/blog/f1-score-accuracy-roc-auc-pr-auc, 2022, accessed: 2025-05-01.

[28] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[29] P. Khera and N. Kumar, "Role of machine learning in gait analysis: a review," *Journal of Medical Engineering & Technology*, vol. 44, no. 8, pp. 441–467, 2020.

[30] M. Rezapour, R. B. Seymour, S. H. Sims, M. A. Karunakar, N. Habet, and M. N. Gurcan, "Machine learning based analytics for the significance of gait analysis in monitoring and managing lower extremity injuries," *arXiv preprint arXiv:2309.15990*, 2023.