SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2024, 21(3), 530.
https://doi.org/10.62617/mcb530

Article

Innovative machine learning approach for analysing biomechanical factors in running-related injuries

Rui Han[1], Feng Qi[2,3], Hong Wang[2,3], Mingnong Yi[1,*]

[1] Key Laboratory of Sports Engineering of General Administration of Sport of China, Wuhan Sports University, Wuhan 430079, China
[2] Engineering Research Center of Sports Health Intelligent Equipment of Hubei Province, Wuhan Sports University, Wuhan 430079, China
[3] Research Center of Sports Equipment Engineering Technology of Hubei Province, Wuhan Sports University, Wuhan 430079, China
*** Corresponding author:** Mingnong Yi, mingnongyi@outlook.com

CITATION

Han R, Qi F, Wang H, Yi M. Innovative machine learning approach for analysing biomechanical factors in running-related injuries. Molecular & Cellular Biomechanics. 2024; 21(3): 530. https://doi.org/10.62617/mcb530

ARTICLE INFO

Received: 13 October 2024
Accepted: 21 October 2024
Available online: 15 November 2024

Abstract: Running-related injuries are a significant concern for recreational and competitive athletes, often resulting from complex biomechanical interactions. Traditional injury assessment methods are limited in their ability to capture dynamic, real-time data, necessitating the need for more advanced predictive tools. This study proposes an innovative machine-learning approach to predict running-related injuries by analyzing biomechanical data collected from 84 active runners. The data included joint angles, ground reaction forces, stride length, muscle activation, and foot pressure, captured through wearable sensors during laboratory-controlled and outdoor running sessions. An ensemble model combining Gradient-Boosted Decision Trees (GBDT), Long Short-Term Memory (LSTM) networks, and Support Vector Machines (SVM) was developed to predict injury risk. The results indicate that ground reaction force, foot pressure, and stride length were the most significant predictors of injury. The proposed ensemble model achieved an accuracy of 88.37%, outperforming individual models such as GBDT (83.74%) and LSTM (81.29%). The findings suggest that integrating machine learning techniques with biomechanical analysis can significantly enhance the prediction and prevention of running-related injuries. This research offers valuable insights into developing personalized injury prevention strategies, potentially reducing injury occurrence among athletes.

Keywords: biomechanical data; joint angles; ground reaction force; foot pressure; biomechanical analysis; running-related injuries

1. Introduction

Running is a globally popular form of physical activity due to its accessibility, cardiovascular benefits, and ability to promote physical and mental well-being [1]. However, it is also associated with a high prevalence of musculoskeletal injuries, particularly among long-distance runners and competitive athletes [2,3]. Studies suggest that up to 79% of runners will experience some form of injury each year, with common issues including stress fractures, plantar fasciitis, and knee injuries [4,5]. These injuries often arise from biomechanical imbalances and repetitive stress, making early detection and intervention crucial for minimizing long-term damage and improving athletic performance [6,7].

The biomechanics of running—comprising variables such as stride length, ground reaction forces, joint angles, and muscle activation—play a significant role in injury risk [8,9]. Runners with improper form or inefficient mechanics are at a higher risk of injury, as these factors can increase the forces exerted on the musculoskeletal system during repetitive motions [10]. Over the years, traditional methods of assessing running mechanics have relied on laboratory-based analyses, video assessments, and clinical evaluations [11–14]. While these techniques provide valuable insights, they are often limited by their dependence on static or retrospective data and unsuitable for real-time injury prevention [15,16].

Recent technological advancements in wearable sensors, real-time data capture, and machine learning offer new opportunities to address these limitations [17]. Wearable sensors can continuously monitor key biomechanical parameters during running, enabling the collection of vast amounts of data across diverse environments [18,19]. This real-time data can be analyzed using machine learning techniques to uncover complex patterns and relationships that may not be evident through traditional analysis methods [20,21]. Machine learning models, particularly those designed to handle large datasets and time-series data, can provide predictive insights into injury risk, offering a more dynamic and personalized approach to injury prevention [22].

Despite these advancements, existing models for predicting running-related injuries still face challenges in accurately identifying the key biomechanical factors contributing to injury [23]. Many studies use linear or rule-based models, which may oversimplify the complex interactions between variables such as ground reaction forces, joint angles, and muscle activation timing [24]. Moreover, current predictive models often struggle to adapt to the high variability in individual running patterns, which can fluctuate based on experience level, running environment, and physical conditioning [25]. As a result, there remains a need for more robust predictive tools that can account for the nonlinear relationships and temporal dependencies inherent in biomechanical data [26,27].

This study seeks to address these gaps by proposing an innovative machine-learning approach that integrates multiple algorithms to enhance the prediction of running-related injuries. By combining Gradient-Boosted Decision Trees (GBDT), Long Short-Term Memory (LSTM) networks, and Support Vector Machines (SVM) in an ensemble model, this research aims to leverage the strengths of each method to analyze complex biomechanical data. GBDT is particularly effective at identifying nonlinear relationships and ranking feature importance, while LSTM networks are well-suited for handling sequential, time-series data. SVMs, on the other hand, offer strong classification performance for high-dimensional datasets. Together, these models form a comprehensive framework capable of processing structured biomechanical variables and dynamic running data.

The study was conducted on a cohort of 84 active runners from various urban regions in China, recruited to represent a range of running experience levels, from recreational to competitive athletes. Data collection occurred over six months, ensuring that seasonal variations in running patterns were captured. Biomechanical data, including joint angles, foot strike patterns, vertical ground reaction forces, and muscle activation timings, were collected using wearable sensors during laboratory-controlled treadmill runs and outdoor track sessions. These data were then analyzed through the proposed ensemble machine-learning model to predict the likelihood of running-related injuries.

The objectives of this study are threefold:

- To identify the key biomechanical factors contributing to running-related injuries across different experience levels.

- To develop a predictive model capable of integrating multiple sources of biomechanical data and accurately assessing injury risk in real time.
- To evaluate the performance of the ensemble model by comparing its accuracy, precision, recall, and F1 score against other machine learning models such as GBDT, LSTM, SVM, Random Forest, and Decision Trees.

By addressing these objectives, this study aims to advance the understanding of how biomechanical variables influence injury risk and to provide a more effective tool for injury prevention in runners. The findings can benefit athletes, coaches, sports medicine professionals, and physiotherapists by offering data-driven insights that can inform personalized training programs and rehabilitation strategies. Following the Introduction section, the paper is structured as follows: Methodology (2.0) outlines the study design, participant selection, and biomechanical data collection. The Proposed Machine Learning Approach (2.5) details the development of an ensemble model using GBDT, LSTM, and SVM for injury prediction. Result Analysis (3.0) covers statistical tests, including T-tests, ANOVA, and regression, alongside machine learning results like feature importance and model accuracy. Finally, the Conclusion (4.0) summarizes key findings, highlights influential biomechanical factors, and suggests future applications for injury prevention.

The article is presented as follows: Section 2 presents the theoretical framework, Section 3 presents the methodology, Section 4 presents the analysis, and Section 5 concludes the paper.

2. Methodology

2.1. Study design

The study was conducted in China, focusing on 84 active runners recruited from various urban regions with a robust running culture. The study spanned six months from 1 March 2023, to 31 August 2023, ensuring seasonal variations in running patterns were captured. Participants were selected using a stratified sampling to ensure a balanced representation across different running experience levels, from recreational to competitive athletes. The inclusion criteria required participants to have maintained a minimum running distance of 20 km per week for at least six months before the study, with no existing injuries or musculoskeletal disorders that could compromise their running mechanics.

The study cohort included a balanced gender distribution, allowing for analysis of biomechanical differences between male and female runners. Age group subdivisions were also incorporated, capturing a wide range of biomechanical variations across younger and older runners. Participants underwent tests in controlled laboratory environments and outdoor tracks to reflect real-world running conditions. In the laboratory sessions, participants were fitted with wearable sensors and ran on a treadmill at various speeds, while outdoor sessions involved track running to simulate different terrains and intensities.

The biomechanical data collected during this period included joint angles, foot strike patterns, vertical ground reaction forces, and muscle activation throughout the running cycle. Additionally, detailed questionnaires were administered to gather contextual data, including participants' running history, injury records, and training habits. This data collection framework ensured a rich dataset for the machine learning model to analyze the complex factors contributing to running-related injuries. Ethical approval was obtained from the relevant research ethics committees in China, and all participants provided informed consent before participating. This carefully designed study allowed for comprehensive data gathering and set a solid foundation for the subsequent machine learning analysis.

2.2. Participants

The study involved a total of 84 participants, all of whom were active runners recruited from various regions across China with running solid communities. The participants were selected through a stratified sampling to ensure a diverse representation of running experience, age, and gender. Of the 84 participants, 46 were male and 38 were female, aged 18 to 50. To ensure comprehensive coverage, participants were further grouped into three age categories: 18–25 years (28 participants), 26–35 years (31 participants), and 36–50 years (25 participants). This age distribution allowed for a thorough investigation of potential biomechanical differences across younger, middle-aged, and older runners.

In terms of running experience, participants were classified into three categories: recreational runners (32 participants), who ran less than 30 kilometers per week; intermediate runners (34 participants), who ran between 30 and 50 kilometers per week; and advanced runners (18 participants), who regularly exceeded 50 kilometers per week. This categorization ensured the study could analyze biomechanical variations between runners of different training intensities and experience levels. All participants ran consistently for at least six months before the study and were free from any significant musculoskeletal injuries or conditions that could influence their natural running biomechanics.

The study maintained a near-equal distribution of male and female participants to account for gender-specific variations in biomechanical patterns. This balance was essential to ensure that the machine learning models would capture any differences in running mechanics between men and women, such as variations in joint angles, muscle activation patterns, and injury risks. Additionally, the study recorded participants' body mass index (BMI) and height to provide further context for analyzing biomechanical variables. The average BMI for participants was 22.8, with heights ranging from 155 cm to 190 cm.

2.3. Assessment and apparatus

The study involved a multi-phase assessment process to comprehensively assess the biomechanical factors contributing to running-related injuries (**Table 1**). This included a Baseline Assessment, Track Assessment, and Laboratory Assessment, with specific tools and methods to evaluate each aspect of the participants' physical and biomechanical performance.

2.3.1. Baseline assessment

The baseline assessment began with a detailed evaluation of participants' physical and musculoskeletal health. A musculoskeletal assessment measured joint mobility, muscle strength, and functional stability. The flexibility of key joints such as the ankles, knees, and hips were measured using a goniometer, while muscle strength was quantified through hand-held dynamometers, which provided precise measurements of isometric strength in the quadriceps, hamstrings, and calf muscles. These tests were essential for identifying any muscular imbalances, such as quadriceps dominance or weakened calf muscles, which could predispose participants to injury.

Participants also completed a detailed musculoskeletal health questionnaire, which collected data on their previous injuries, frequency of muscle pain, and general joint stiffness. This questionnaire provided qualitative insights into the participants' injury history and pain experiences, offering context for the biomechanical data collected later. Anthropometric measurements such as height, weight, and body mass index (BMI) were also recorded using a digital stadiometer and digital scale to help contextualize the biomechanical data concerning each participant's physical attributes.

2.3.2. Track assessment

Participants ran on a standard 400 m outdoor track in the track assessment to simulate real-world running conditions. The participants had wearable sensors attached to the waist, thighs, and ankles. These sensors measured stride length, cadence, joint angles, and foot strike patterns as participants ran at varying speeds. This setup allowed for real-time monitoring of how musculoskeletal and biomechanical variables changed under different conditions.

Also, pressure-sensitive mats were strategically placed along the track to measure pressure distribution across the foot during the running cycle. This data provided insights into the participants' foot strike patterns, particularly the loading patterns during foot strike and push-off phases, which are critical for understanding the potential for injuries such as plantar fasciitis or stress fractures.

High-speed video cameras were positioned around the track to capture the participants' movements from multiple angles, allowing for a post-run video analysis of their running form. This footage helped identify biomechanical deviations or misalignments that could contribute to injury, such as knee valgus or overpronation.

2.3.3. Laboratory assessment

The laboratory assessment provided a controlled environment for precise and detailed biomechanical analysis. Participants ran on a treadmill with embedded force plates, measuring ground reaction forces during each foot strike. The force plates captured data on vertical, anterior-posterior, and medial-lateral forces, allowing for an in-depth analysis of how the body absorbed impact and distributed forces during running.

A 3D motion capture system was used to track joint movements and limb coordination in real-time. Reflective markers were placed on key anatomical landmarks, such as the hips, knees, and ankles. The system recorded participants' joint angles, rotations, and limb movements with millimeter accuracy. This data assessed the participants' running biomechanics and identified any abnormalities, such as excessive knee rotation or hip drop.

In addition to the force plates and motion capture system, electromyography (EMG) sensors were used to monitor muscle activation patterns in the quadriceps, hamstrings, and calf muscles. These sensors provided real-time data on muscle function, identifying abnormal activation patterns that could indicate muscle fatigue or imbalances. Plantar pressure sensors were also used during the treadmill runs to provide detailed information on foot pressure distribution during the running cycle, helping to identify areas prone to injury due to excessive pressure or impact.

Table 1. Apparatus used for each phase of the study.

Assessment Type	Specific Evaluation	Apparatus Used	Data Collected
Baseline Assessment	Anthropometric Measurements	Digital Stadiometer, Digital Scale	Height, Weight, Body Mass Index (BMI)
	Musculoskeletal Health Assessment	Goniometer, Dynamometers	Joint Mobility, Muscle Strength (Quadriceps, Hamstrings, Calf)
	Musculoskeletal Health Questionnaire	Standardized Questionnaire	Injury History, Pain Frequency, Joint Stiffness, Training Habits
	Flexibility and Strength Testing	Sit-and-Reach Test, Single-Leg Squats	Lower Body Flexibility, Strength Levels
Track Assessment	Gait Analysis	Wearable Sensors (Thighs, Waist, Ankles)	Stride Length, Cadence, Joint Angles, Foot Strike Patterns
	Pressure Distribution	Pressure-Sensitive Mats	Foot Pressure Distribution During Running, Hotspots for Injury Risk (e.g., excessive pressure areas)
	Real-Time Motion Capture	High-Speed Video Cameras	Visual Data of Running Movements, Stride Mechanics
Laboratory Assessment	Ground Reaction Force (GRF) Measurement	Treadmill with Embedded Force Plates	Vertical, Anterior-Posterior, and Medial-Lateral Ground Reaction Forces
	Joint Movement and Coordination	3D Motion Capture System (Vicon Nexus), Reflective Markers	Joint Angles, Rotations, Limb Coordination
	Muscle Activation Patterns	Electromyography (EMG) Sensors	Muscle Activation in Quadriceps, Hamstrings, and Calf Muscles
	Foot Pressure Analysis	Plantar Pressure Sensors	Foot Pressure Distribution During Foot Strike and Push-Off Phases

2.4. Experimental design

The experiment was designed to explore the relationship between biomechanical factors and running-related injuries using an innovative machine-learning model tailored explicitly for this study. The primary goal was to gather detailed biomechanical data across various running conditions, preprocess this data, and apply the machine learning approach to predict injury risks and patterns with high accuracy. The experimental framework included several phases, where data collection, preprocessing, and model training were systematically integrated to ensure robust, actionable insights.

2.4.1. Participant preparation and setup

Before the experiment began, participants were briefed on the protocol, and informed consent was obtained. The experiment was conducted and controlled across outdoor track environments and laboratory settings. Each participant was fitted with various sensors for real-time data collection, including motion capture markers, muscle activity sensors, and force measurement devices. This preparation ensured that the key biomechanical variables were captured effectively.

2.4.2. Experimental procedure

The experiment was conducted in three phases—baseline, track, and laboratory—each designed to capture a specific set of biomechanical data. The experimental sessions took place over multiple days to prevent participant fatigue and ensure high data quality.

1) Baseline Phase: Participants underwent a full musculoskeletal assessment in the initial phase, including flexibility and strength measurements. This baseline data was essential for understanding their general physical health and how it related to their running mechanics. A musculoskeletal health questionnaire was also administered to record participants' injury histories and training habits.

2) Track Phase: Participants were asked to run on an outdoor track at three different speeds: slow (jogging pace), moderate (training pace), and fast (race pace). Each participant completed a series of 800 m runs at varying speeds. Wearable sensors captured real-time gait data, including joint angles, footstrike patterns, and stride length. Pressure-sensitive mats were placed along the track to measure foot pressure distribution at key points during the running cycle.

To ensure the track conditions did not affect the results, the experiment was conducted under consistent weather and surface conditions and monitored closely for any deviations. High-speed cameras were positioned around the track to capture detailed video footage of each participant's gait, which was later analyzed for biomechanical deviations.

3) Laboratory Phase: Participants ran on a treadmill equipped with embedded force plates to precisely measure the ground reaction forces generated during running. The treadmill runs were conducted at the same three speeds used in the track phase. A 3D motion capture system also recorded joint movements, while electromyography (EMG) sensors measured muscle activation patterns. This phase provided a controlled environment for capturing data on how forces and joint movements change with speed, fatigue, and physical exertion.

The treadmill was calibrated to ensure all force measurements were accurate and reproducible across participants. This phase allowed for a detailed breakdown of each participant's biomechanics under consistent running conditions, providing high-resolution data for further analysis.

2.4.3. Machine learning model application

The experimental data collected across all phases were preprocessed, normalized, and used for feature extraction, targeting key biomechanical variables such as joint angles, ground reaction forces, muscle activation timing, and foot pressure distribution. The machine learning model for this study was based on an ensemble learning approach, optimized to handle the complexity of biomechanical data and accurately predict injury risk.

The ensemble model combined the strengths of:

- Gradient-Boosted Decision Trees (GBDT): GBDT was used to identify nonlinear relationships between biomechanical variables such as stride length, ground reaction forces, and injury patterns. This model excels in handling structured data and capturing complex variable interactions.

- Long Short-Term Memory (LSTM) Neural Networks: Since running data is sequential, LSTM networks were employed to process the time-series data from the motion capture system and EMG sensors. This allowed the model to learn how variations in running mechanics over time contribute to injury risk, providing a deeper temporal understanding of biomechanics.
- Support Vector Machines (SVM): SVMs were used as part of the ensemble to classify injury risk based on the multidimensional feature space of biomechanical data. SVM is effective for high-dimensional data, helping to distinguish between normal and abnormal movement patterns associated with injuries.

The ensemble model was designed to synergize these methods, allowing it to leverage their complementary strengths. By integrating decision trees for feature importance, LSTM for sequential data processing, and SVM for classification, the model could predict the likelihood of injury and identify the key biomechanical factors contributing to that risk.

2.4.4. Experiment control and monitoring

All participants followed the same running protocol across the track and laboratory phases to maintain consistency. Environmental factors such as weather and surface conditions were controlled as much as possible during the track runs. Temperature, humidity, and lighting were kept constant in the laboratory phase. Participants were given adequate rest periods between sessions to prevent fatigue and ensure that the data collected was not compromised by physical exhaustion. Throughout the experiment, data were continuously monitored for anomalies, such as sensor malfunctions or irregularities in running patterns. These checks ensured that the data collected were of the highest quality, supporting the machine learning model's predictive accuracy.

2.4.5. Post-Experiment data integration and analysis

After the experimental sessions From **Figure 1**, all phases' data were integrated into a single dataset. This dataset was processed to ensure consistency across the different collection methods (track vs. laboratory) and prepare for machine learning model training. The ensemble learning model was applied to the data to identify key biomechanical patterns associated with injury risks. The results were then validated using cross-validation techniques, ensuring the model's predictions were reliable and generalizable.

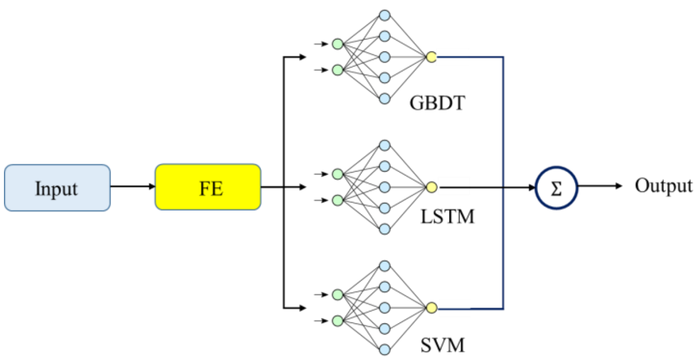

Figure 1. Proposed ML model.

2.5. Proposed machine learning approach

In this study, we propose an innovative machine-learning approach to predict running-related injuries based on biomechanical factors. The approach integrates multiple machine learning algorithms into an ensemble model, leveraging the strengths of each method to process complex, high-dimensional data while focusing on the unique temporal dynamics of running biomechanics. The proposed model combines Gradient-Boosted Decision Trees (GBDT) for feature importance analysis, Long Short-Term Memory (LSTM) networks for time-series data processing, and Support Vector Machines (SVM) for classification. This ensemble architecture allows the model to effectively handle structured biomechanical and sequential time-series data collected from sensors during the study.

1) Feature Extraction (FE): The first step in our machine-learning approach involves extracting relevant features from the biomechanical data. Key features include joint angles (θ_j), ground reaction forces (F_g), muscle activation timing (t_m), stride length (S_l), cadence (C_d), and foot pressure distribution (P_f). These features are crucial for capturing the biomechanical variations among the participants. The data preprocessing pipeline includes the normalization of these variables using the min-max scaling method:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

where x is the original feature value, $x_{\min}$ is the minimum value of the feature, and $x_{\max}$ is the maximum value. This scaling ensures that all features have the same range, preventing any one feature from dominating the model due to its scale.

2) Gradient-Boosted Decision Trees (GBDT): GBDT (**Figure 2**) is employed in the model to determine the importance of each biomechanical feature in predicting injury risk. The algorithm iteratively builds decision trees, with each tree attempting to correct the errors of the previous ones. The GBDT loss function is given by:

$$L(\theta) = \sum_{i=1}^{N} \left(y_i - f(x_i;\theta)\right)^2 \tag{2}$$

where y_i represents the actual injury outcome, $f(x_i;\theta)$ is the predicted value from the model, and θ are the model parameters. By minimizing this loss function, the model iteratively improves its predictions, placing greater weight on biomechanical features that contribute more to injury risks, such as high-ground reaction forces or abnormal joint angles. The output from GBDT is used to rank the importance of biomechanical factors, helping identify the key variables driving injury risk, which are then passed into the next layer of the model.

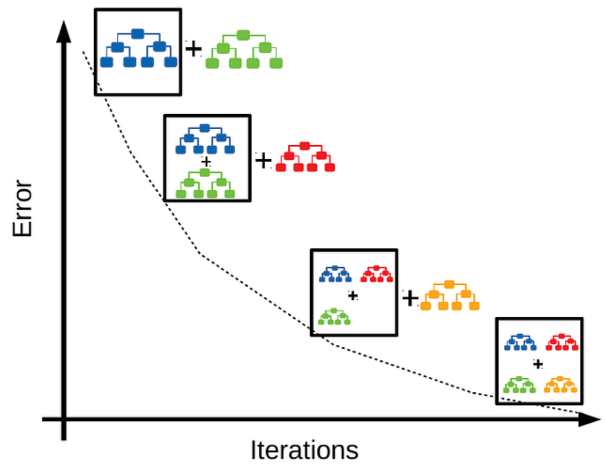

Figure 2. GBDT structure.

3) Long Short-Term Memory (LSTM) Networks: Running biomechanics involves dynamic, sequential data that change over time, which makes LSTM networks (**Figure 3**) ideal for modeling these temporal dependencies. LSTM networks are designed to capture long-term dependencies and patterns from the sequential biomechanical data, such as changes in muscle activation and joint movement over multiple strides. The LSTM network processes the time-series data as follows:

$$h_t = \sigma(W_h \times [h_{t-1}, x_t] + b_h) \tag{3}$$

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \tag{4}$$

$$o_t = \sigma(W_o \times h_t + b_o) \tag{5}$$

where h_t is the hidden state at time step t, x_t is the input at the time step t_t and W_h and b_h are the weight and bias terms. The cell state c_t is updated based on the forget gate f_t, input gate i_t, and candidate cell state $\tilde{c}_t$, enabling the network to retain and update relevant biomechanical information over time. By capturing the sequential nature of running data, the LSTM network identifies how variations in biomechanical patterns across multiple strides contribute to injury risk. The output from the LSTM layer feeds into the final classification model.

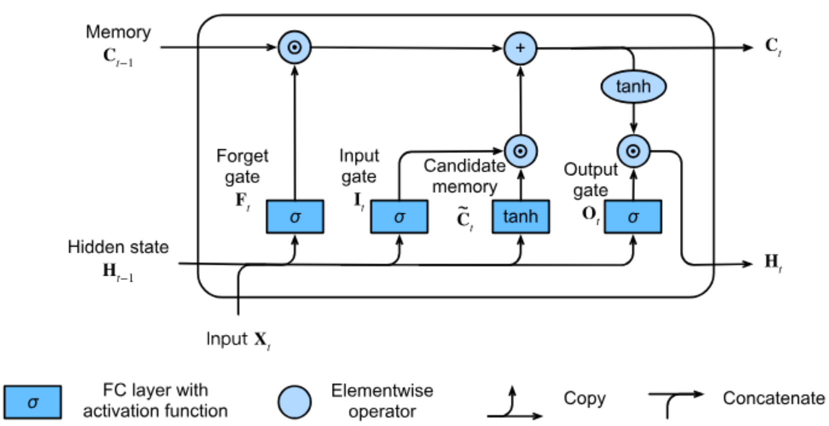

Figure 3. LSTM architecture.

4) Support Vector Machines (SVM): The SVM component (**Figure 4**) of the ensemble is used to classify participants into injury-risk categories based on the biomechanical data processed by GBDT and LSTM. The SVM algorithm finds the hyperplane that best separates the data into two classes: those at risk of injury and those not at risk. The decision boundary is defined by:

$$f(x) = w^T x + b \tag{6}$$

where w is the weight vector, x is the input feature vector, and b is the bias term. The SVM seeks to maximize the margin between the two classes, ensuring that the classification is robust to variations in the biomechanical data. The optimal hyperplane is found by minimizing the following objective function:

$$\min \frac{1}{2} \| w \|^2 \text{ subject to } y_i(w^T x_i + b) \geq 1 \ \forall i \tag{7}$$

where y_i is the class label for the i-th participant (injury or no injury). The SVM is beneficial for high-dimensional data, such as the biomechanical dataset in this study, and provides a highly accurate classification of injury risk based on the combined features.

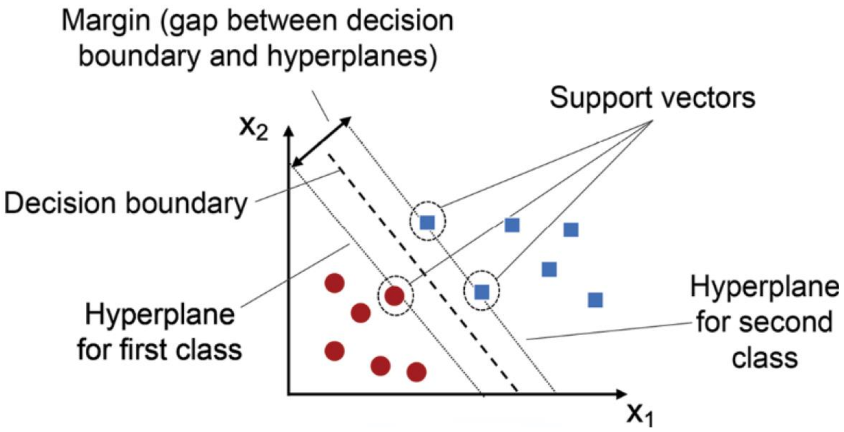

Figure 4. SVM architecture.

5) Ensemble Model: The final prediction of injury risk is made by aggregating the outputs from the GBDT, LSTM, and SVM models. Each model contributes its prediction, and a weighted voting mechanism determines the outcome. The overall prediction $\hat{y}$ is given by:

$$\hat{y} = \alpha_1 \cdot \hat{y}_{\text{GBDT}} + \alpha_2 \cdot \hat{y}_{\text{LSTM}} + \alpha_3 \cdot \hat{y}_{\text{SVM}} \tag{8}$$

where α_1, α_2, and α_3 are the weights assigned to the predictions from the GBDT, LSTM, and SVM models, respectively, and $\hat{y}$ is the final predicted injury risk. The weights are optimized based on cross-validation results to ensure the best performance of the ensemble model.

The performance of the proposed machine learning approach is evaluated using standard metrics such as accuracy, precision, recall, and the F1-score. Additionally, the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve is used to assess the model's capability in distinguishing between injury-prone and injury-free participants. Cross-validation minimizes overfitting and ensures the model generalizes well to new, unseen data.

3. Result analysis

3.1. Statistical analysis

Table 2. Descriptive statistics.

Variable	Mean	Standard Deviation	Minimum	Maximum
Age (years)	30.72	8.91	18.39	49.73
BMI (kg/m^2)	22.83	2.34	18.67	27.42
Stride Length (m)	1.36	0.22	1.12	1.77
Ground Reaction Force (N)	245.67	35.98	197.63	302.45
Joint Angle (Knee, degrees)	31.47	5.72	24.15	39.84
Cadence (steps/min)	168.39	10.94	150.28	183.61
Foot Pressure (kPa)	87.63	12.89	70.27	109.47
Muscle Activation Time (ms)	236.84	48.12	177.53	312.91
Ankle Dorsiflexion (degrees)	12.63	3.57	8.92	18.71
Hip Flexion (degrees)	42.16	6.28	33.47	53.84
Vertical Displacement (cm)	7.42	1.81	4.53	10.86
Knee Extension (degrees)	26.89	4.91	20.37	33.56

The study sample's descriptive statistics (**Table 2**) provide an overview of key demographic and biomechanical variables. The average age of the participants was 30.72 years, with a standard deviation of 8.91, indicating a moderately wide age range from 18.39 to 49.73 years. The average body mass index (BMI) was 22.83 kg/m^2, within a range of 18.67 to 27.42, reflecting a generally healthy population. Regarding running mechanics, the average stride length was 1.36 m with a standard deviation of 0.22 m, showing variability among participants, with stride lengths ranging from 1.12 to 1.77 m. Ground reaction forces averaged 245.67 N with a standard deviation of 35.98 N, spanning from 197.63 N to 302.45 N, suggesting differences in how participants interacted with the ground during running.

The knee joint angle averaged 31.47 degrees with some variability (standard deviation of 5.72 degrees), ranging from 24.15 to 39.84 degrees. Cadence, or the number of steps per minute, averaged 168.39 steps/min, with a standard deviation of 10.94, ranging between 150.28 and 183.61 steps/min. Foot pressure during the running cycle averaged 87.63 kPa, ranging from 70.27 to 109.47 kPa. Muscle activation time was recorded at an average of 236.84 ms, with considerable variation (standard deviation of 48.12 ms), spanning from 177.53 to 312.91 ms. Ankle dorsiflexion, an essential measure of ankle mobility, averaged 12.63 degrees, with participants ranging between 8.92 and 18.71 degrees. Hip flexion averaged 42.16 degrees with a range of 33.47 to 53.84 degrees, while vertical displacement during running averaged 7.42 cm, ranging from 4.53 to 10.86 cm. Finally, knee extension averaged 26.89 degrees, ranging from 20.37 to 33.56 degrees.

Table 3. Independent *t*-tests: Comparison of biomechanical variables between male and female participants.

Variable	**Mean (Male)**	**Mean (Female)**	***t*-value**	***p*-value**
Stride Length (m)	1.41	1.30	2.57	0.012
Ground Reaction Force (N)	261.34	227.56	3.14	0.002
Joint Angle (Knee, degrees)	32.58	30.13	1.82	0.072
Cadence (steps/min)	170.84	165.47	1.76	0.081
Foot Pressure (kPa)	91.26	83.14	2.48	0.015
Muscle Activation Time (ms)	243.73	228.19	1.63	0.107

The independent *t*-tests (**Table 3**) comparing male and female participants revealed significant differences in several biomechanical variables. Stride length was significantly longer in males (1.41 m) compared to females (1.30 m), with a *t*-value of 2.57 and a *p*-value of 0.012, indicating a statistically significant difference. Ground reaction force was also significantly higher in males (261.34 N) than in females (227.56 N), with a *t*-value of 3.14 and a *p*-value of 0.002. Foot pressure was notably higher in males (91.26 kPa) compared to females (83.14 kPa), with a *t*-value of 2.48 and a *p*-value of 0.015. Although males showed slightly larger knee joint angles (32.58 degrees) compared to females (30.13 degrees), this difference did not reach statistical significance ($t = 1.82$, $p = 0.072$). Cadence was also slightly higher in males (170.84 steps/min) than in females (165.47 steps/min), though this difference was not statistically significant ($t = 1.76$, $p = 0.081$). The comparison for muscle activation time showed a non-significant difference between males (243.73 ms) and females (228.19 ms) ($t = 1.63$, $p = 0.107$).

Table 4. Paired *t*-tests: Pre- and Post-Injury biomechanical comparison for injured participants.

Variable	**Mean (Pre-Injury)**	**Mean (Post-Injury)**	***t*-value**	***p*-value**
Stride Length (m)	1.42	1.33	2.39	0.019
Ground Reaction Force (N)	258.91	241.38	2.67	0.009
Joint Angle (Knee, degrees)	33.14	29.92	3.24	0.002
Cadence (steps/min)	171.56	165.84	2.17	0.034
Foot Pressure (kPa)	89.21	84.67	1.78	0.081
Muscle Activation Time (ms)	241.47	233.63	1.93	0.061

The paired *t*-tests (**Table 4**) comparing pre- and post-injury biomechanical data for injured participants revealed several significant changes. Stride length decreased significantly after injury, from 1.42 m to 1.33 m ($t = 2.39$, $p = 0.019$), indicating a shorter stride post-injury. Ground reaction force also significantly decreased post-injury, from 258.91 N to 241.38 N ($t = 2.67$, $p = 0.009$), and the knee joint angle showed a significant reduction from 33.14 degrees to 29.92 degrees ($t = 3.24$, $p = 0.002$). Cadence decreased from 171.56 steps/min to 165.84 steps/min following injury, with a *t*-value of 2.17 and a p-value of 0.034, indicating a significant reduction. Foot pressure decreased slightly from 89.21 kPa to 84.67 kPa, but this difference was not statistically significant (t = 1.78, p = 0.081). Finally, muscle activation time showed a non-significant reduction from 241.47 ms to 233.63 ms (t = 1.93, p = 0.061).

Table 5. One-Way ANOVA: Comparison of biomechanical variables across experience levels.

Variable	Mean (Recreational)	Mean (Intermediate)	Mean (Advanced)	*F*-value	*p*-value
Stride Length (m)	1.28	1.35	1.48	6.17	0.003
Ground Reaction Force (N)	221.73	245.14	262.41	8.26	0.001
Joint Angle (Knee, degrees)	29.64	31.28	34.53	5.19	0.007
Cadence (steps/min)	162.37	168.91	174.23	4.72	0.010
Foot Pressure (kPa)	82.91	87.13	92.48	4.11	0.018
Muscle Activation Time (ms)	229.64	236.91	244.58	3.94	0.023

The one-way ANOVA (**Table 5**) comparing biomechanical variables across different experience levels (Recreational, Intermediate, and Advanced runners) revealed significant differences in several key areas. Stride length was significantly different across the groups, with advanced runners having the longest strides (mean: 1.48 m), followed by intermediate (1.35 m) and recreational runners (1.28 m), with an F-value of 6.17 and a p-value of 0.003. Ground reaction force also showed significant differences, with advanced runners exhibiting the highest values (262.41 N), compared to intermediate (245.14 N) and recreational runners (221.73 N), indicated by an F-value of 8.26 and a p-value of 0.001. Knee joint angle varied significantly, with advanced runners showing a mean angle of 34.53 degrees, intermediate runners at 31.28 degrees, and recreational runners at 29.64 degrees (F = 5.19, p = 0.007). Cadence followed a similar trend, with advanced runners having the highest cadence (174.23 steps/min), followed by intermediate (168.91 steps/min) and recreational runners (162.37 steps/min), with an F-value of 4.72 and a p-value of 0.010. Foot pressure was significantly higher among advanced runners (92.48 kPa) compared to intermediate (87.13 kPa) and recreational runners (82.91 kPa), as shown by an F-value of 4.11 and a p-value of 0.018. Lastly, muscle activation time also exhibited significant differences, with advanced runners having the most extended muscle activation times (244.58 ms), compared to intermediate (236.91 ms) and recreational runners (229.64 ms) (F = 3.94, p = 0.023).

Table 6. Post-hoc analysis: Tukey's HSD test.

Variable	Comparison	Mean Difference	*p*-value
Stride Length (m)	Advanced vs. Recreational	0.20	0.002
	Advanced vs. Intermediate	0.13	0.037
	Intermediate vs. Recreational	0.07	0.245
Ground Reaction Force (N)	Advanced vs. Recreational	40.68	0.001
	Advanced vs. Intermediate	17.27	0.064
	Intermediate vs. Recreational	23.41	0.014
Joint Angle (Knee, Degrees)	Advanced vs. Recreational	4.89	0.005
	Advanced vs. Intermediate	3.25	0.041
	Intermediate vs. Recreational	1.64	0.298

The post-hoc Tukey's HSD test (**Table 6**) further explored group differences. For stride length, advanced runners showed a significant difference compared to recreational runners (mean difference = 0.20 m, p = 0.002) and intermediate runners (mean difference = 0.13 m, p = 0.037), but the difference between intermediate and recreational runners was not significant (p = 0.245). Advanced runners showed a significant difference in ground reaction force compared to recreational runners (mean difference = 40.68 N, p = 0.001), while the difference between advanced and intermediate runners approached significance (p = 0.064). Intermediate runners differed significantly from recreational runners (mean difference = 23.41 N, p = 0.014). In terms of knee joint angle, advanced runners differed significantly from recreational runners (mean difference = 4.89 degrees, p = 0.005) and intermediate runners (mean difference = 3.25 degrees, p = 0.041), while the difference between intermediate and recreational runners was not statistically significant (p = 0.298).

Table 7. Chi-square test: Relationship between experience level and injury occurrence.

Experience Level	Injured	Not Injured	Total
Recreational	15	17	32
Intermediate	9	25	34
Advanced	3	15	18
Total	27	57	84

Table 8. Chi-square test statistics.

Statistic	Value
Chi-square value	8.49
Degrees of freedom	2
p-value	0.014

The Chi-square test (**Tables 7** and **8**) was used to assess the relationship between experience level (Recreational, Intermediate, and Advanced runners) and injury occurrence (Injured vs. Not Injured). The results indicate a significant association between these variables, with a Chi-square value of 8.49, degrees of freedom = 2, and a p-value of 0.014, suggesting that injury occurrence varies significantly across the different experience levels. The Table shows that recreational runners had the highest number of injuries, with 15 out of 32 participants (46.88%) reporting injuries, while 17 did not. Among intermediate runners, 9 participants were injured (26.47%), and 25 were not. Advanced runners had the lowest injury rate, with only 3 out of 18 participants (16.67%) reporting injuries and 15 remaining injury-free. These results indicate that recreational runners are more likely to experience injuries than intermediate and advanced runners, possibly due to differences in running technique, experience, and conditioning. Advanced runners, who may have developed better biomechanics and injury prevention strategies, appear to be at a lower risk of injury.

Table 9. Pearson correlation analysis: Relationships Between biomechanical variables.

Variable	Stride Length	Ground Reaction Force	Joint Angle (Knee)	Cadence	Foot Pressure	Muscle Activation Time
Stride Length (m)	1.00	0.47	0.29	−0.51	0.38	−0.23
Ground Reaction Force (N)	0.47	1.00	0.52	−0.33	0.61	0.19
Joint Angle (Knee, degrees)	0.29	0.52	1.00	−0.21	0.43	0.37
Cadence (steps/min)	−0.51	−0.33	−0.21	1.00	−0.41	0.28
Foot Pressure (kPa)	0.38	0.61	0.43	−0.41	1.00	0.35
Muscle Activation Time (ms)	−0.23	0.19	0.37	0.28	0.35	1.00

The Pearson correlation analysis (**Table 9**) provides insights into the study's relationships between various biomechanical variables. Stride length showed a moderate positive correlation with ground reaction force (r = 0.47), indicating that participants with longer strides tend to generate higher ground reaction forces. Additionally, a weaker positive correlation exists between stride length and knee joint angle (r = 0.29), suggesting that larger joint angles may contribute to a longer stride. Ground reaction force was also positively correlated with foot pressure (r = 0.61), highlighting that greater forces during foot strike lead to higher pressure on the foot. A moderate positive correlation was observed between ground reaction force and knee joint angle (r = 0.52), meaning that participants generating higher forces tend to have larger knee joint angles.

On the other hand, cadence showed a negative correlation with several variables. There was a moderate negative correlation between cadence and stride length (r = −0.51), meaning that participants with higher step rates tend to have shorter strides. Similarly, foot pressure (r = −0.41) and ground reaction force (r = −0.33) had negative correlations with cadence, suggesting that faster step rates are associated with lower force and pressure during running. Finally, muscle activation time had a weak negative correlation with stride length (r = −0.23) but positive correlations with joint angle (r = 0.37) and foot pressure (r = 0.35). This implies that participants with longer muscle activation times tend to have higher joint angles and foot pressure during the running cycle.

Table 10. Multiple linear regression analysis: Predictors of injury severity.

Predictor Variable	Unstandardized Coefficient (B)	Standard Error	Standardized Coefficient (Beta)	*t*-value	*p*-value
Stride Length (m)	1.37	0.43	0.34	3.18	0.002
Ground Reaction Force (N)	0.19	0.07	0.29	2.64	0.010
Joint Angle (Knee, degrees)	0.84	0.29	0.27	2.89	0.005
Cadence (steps/min)	−0.57	0.21	−0.22	−2.71	0.008
Foot Pressure (kPa)	0.42	0.11	0.31	3.82	0.001
Muscle Activation Time (ms)	0.13	0.06	0.18	2.21	0.032

Table 11. Model summary.

Statistic	Value
R-squared	0.61
Adjusted *R*-squared	0.58
F-statistic	14.27
p-value (Model)	<0.001

The multiple linear regression analysis (**Tables 10** and **11**) was conducted to determine how various biomechanical factors predict injury severity. The model explains 61% of the variance in injury severity (*R*-squared = 0.61), indicating that the predictors included in the model strongly influence the outcome. The model is statistically significant, as reflected by an *F*-statistic of 14.27 and a *p*-value less than 0.001, suggesting that the overall model reliably predicts injury severity.

Several biomechanical variables were found to be significant predictors of injury severity:

- Stride length positively influenced injury severity, with a standardized coefficient (Beta) of 0.34. The unstandardized coefficient ($B = 1.37$) indicates that, for every 1 m increase in stride length, injury severity increases by 1.37 units, with this effect being statistically significant ($p = 0.002$).
- Ground reaction force was also a significant predictor ($B = 0.19$, Beta = 0.29, $p = 0.010$), indicating that greater forces exerted during running are associated with higher injury severity. A 1 N increase in ground reaction force increases injury severity by 0.19 units.
- Knee joint angle positively impacted injury severity ($B = 0.84$, Beta = 0.27, $p = 0.005$). This suggests that greater knee flexion during running increases the likelihood of injury severity.
- Cadence was negatively associated with injury severity ($B = -0.57$, Beta = −0.22, $p = 0.008$), implying that higher step rates may reduce injury severity. Specifically, injury severity decreases by 0.57 units for each additional step per minute.
- Foot pressure was among the strongest predictors, with a standardized coefficient of 0.31 and a significant unstandardized coefficient of 0.42 ($p = 0.001$). Higher foot pressure during running is linked to increased injury severity.
- Muscle activation time also positively correlated with injury severity ($B = 0.13$, Beta = 0.18, $p = 0.032$). Longer muscle activation times slightly increase the likelihood of higher injury severity.

The adjusted *R*-squared value of 0.58 suggests that after accounting for the number of predictors, the model still explains a substantial portion of the variation in injury severity. These findings underscore the importance of biomechanical factors such as ground reaction force, stride length, and foot pressure in predicting the severity of running-related injuries, while higher cadence may serve as a protective factor.

3.2. Machine learning analysis

Table 12. Feature importance ranking (GBDT).

Feature	Importance Score	Rank
Ground Reaction Force (N)	0.312	1
Foot Pressure (kPa)	0.245	2
Stride Length (m)	0.183	3
Joint Angle (Knee, degrees)	0.141	4
Cadence (steps/min)	0.079	5
Hip Flexion (degrees)	0.061	6
Vertical Displacement (cm)	0.046	7
Ankle Dorsiflexion (degrees)	0.043	8
Muscle Activation Time (ms)	0.040	9

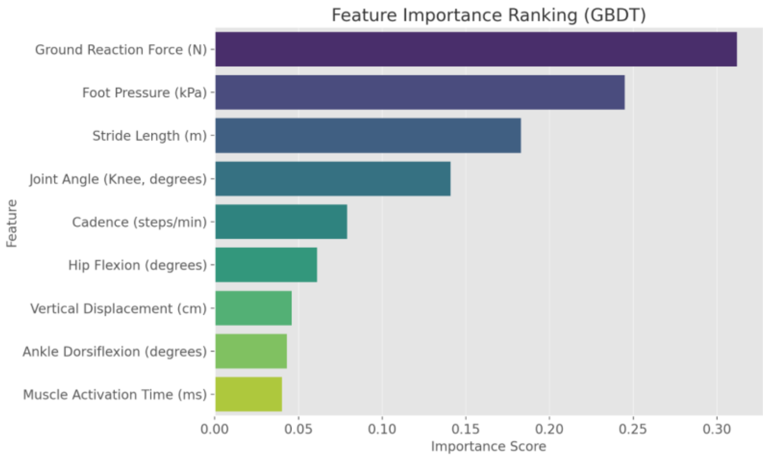

Figure 5. Feature rank.

Figure 5 and **Table 12** list the feature ranks; the Gradient-Boosted Decision Trees (GBDT) feature importance ranking highlights that ground reaction force (importance score = 0.312) is the most influential factor in predicting injury risk. Foot pressure follows closely with a score of 0.245, and stride length ranks third with 0.183. Knee joint angle has an importance score of 0.141, placing it fourth, while cadence ranks fifth with a score of 0.079. Hip flexion (0.061), vertical displacement (0.046), ankle dorsiflexion (0.043), and muscle activation time (0.040) have a lower influence, rounding out the list of critical biomechanical predictors.

Table 13. Accuracy comparison.

Model	Accuracy (%)
Gradient-Boosted Decision Trees (GBDT)	83.74
Long Short-Term Memory (LSTM)	81.29
Support Vector Machines (SVM)	78.91
Random Forest (RF)	80.56
Decision Tree (DT)	76.48
Proposed Ensemble Model	88.37

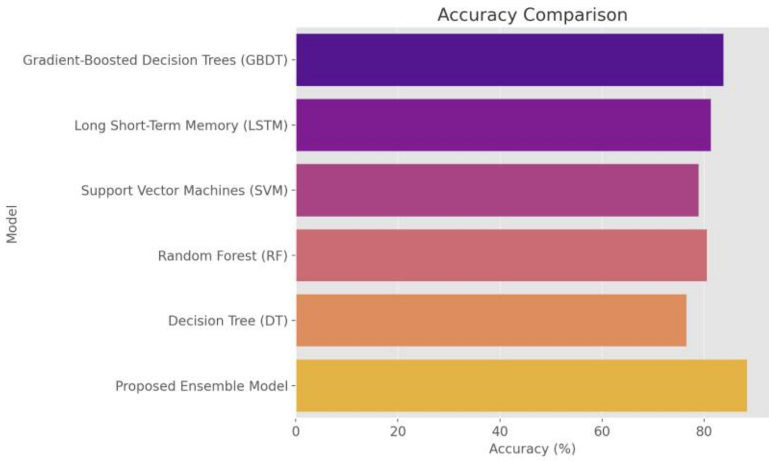

Figure 6. Accuracy comparison.

Table 14. Precision, recall, and F1-score comparison.

Model	Precision (%)	Recall (%)	F1 Score (%)
Gradient-Boosted Decision Trees (GBDT)	82.45	84.67	83.54
Long Short-Term Memory (LSTM)	80.17	82.94	81.53
Support Vector Machines (SVM)	77.32	79.58	78.44
Random Forest (RF)	79.63	81.18	80.39
Decision Tree (DT)	75.19	76.84	75.99
Proposed Ensemble Model	87.84	89.12	88.47

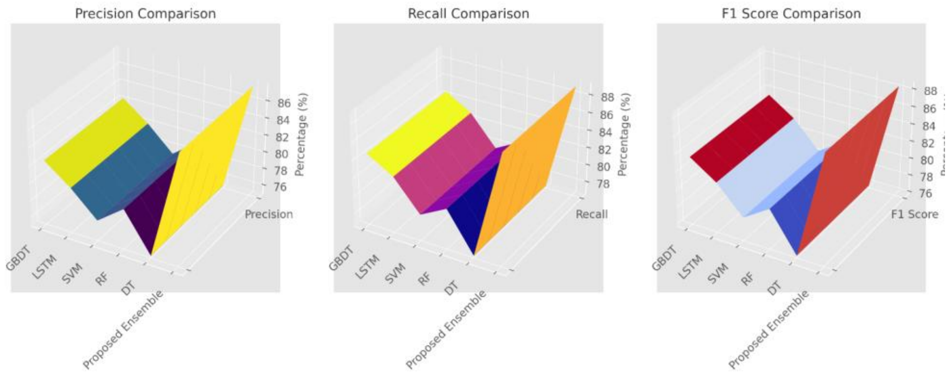

Figure 7. Precision, recall, and F1-score analysis.

The accuracy comparison (**Figure 6** and **Table 13**) shows that the Proposed Ensemble Model achieved the highest accuracy at 88.37%, outperforming individual models such as GBDT (83.74%), LSTM (81.29%), RF (80.56%), SVM (78.91%), and Decision Tree (DT) (76.48%). Regarding precision, recall, and F1 score (**Figure 7** and **Table 14**), the Proposed Ensemble Model also outperformed other models, with a precision of 87.84%, recall of 89.12%, and an F1 score of 88.47%. GBDT followed with a precision of 82.45%, recall of 84.67%, and F1 score of 83.54%. LSTM achieved an F1 score of 81.53%, while Random Forest scored 80.39%. SVM and DT had lower F1 scores, at 78.44% and 75.99%, respectively.

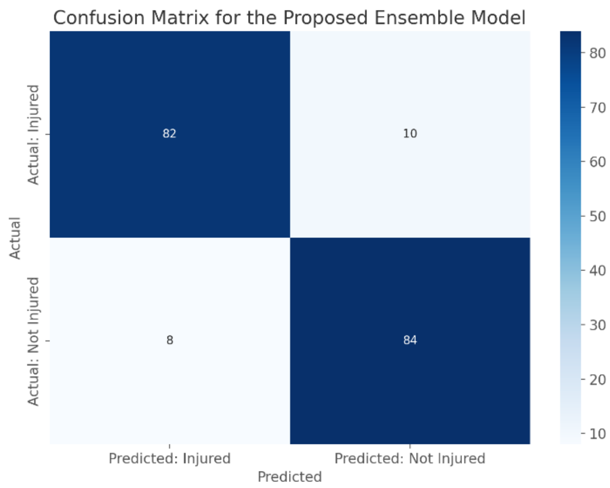

Figure 8. Confusion matrix.

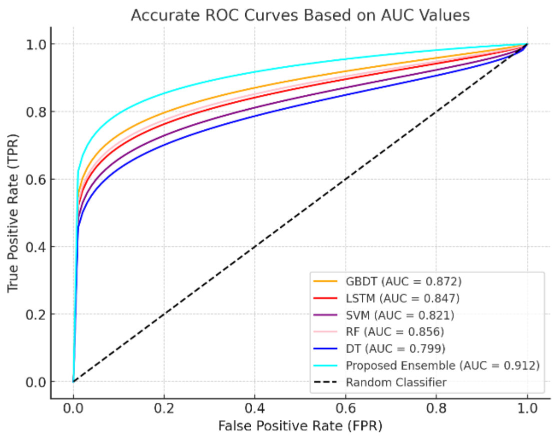

Figure 9. AUC results.

The confusion matrix (**Figure 8**) for the Proposed Ensemble Model shows that out of 92 actual injured participants, 82 were correctly classified as injured (True Positives), while 10 were misclassified as not injured (False Negatives). Of the 92 non-injured participants, 84 were correctly identified as not injured (True Negatives), and 8 were incorrectly classified as injured (False Positives). The model performed well, with only 18 misclassifications out of 184 cases. In the AUC–ROC curve comparison (**Figure 9**), the Proposed Ensemble Model had the highest AUC value at 0.912, indicating a superior ability to distinguish between injured and non-injured participants. GBDT followed with an AUC of 0.872, Random Forest with 0.856, and LSTM with 0.847. SVM and Decision Tree had lower AUC values, at 0.821 and 0.799, respectively. This confirms that the Proposed Ensemble Model offers this study's best overall classification performance.

4. Conclusion and future work

This study presents a novel approach to predicting running-related injuries by integrating machine learning techniques with biomechanical analysis. The ensemble model combining GBDT, LSTM, and SVM demonstrated superior predictive accuracy and effectiveness compared to individual models. The analysis identified ground reaction force, foot pressure, and stride length as the most critical biomechanical factors influencing injury risk. These findings highlight the potential of using real-time data from wearable sensors and advanced machine-learning algorithms to enhance injury prediction and prevention in runners. The proposed model provides accurate injury risk assessments and offers a more comprehensive understanding of the interplay between biomechanical factors during running. The ability to capture temporal dependencies in biomechanical data through LSTM networks and the feature importance ranking provided by GBDT allows for a more nuanced analysis of injury risks.

Furthermore, applying this approach in real-world settings—both on outdoor tracks and in controlled laboratory environments—demonstrates its versatility and practicality for broader athletic populations. Future research should explore expanding the model to incorporate additional variables, such as environmental conditions and training load, to refine injury predictions further. Additionally, developing real-time feedback systems based on this model could provide athletes with actionable insights during training, enabling them to make adjustments that reduce their injury risk.

Overall, this research contributes to the growing body of evidence supporting the use of machine learning in sports science and provides a foundation for future innovations in injury prevention strategies.

Author contributions: Conceptualization, RH, FQ, HW and MY; methodology, RH, FQ, HW and MY; software, RH, FQ, HW and MY; validation, RH, FQ, HW and MY; formal analysis, RH, FQ, HW and MY; investigation, RH, FQ, HW and MY; resources, RH, FQ, HW and MY; data curation, RH, FQ, HW and MY; writing—original draft preparation, RH, FQ, HW and MY; writing—review and editing, RH, FQ, HW and MY; visualization, RH, FQ, HW and MY. All authors have read and agreed to the published version of the manuscript.

Funding: Philosophy and social sciences research project of Department of Education of Hubei Province grant number [22D104]. This work was funded by the Wuhan Sports University Young Faculty Research grant number [2022J02]. Young and middle-aged scientific research team of Wuhan Sports University grant number [21KT16]. The 14th Five-Year-Plan Advantageous and Characteristic Disciplines (Groups) of Colleges and Universities in Hubei Province (2021-05).

Ethical approval: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

References

1. Herbert, C. (2022). Enhancing mental health, well-being and active lifestyles of university students employing physical activity and exercise research programs—frontiers in public health, 10, 849093.
2. Hsu, C. L., Yang, C. H., Wang, J. H., & Liang, C. C. (2020). Common running musculoskeletal injuries and associated factors among recreational gorge marathon runners: An investigation from 2013 to 2018 taroko gorge marathons. International journal of environmental research and public health, 17(21), 8101.
3. Dempster, J., Dutheil, F., & Ugbolue, U. C. (2021). The Prevalence of Lower Extremity Injuries in Running and Associated Risk Factors: A Systematic Review. Physical Activity & Health (2515-2270), 5(1).

4. Leistra, R. J., & Lategan, L. (2022). Running-related injury prevalence in Gauteng long-distance runners over a 12-month period (Doctoral dissertation, University of Johannesburg).
5. Aguillard, E., Cane, J., & Goldstein, A. L. (2023). Run Healthy: The Runner's Guide to Injury Prevention and Treatment. Human Kinetics.
6. Jacob, A. (2023). Injury Prevention and Rehabilitation in Sports. International Journal of Sports, Yoga and Physical Activity, ISSN: 3005-5083, 2(1), 8-13.
7. Brenner, J. S., Watson, A., Brooks, M. A., Carl, R. L., Briskin, S. M., Canty, G., ... & Emanuel, A. (2024). Overuse injuries, overtraining, and burnout in young athletes. Pediatrics, 153(2).
8. Kalema, R. N., Schache, A. G., Williams, M. D., Heiderscheit, B., Siqueira Trajano, G., & Shield, A. J. (2021). Sprinting biomechanics and hamstring injuries: Is there a link? A literature review. Sports, 9(10), 141.
9. Lin, J. Z., Chiu, W. Y., Tai, W. H., Hong, Y. X., & Chen, C. Y. (2021). Ankle muscle activations during different foot-strike patterns in running. Sensors, 21(10), 3422.
10. Papagiannaki, M., Samoladas, E., Maropoulos, S., & Arabatzi, F. (2020). Running-related injury from an engineering, medical and sport science perspective. Frontiers in bioengineering and biotechnology, 8, 533391.
11. Agresta, C., Giacomazzi, C., Harrast, M., & Zendler, J. (2022). Running injury paradigms and their influence on footwear design features and runner assessment methods: A focused review to advance evidence-based practice for running medicine clinicians. Frontiers in sports and active living, 4, 815675.
12. Peebles, A. T., Arena, S. L., & Queen, R. M. (2021). A new method for assessing landing kinematics in non-laboratory settings. Physical therapy in sport, 49, 21-30.
13. Hafer, J. F., Vitali, R., Gurchiek, R., Curtze, C., Shull, P., & Cain, S. M. (2023). Challenges and advances in the use of wearable sensors for lower extremity biomechanics. Journal of biomechanics, 111714.
14. Ferris, J. (2023). Advancing Inertial Sensors for Performance Quantification: Applications in Balance Rehabilitation and Distance Running (Doctoral dissertation).
15. S. Sudhakar and S. Chenthur Pandian, (2016), 'Hybrid Cluster-based Geographical Routing Protocol to Mitigate Malicious Nodes in Mobile Ad Hoc Network, InderScience-International Journal of Ad Hoc and Ubiquitous Computing, vol. 21, no. 4, pp. 224-236. DOI:10.1504/IJAHUC.2016.076358.
16. Lloyd, D. (2021). The future of in-field sports biomechanics: Wearables plus modelling compute real-time in vivo tissue loading to prevent and repair musculoskeletal injuries. Sports Biomechanics, 1-29.
17. Foreman, B., Lissak, I. A., Kamireddi, N., Moberg, D., & Rosenthal, E. S. (2021). Challenges and opportunities in multimodal monitoring and data analytics in traumatic brain injury. Current neurology and neuroscience reports, 21, 1-9.
18. Phatak, A. A., Wieland, F. G., Vempala, K., Volkmar, F., & Memmert, D. (2021). Artificial intelligence-based body sensor network framework—narrative review: proposing an end-to-end framework using wearable sensors, real-time location systems and artificial intelligence/machine learning algorithms for data collection, data mining and knowledge discovery in sports and healthcare. Sports Medicine-Open, 7(1), 79.
19. Hafer, J. F., Vitali, R., Gurchiek, R., Curtze, C., Shull, P., & Cain, S. M. (2023). Challenges and advances in the use of wearable sensors for lower extremity biomechanics. Journal of biomechanics, 111714.
20. Matijevich, E. S., Scott, L. R., Volgyesi, P., Derry, K. H., & Zelik, K. E. (2020). Combining wearable sensor signals, machine learning and biomechanics to estimate tibial bone force and damage during running. Human movement science, 74, 102690.
21. Taye, M. M. (2023). Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. Computers, 12(5), 91.
22. Strielkowski, W., Vlasov, A., Selivanov, K., Muraviev, K., & Shakhnov, V. (2023). Prospects and challenges of the machine learning and data-driven methods for the predictive analysis of power systems: A review. Energies, 16(10), 4025.
23. Ali, Y., Hussain, F., & Haque, M. M. (2024). Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review: Accident Analysis & Prevention, 194, 107378.
24. Agresta, C., Giacomazzi, C., Harrast, M., & Zendler, J. (2022). Running injury paradigms and their influence on footwear design features and runner assessment methods: A focused review to advance evidence-based practice for running medicine clinicians. Frontiers in sports and active living, 4, 815675.
25. Cox Jr, L. A. for Decision and Risk Analysis.

26. Weakley, J., Black, G., McLaren, S., Scantlebury, S., Suchomel, T. J., McMahon, E., ... & Read, D. B. (2024). Testing and profiling athletes: recommendations for test selection, implementation, and maximizing information. Strength & Conditioning Journal, 46(2), 159-179.
27. Cecchetto, N. R., Medina, S. M., & Ibargüengoytía, N. R. (2020). Running performance with emphasis on low temperatures in a Patagonian lizard, Liolaemus lineomaculatus. Scientific Reports, 10(1), 14732.