Proactive Injury Risk Assessment in High-Performance Athletics Using Multi-Modal Wearable Sensor Data: A Comparative Analysis of Ensemble Learning and Deep Temporal Architectures

MSc Research Project
MsC in Data Analytics

Akshada Warik
Student ID: x23399155

School of Computing
National College of Ireland

Supervisor: Dr. Abid Yaqoob

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Akshada Warik
Student ID:	x23399155
Programme:	Msc in Data Analytics
Year:	2025
Module:	MSc Research Project
Supervisor:	Dr. Abid Yaqoob
Submission Due Date:	11/12/2025
Project Title:	Proactive Injury Risk Assessment in High-Performance Athletics Using Multi-Modal Wearable Sensor Data: A Comparative Analysis of Ensemble Learning and Deep Temporal Architectures
Word Count:	XXX
Page Count:	24

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	27th January 2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Proactive Injury Risk Assessment in High-Performance Athletics Using Multi-Modal Wearable Sensor Data: A Comparative Analysis of Ensemble Learning and Deep Temporal Architectures

Akshada Warik
x23399155

Abstract

In the contemporary landscape of high-performance sports, the margin between peak physical conditioning and musculoskeletal injury is increasingly narrow. As athletes push physiological boundaries, the necessity for a paradigm shift from reactive medical treatment to proactive risk assessment has become paramount. This research presents a robust predictive analytics framework leveraging multi-modal biometric data harvested from wearable sensors. Utilizing a synthetic dataset of 1,000 high-fidelity training sessions, we monitor critical physiological indicators including Heart Rate Variability (HRV), Blood Oxygen Saturation (SpO2), Skin Temperature, Biomechanical Impact Force, and a derived Cumulative Fatigue Index.

We conduct a rigorous comparative analysis between traditional feature-based machine learning models (Logistic Regression, Support Vector Machines, Random Forests) and advanced deep learning architectures (Baseline Long Short-Term Memory networks and LSTM Autoencoders) to detect pre-injury states. Our Exploratory Data Analysis (EDA), reveals statistically significant differences (p ¡ 0.001) in Cumulative Fatigue Index, Impact Force, and Injury Risk Score between injured and non-injured sessions, while other metrics such as SpO2 and Respiratory Rate remain non-discriminative. Furthermore, we introduce a granular time-stamp prediction mechanism capable of identifying risk probabilities at specific millisecond intervals using sequence models.

LSTM Autoencoder attains a Recall of 1.00 with an F1-Score of 0.7616, prioritizing sensitivity to injury cases, while SVM and Random Forest baselines consistently achieve Recall values above 0.88. This study provides a blueprint for integrating real-time AI into coaching dashboards, combining interpretable ensemble methods with highly sensitive temporal deep learning pipelines.

Index Terms

Sports Analytics, Wearable Technology, LSTM, LSTM Autoencoder, Random Forest, SVM, Logistic Regression, Injury Prevention, Biometrics, Time-Series Analysis.

1 Introduction

1.1 Research Background and Motivation

The fundamental nature of the competitive sports world is characterized by a relentless drive to achieve optimal performance and push human physiological limits. This pursuit of excellence necessitates constant innovation in training methodologies, recovery protocols, and, most critically, injury mitigation strategies. Historically, sports medicine has operated on a reactive paradigm: treating injuries after they occur. However, the last decade has witnessed a technological revolution, with wearable technologies becoming ubiquitous in the sports sector. These devices, ranging from GPS trackers to bio-harnesses, provide continuous, high-resolution monitoring of an athlete's physiological condition.

Artificial intelligence (AI) integrating with biomechanics and sports medicine is a new dawn in performance optimization and injury prevention. The technology of wearable devices, powered by AI, does not only provide raw numbers; it can give detailed real time analysis and predictive feedback loops, as well as customized recommendations, which are aimed at both improving athletic performance and, at the same time, reducing inherent risks (1). The strategic implementation of wearable devices technology with advanced analytics reinvents the classic reactive approach to athletic management. This technology does not merely make a solution to a problem apparent and consequently seen but rather results in a paradigm shift towards preventative actions and ensures that the risk-indicating factors are identified long before they translate into clinical manifestations.

1.2 Defining the Research Problem and Gap

Although sensors that measure Heart Rate (HR), Respiratory Rate, and SpO 2 have become numerous, the existing commercial approaches are mostly descriptive, as opposed to being predictive. There is a serious disparity in the multi-modal sensor information translation into credible real-time injury danger scores. Besides, a debate exists on the most appropriate modeling technique, whether interpretable machine learning models including Decision Trees and Logistic Regression are adequate, or more complicated deep learning models are necessitated to explain the temporal dynamics of fatigue accumulation.

It is also evident that there is a significant disparity between the translation of the multi-modal biometric data that is rich and highly reactive and the proactive real-time software that can be used to measure musculoskeletal injury risk in high-performance athletes who are constantly subjected to overtraining and cumulative physical stress (2). Current injury prevention approaches tend to be small scale and use single measures of biomechanical sensors or parameter models. These methods do not reflect the multidimensionality of the occurrence of injury that is not often explained by a single acute determinant.

The main goal of the given research is to create and test a predictive, real-time injury risk model, which would combine physiological fatigue and biomechanical stress, based on machine learning and deep learning algorithms. The study, specifically, aims at finding out whether temporal deep learning models tend to outperform traditional interpretable machine learning models in the detection of early injury risk during the scenario of cumulative physical load and insufficient rest.

This goal is driven by the desire to simulate the interaction synergism of internal physiological exhaustion and external biomechanical stress. Indicatively, high intensity of training in terms of ImpactForceNewtons can be tolerable at specific stages in training but at the same time accompanied by reduced Heart Rate Variability (HRV) or high CumulativeFatigueIndex, the risk of injury becomes significantly high. This study seeks to give a contextualized and continuous risk assessment as opposed to parameter-based alerts that are isolated by doing the 3 factors jointly.

According to epidemiological studies, approximately 30-40 percent of athletes suffer at least one musculoskeletal injury in a year, and the rates go higher to over 60 percent in sporting activities with high intensity because it puts strains on the body and causes recurrent stresses (3) (4). The current development of wearable sensing technologies allows a permanent status of heart rate, oxygen saturation, skin temperature, and mechanical forces, and it creates a possibility to identify abnormal physiological and biomechanical tendencies before the event of clinical injury. The existing studies indicate that variations in these indicators are likely to pre-empt injury accidents, which support the possibility of predicting the risk proactively and sensor-immediately.

In line with this, this paper discusses a 1,000 athletic training session dataset and uses statistical and machine learning libraries in Python (pandas, seaborn, scipy). Classical algorithms such as Logistic Regression, Support Vector Machines (SVM) and Random Forests are compared to Long Short-Term Memory (LSTM) networks and LSTM Autoencoders to measure the trade-offs between interpretability and temporal predictive error. One of the contributions of this work is the fact that it introduced a timestamp-level prediction capability that was able to define the specific point in a training session where the risk of injury exceeds a particular threshold.

2 Related Work

Recent studies are becoming more and more indicative that wearable sensors and machine learning can be considered powerful tools to comprehend fatigue, load, and injury risk among athletes. Research in both fields of sports and healthcare shows how biometric and biomechanical data can be used to provide early indications before injuries happen. This section covers the major contributions that guide the present trends in predicting injuries with data.

In this vein, Kovoor et al. (2024) (5) introduced a prospective overview of the next generation wearable systems that are upgraded with the help of IoT systems and automatic analytics. In their work, they focus on the way multimodal sensors and real-time data streaming with the assistance of ML pipelines may help to monitor the athletes and provide them with information on the field. It is conceptual, but it describes the technological path toward the completely automated injury prevention systems. Their vision is in line with the objectives of the present project, which tries to move forward without theory frameworks to practical predictive modelling based on fused physiological and mechanical sensor data.

The article by Subramaniam et al. (2022) (4) is an exhaustive review of wearable IMU-based systems that detect fall-risks in the elderly. The target population varies with the athletes; however, the theory behind gait monitoring, identifying anomalies, and instability in the movement patterns are directly applicable to the sports setting. A great number of sports related injuries occur due to asymmetry, deviations caused by fatigue, or unusual patterns of movement, and the approaches reviewed are useful to inform the elements of a movement-analysis in the current systems of injury prediction.

Zadeh et al. (2021) (3) asked the question of whether musculoskeletal injury in military cadets could be forecasted based on the data gathered with wearable sensors commercially available. They were measured by Zephyr BioHarness equipment, which provided information on physiological parameters, heart rate, and respiration, as well as biomechanical, acceleration, and external loads. Body mass index (BMI) and repetitive mechanical loading were found to be important predictors of early-season injuries (their results). Even though the sample size was small and the modelling relatively simple, the study proved that the low-complexity sensor data can be used to detect the risk of injury early. This will be a good starting point to the current study, indicating that wearable physiological and bio-mechanical indicators can be effective predictive modelling signals.

Iduh et al. (2024) (2) suggested a conceptual predictive framework of determining injury risk in athletes in 2024. In their model, they combine external load, training history and physiological factors in a single risk score. Although the framework is yet to be experimentally tested and has no performance metrics, it is aligned with modern trends in sports analytics that support multifactorial models on single variable measurements. Their work offers conceptual basis to the current study, in terms of integration of various biometric and workload measures.

Pilka et al. (2023) (6) worked with GPS-wearing instruments to track players in the field of football and anticipate soft-tissue injuries. Machine learning of sprint counts, acceleration changes, and other variables of training-loads demonstrated high correlations between load patterns and injury occurrence. Their research approach is practical and field-based, which enables the realistic feasibility of wearable-driven injury prediction. The study will be useful as it will support the project at hand to explain how physiological time-series data can be used together with external-load metrics to enhance the prediction accuracy.

Shah et al. (2025) (7) discussed the opportunities of using big data and AI in improving both strategic performance analysis and injury prevention. Their model combines contextual game information and biometric measurements to develop a holistic system of monitoring athletes. In its impactful and still conceptual form, the paper identifies the advantage of combining performance analytics with personalised risk modelling. The multidisciplinary view corresponds to the objectives of the current study that is aimed at producing actionable information to aid in the optimisation of the performance and the minimisation of injuries.

Albahri et al. (2019) (8) concentrated on the strength of real-time data streams within mHealth IoT systems, in which the authors performed research on issues of transmission reliability, missing data, and noise. Their setting is the healthcare industry, but the system-related problems they consider, including data integrity, redundancy, or fault tolerance, apply to sports applications that also demand constant wearable and monitoring. Their results highlight the significance of sound data pipelines, which is a major prerequisite of any predictive injured monitoring system.

Eisenkraft et al. (2023) (9) created a multi-parameter monitoring device which is a wearable that generates early warning scores based on physiological indicators of stress. The system proved to be capable of identifying the first indicators of medical environment deterioration. Though acute clinical incidences were used as the main target, and not sporting injuries, their study proves the viability of real-time physiological monitoring and early-warning systems, which the current study will seek to generalize to proactive determination of injury risk in the context of sporting settings.

Alagdeve et al. (2024) (10) conducted a review of the recent sensing technologies in wearables with regard to real-time biomechanical analysis in high-level sport. They emphasized the value of the high-frequency motion tracking, in collaboration with IoT systems, to improve the performance monitoring and decrease the risk of injuries. Their review, although with less field validation, supports the increasing centrality of wearable biomechanical data in high-performance sport and provides the biomechanical aspects of the predictive model in the current project.

Alghamdi (2023) (11) used deep learning, namely recurrent neural networks when the authors used the wearable sensor time-series data to make predictions of athlete health. They find that RNNs are able to model temporal dependencies better than conventional ML models, which provide better classification accuracy. Although issues, including data sparsity, persist, this literature is highly in favour of using temporal deep learning models, such as LSTM and CNN-LSTM, to form the core of the current project approach to sequential biometric data modelling.

Tahir et al. (2023) (12) came up with a multi-modal wearable mechanism that can identify the risky upper-limb movements that are related to musculoskeletal injury. Their platform detects dangerous postures and patterns of motion by means of combining biomechanical and physiological sensors. Despite the fact that it only evaluates upper limbs, the study illustrates the usefulness of the fine-grained motion analytics in the early detection of injuries. It is in this view that biomechanical indicators are incorporated into the larger predictive model that is applied in this dissertation.

Velusamy et al. (2023) (13) have conducted a review of wearable IMU-based gait analysis systems on real-time fall prediction systems. The focus on low-latency pipelines and fast inference is relevant to the sports setting in direct proportion, where the detection of the gait irregularity or instability is crucial in a short time. Although the population varies, the methodological insights guide the real-time aspect of processing of injury risk modelling.

Uddin and Syed-Abdul (2020) (14) gave a general summary of wearable sensor analytics in the health care sector, including the detection of anomalies, feature identification, and the process of processing data on a large scale. Their effort is applicable to sports analytics since it describes effective methods of handling large sensor data volumes, training of models with high accuracy and effectiveness. These guidelines stand behind the creation of strong data pipelines of the present project.

Aswini et al. (2024) (15) emphasised the notion that wearable sensor systems have become more relevant in professional sports in order to track physiological and biomechanical performance indicators. In their article, a demonstration of how fatigue, overtraining, and abnormal patterns may be detected at the initial stage of development and prevented to avoid injury and enhance performance is presented. This is in direct correlation with the current study, which will also aim at using multi-sensor data streams to preempt injury risk.

Bashkaran et al. (2023) (16) analyzed the application of wireless sensor networks as the means of improving the safety of athletes. Their system sends real-time information in order to detect any stress-related movements that may cause a wound. They point out some of the most important technical issues, such as the decrease in latency and the stability of the network, which is necessary in scalable and real-time injury prognosis systems. Their results guide the infrastructural needs of reliable predictive analytics in sports.

Ahmad et al. (2024) (17) wrote about the way wearable systems are integrated to change the landscape of sports performance monitoring, that is, they provide such variables as the heart rate, acceleration, posture, and impact forces. They stress the need to combine multimodal data with predictive modelling in the assessment of injury risks. This is directly conducive to the direction followed by the current study incorporating physiological and biomechanical characteristics in a hybrid machine learning and deep learning model to achieve a total monitoring of an athlete.

Kathirgamanathan et al. (2023) (18) examine the idea of how fatigue can be recognized in runners and, more importantly, described by time series data on wearable inertial sensors. The article of their work titled Explaining fatigue in runners using time series analysis on wearable sensor data indicates that (i) fatigue-related biomechanical adaptations can be trained on IMU time series, (ii) time-series classification with explanation methods can help understand how and where fatigue occurs during running gait, and (iii) explainable models can be used to bridge the gap between raw wearable data and coaching action. This is closely related to our study objective of applying wearable data and machine/deep learning to foresee injury-related conditions.

Phatak et al. (2021) (1) present an in-depth model of applying wearable sensors, real-time location, and AI to assist in making well-informed decisions in sports and healthcare. In their review, they point out that despite the large number of studies that deploy wearable devices, the information is not usually collected and performed across various types of sensors. To solve this, they suggest a body sensor network, which end to end gathers and synchronises physiological, biomechanical and positional data, which allows better machine learning analysis. This article has significance to our study since it indicates the value of multi-modal, high-quality wearable data to make strong predictions of injuries, and underpins the use of AI/ML models, including the Random Forest, SVM and LSTM-based models we use, in order to draw meaningful conclusions about athlete monitoring systems.

Chouhan et al. (2025) (19) introduce a machine learning approach to predictive injury in physical education classrooms to illustrate how the data on the activities of students can be evaluated to see those who are at higher risk. Their paper compares various ML models and demonstrates that pattern-related to overexertion and improper movement can be successfully ferried by supervised learning, and both of them are frequent causes of injury. The authors point out that predictive analytics will be able to assist teachers and coaches by providing them with early risk alerts ahead of the injury, hence during physical activity sessions making them safer and more customised. This article is connected to our research as it confirms the usefulness of the classical ML classifiers, i.e., the Random Forest and SVM, in the context of the injury prediction tasks, and it also supports our comparison of the classical machine learning and deep learning models in terms of wearable sensor data.

The article by Hu and Liu (2024) (20) presents a real-time athlete health monitoring and early warning system implemented with the use of an Internet of Things (IoT) architecture, which is a combination of wearable biomedical sensors and machine learning models equipped to react to abnormal physiological or training-load conditions. Their system feeds data on a constant stream and implements adaptive modelling and predictive analytics, and creates alerts whenever the metrics of an athlete exceed or do not meet healthy levels. Especially, it is observed that the algorithms such as the Random Forest can effectively determine the risky states which suggests high potential of preventing injuries even before they happen. This article can be applied in our study as it unveils the idea that sensor-based health data and AI models could be applied to detect early warning signs, which helps to support the notion that wearable-based datasets, including the one deployed in our project, can be helpful to develop predictive injury classification and anomaly detection systems.

Overall, AI-enhanced wearable systems have great potential to forecast injuries, as noted in the literature, but lack of an integrated validation, data integration, and temporal modelling was identified. These observations explain why a hybrid machine learning model, LSTM Autoencoder and deep learning system, which is the one implemented in the present research, is necessary to represent the dynamic trends in pre-injury stages.

3 Methodology

The methodology for this project is anchored in a rigorous comparative analytical framework designed to develop, implement, and evaluate a predictive system for injury risk assessment using various biosensor technologies. This section details the architecture of the proposed solution, the characteristics of the dataset utilized, the data preparation and feature engineering pipeline, the model development specifications, and the evaluation protocol.

3.1 Dataset Description

The study utilizes dataset representing 1,000 training sessions (21). The feature set includes:

- Heart Rate (BPM): Indicator of cardiovascular stress.
- Respiratory Rate (BPM): Correlates with anaerobic threshold.
- Skin Temperature ($\circ$C): Surface temperature; rises with exertion, drops with shock/fatigue.
- SpO2 (%) : Blood oxygen saturation.
- Impact Force (Newtons): Vertical ground reaction force; proxy for musculoskeletal load.
- Fatigue Index (0–1): A derived metric combining HRV and workload (Cumulative Fatigue Index).
- Injury Risk Score (0–1): Probabilistic score of injury likelihood.
- Binary Label: `Injury_Occurred` $\in \{0, 1\}$.

3.2 Preprocessing Pipeline

- Raw sensor data was normalized using StandardScaler (z-score standardization) scaling to map all features into a [0, 1] range for the deep learning models. For the LSTM-based architectures, we employed a sliding window approach with a window size of T = 30 time steps, allowing the model to "look back" at the previous 30 data points to predict the state of the current time step.
- We initially trained a baseline LSTM using one timestep per session, and later extended this to a full sequence model using a sliding window of T = 30 steps for temporal modeling and timestamp-specific predictions.

- For classical models (Logistic Regression, Random Forest, SVM), we constructed scikit-learn Pipelines that handle: Column-wise preprocessing (numeric vs. categorical features), Standardization using StandardScaler where required, Model-specific classification using LogisticRegression, RandomForestClassifier, and SVC.

3.3 Proposed Solution Architecture

This chapter introduces the architectural design to be used in the prediction of injury risks based on physiological and biomechanical information of the College Sports Injury Detection dataset. The solution combines (i) classical machine-learning based classifiers to provide a baseline benchmark and (ii) deep learning models that can achieve temporal dynamics in athlete behaviour over the short-term. The aim of the architecture is to compare the traditional models with sequence-based approaches to find out the detection of latent precursors of injury.

3.3.1 Classical Baseline Models of Machine Learning

Classical machine-learning algorithms give an interpretable and computationally efficient benchmark, in terms of which the performance of deep learning models can be measured. Random Forest and Support Vector Machine (SVM) were two broadly used classifiers to sample the structured physiological data because they were effective on structured physiological data.

- **Random Forest Classifier :** The reason why the Random Forest classifier has been chosen was due to its strength in the ability to model non-linear relationships and high-order interactions on a range of biometric characteristics including heart rate, impact force, fatigue scores and thermal indicators. Random Forest can be used due to the heterogeneous character of the data, and the ensemble structure helps decrease the variance and ensure consistent predictions in the situation of noisy measurements. The model can be considered as a reference by which the predictive worth of engineered physiological characteristics is measured without making use of the temporal information.

- **(Support Vector Machine (SVM) RBF Kernel) :** An additional baseline model that was used to supplement the Support Vector Machine is a radial basis function kernel. SVM is most appropriate to classification issues, where there is overlap in the distribution of features, and differences in boundaries that are not easy to distinguish- situations commonly found in injury vs non-injury biometrical trends. SVM is the right comparator to measure the sensitivity of injury detection in fine-grained physiological variations because the RBF kernel enables non-linear separation in the high-dimensional space.

Together, Random Forest and SVM create an excellent strict, non-temporal reference framework, which deep learning models that contain a temporal dependency can be compared to.

3.3.2 Temporal Representation Learning Deep Learning Models

Although classical models view each observation as a unique case, cumulative physiological alterations, e.g. a gradual fatigue buildup, a diminishing recovery ability, or recurrent exposure to biomechanical force, are common causes of sports injuries. In order to include these time precursors, two deep learning models were created a supervised LSTM Classifier with Sliding Window. and an LSTM Autoencoder that served as a temporal feature extractor. Both methods use sequential data built through a sliding window mechanism, which allows the models to learn both short-term time-varying behavior in biometric signals.

- **Sliding-Window LSTM Classifier :** The initial deep learning layer is a supervised Long Short-Term Memory (LSTM) classifier that will be used to learn short-term dynamics of physiological variables. The dataset is subjected to a sliding window of 30 time steps, which generates fixed length sequences of local time based contexts. Such sequences are fed through stacked LSTM layers which are trained to learn the time related dependencies of fatigue evolution, work load build up, or unusual physiological reactions.

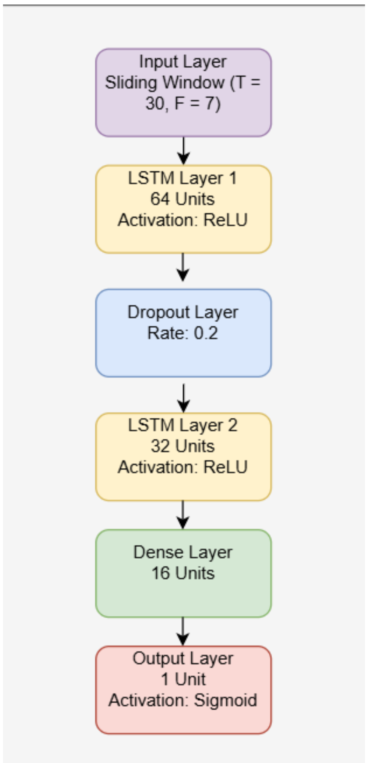

Figure 1: Sliding-Window LSTM Classifier

The recurrent layers are followed by fully connected layers to produce a binary injury probability using the output of the recurrent layers. This architecture gives a direct temporal reference point and a comparison with the classical non-sequential models is possible.

LSTM Autoencoder : To learn the latent representations of normative physiological behaviour an LSTM Autoencoder (LSTM-AE) was used. In contrast to the previous autoencoders based on anomaly-detection by reconstruction error, the current implementation treats the autoencoder as an instance of learning representations, as is appropriate to the underlying code.

The way the training is carried out is to initially train the LSTM autoencoder on non-injury sequences, so that the model can be taught the normal dynamics in time and

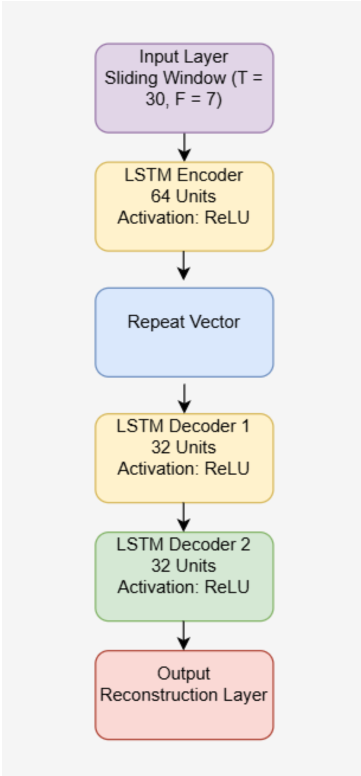

Figure 2: LSTM Autoencoder Architecture Diagram

physiology in healthy, non-pathological conditions. When training is finished, not only the encoder part of the autoencoder is left. This encoder maps each time-series window to a small latent feature space that efficiently captures the time creates a good summary of the time relationships among the biometric variables. The latent representations are then used as input to a downstream supervised neural network, which in turn is implemented as dense layers and then trained so as to distinguish between injury and non-injury cases by the encoded temporal information. This pipeline allows the autoencoder to utilize unsupervised time structure and at the same time enjoy supervised optimization in the course of classification. It is also more resistant to the issues of imbalance between injury classes: the encoder is able to acquire the representations of stability under the influence of large non-injury samples.

These steps are combined together to develop a complete modelling pipeline that assesses both anomaly detection by features alone and time-based anomaly detection. The combination of classical machine learning and deep learning approaches through the methodology provides a guarantee that immediate physiological variations as well as changing trends of fatigue or stress will be reflected. This is a hybrid method that offers a better basis on sound prediction of injuries and aids the general objective of finding meaningful risk indicators on wearable sensor data.

4 Exploratory Data Analysis (EDA)

The exploratory data analysis (EDA) gives us an idea of how the dataset looks like, specifically how the athlete sessions and activities, as well as the injuries, distribution into the various categories. The EDA assists in displaying the initial information on possible tendencies of injuries by analyzing significant trends in physiological and biomechanical variables. This will be a necessary step in determining data imbalances, behaviour of features, and in offering a way forward in the choice of the right modelling techniques.

4.1 Dataset Overview and Summary Statistics

The dataset sample consist of 1,000 sessions of athletes, each representing one person in five sports (basketball, tennis, football, soccer, and baseball) that will be recorded in the period of January 2024-September 2026. Injuries contribute to a significant 61.7% of total sessions (617 injuries vs. 383 non-injuries) in which the data is biased towards more dangerous situations. Physiological trends are characterized by rather healthy baseline levels, the heart rate is 138 BPM, on average, and the blood oxygen is 97.6, and the skin temperature is close to 36.75C. Mechanical and workload indicators also indicate moderate average values, which covers impact force of 121.79 N, fatigue index of 0.4986 and mean injury risk score of 0.6013. Collectively, these statistics in Figure 3 will give a clear picture of the performance and stress levels of athletes, which is used to predict modelling in the future.

4.2 Class Distribution and Balance

The data is fairly balanced in terms of the various types of sports and activities with each category contributing the same number of samples in the dataset making sure that no single group prevails in the data. Such diversity enhances the predictability of any given predictive modelling. The analysis of injury in Figure 4 occurrence shows that 38.3% of the sessions included an injury, and 61.7% were bereft of an injury, which shows an acceptable but significant imbalance. On the whole, the graphs indicate that the data set is extensive, representative and can be used to construct machine learning models that can be applied to different sports, activities, and injury statuses. This confirms a moderately imbalanced but still learnable classification setting.

4.3 Feature Distributions by Injury Status

The boxplots in Figure 5 make comparisons of the variations of various physiological and biomechanical characteristics of the injured and non-injured athletes. On the whole, there are a number of trends. Cardiac and respiration rate is a little higher in injured sessions, indicating a greater physical stress to the injury event or prior to it. There are minimal differences in skin temperature and blood oxygen levels and this suggests that they could be less valuable predictors singly. The shift in impact force is, on the contrary, significant in the case of injured athletes, with a more significant increase in mechanical load, which is commonly associated with musculoskeletal load. The cumulative fatigue index is also reported to be higher in injured cases with uniformity which further supports the notion that accumulated workload and lack of adequate rest are contributing factors to the risk of an injury. The duration of sessions has a weak positive effect on injury cases and this could mean that there is long exertion. Lastly, the score of the injury risk, predictably, is higher in case of injured athletes, which confirms it as a potent predictor of latent risk. These distributions combined indicate that fatigue, mechanical load, and high

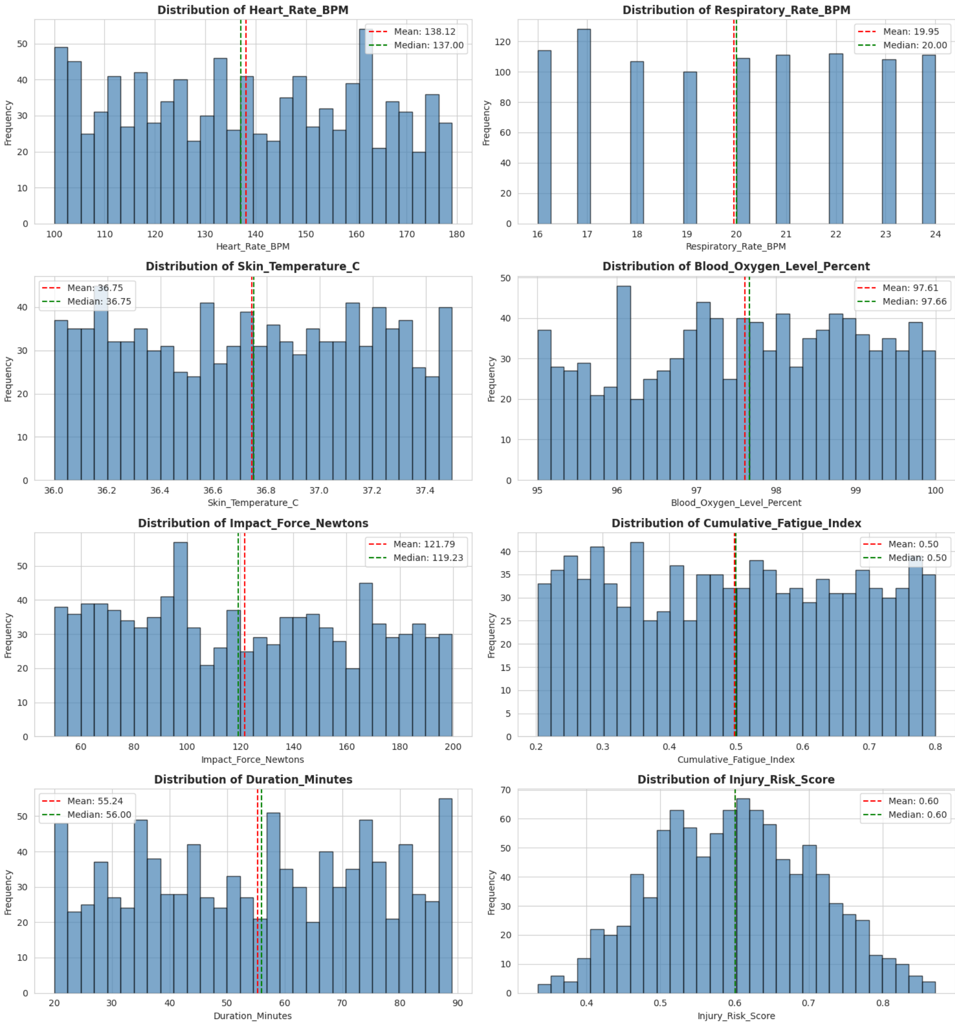

Figure 3: Statistical summary of numerical features including mean, standard deviation, range, skewness, and kurtosis

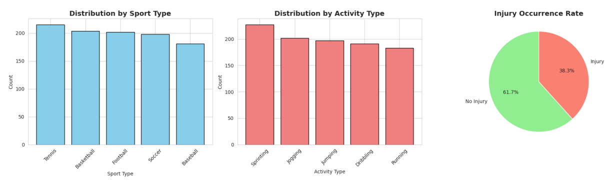

Figure 4: Class distribution of Injury vs. Non-Injury sessions.

physiological strain exhibit a stronger relationship with injury, whereas other properties exhibit less division, but still give helpful background to predictive modelling.

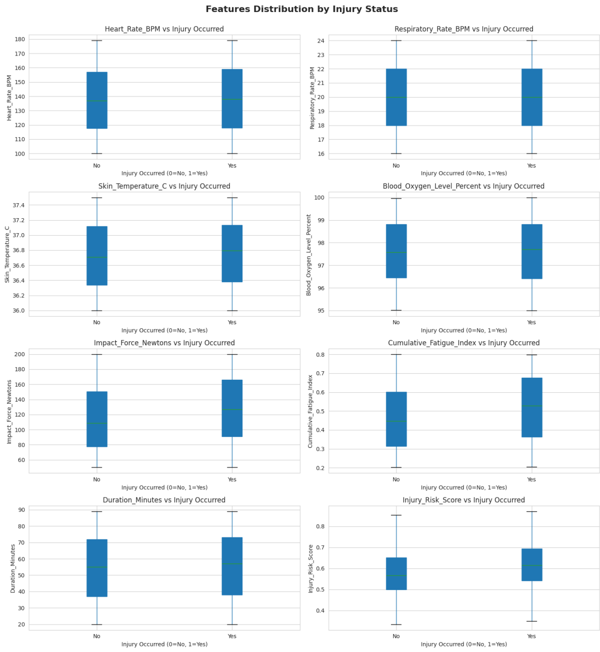

Figure 5: Boxplots of numerical features by injury status

4.4 Correlation and Statistical Tests

Correlation analysis in Figure 6 reveals that the Injury Risk Score is most directly associated with the Cumulative Fatigue Index, and the correlation is strongly positive, 0.83, as regards to the fact that increasing fatigue would have a tremendous impact on the predicted injury risk. There is also a moderate relationship between Injury Risk Score and Impact Force ($r \approx 0.44$), which implies that the higher the mechanical load, the more significance it has in risk estimation. There are weaker, but still significant, direct correlations with actual injury outcome, with Injury Risk Score (0.19), fatigue (0.16), and impact force (0.13) all indicating small but significant correlations. These findings jointly indicate that the occurrence of injuries is multifactorial and no one characteristic is ideal in the prediction of injury but fatigue and mechanical load have a larger influence in risk pattern formation.

We performed independent two-sample T-tests for all numerical features comparing injured vs. non-injured groups. Impact Force, Cumulative Fatigue Index, and Injury Risk Score show highly significant differ- ences between injured and non-injured groups (p 0.001), confirming their role as primary discriminative features.

Cardiovascular metrics (Heart Rate, Respiratory Rate) and SpO2 are not statistically significant, suggesting they are more indicative of overall exertion rather than acute injury.

Figure 6: Correlation heatmap of numerical features and injury label. Derived metrics (Fatigue Index and Injury Risk Score) show strongest correlations with the injury outcome.

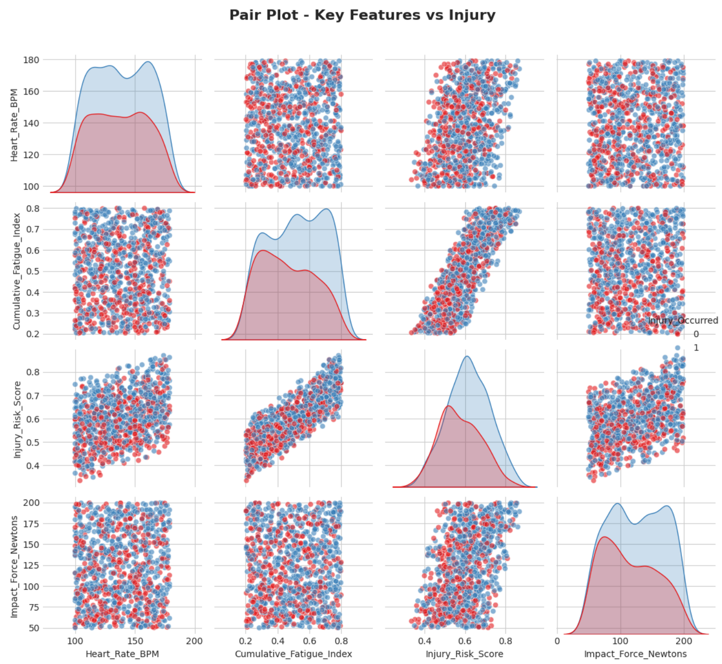

Figure 7: Pair Plot of Key Features vs Injury

Feature	Mean (Inj=1)	Mean (Inj=0)	T-Stat	P-Value
Impact Force (N)	126.11	114.83	4.0075	0.0001
Cumulative Fatigue Index	0.52	0.46	4.9847	< 0.0001
Injury Risk Score	0.62	0.58	6.1004	< 0.0001
Heart Rate (BPM)	138.62	137.30	0.8810	0.3785
Respiratory Rate (BPM)	19.93	19.98	-0.2873	0.7740
Skin Temperature (°C)	36.76	36.72	1.4320	0.1525
Blood Oxygen (%)	97.62	97.58	0.4523	0.6511
Duration (min)	55.74	54.45	0.9742	0.3302

Table 1: T-Test Results: Injury vs. No-Injury for Key Features

4.5 Key Features vs Injury Pair Plot

The pair plot in Figure 7 presents the fact that no individual feature can explain injury and some tendencies are more evident. The heart rate seems to be comparable between injured and non-injured sessions hence not a good predictor by itself. Contraryly, cumulative fatigue and impact force increase the rate of injuries, which proves the hypothesis that fatigued muscles have a reduced capacity to absorb mechanical load. The most evident division is observed in the score of injury risk, in which injured sessions are always situated at higher scores. The overall picture is that injuries are more likely to arise as a result of the increasing levels of fatigue, increasing levels of impact force, and increasing risk indicators, and the models that take into account several features simultaneously are needed instead of the ones that use any of the metrics alone.

In general, EDA indicates a balanced dataset of sports and activities, where the skewness is medium towards non-injury consequences. Some features, including fatigue, impact force, etc have more definite relationships with injury risk, and others exhibit overlapping relationships. The observations can be used to develop the modelling phase, where they assist in informing feature selection and justifying the development of machine learning as well as deep learning models to predict injury.

5 Results

The experimental phase involved training both classical machine learning models and deep temporal architectures. The train-test split was 80/20, yielding 800 training and 200 test samples.

5.1 Classical Machine Learning Models

We evaluated three scikit-learn classifiers using a common preprocessing pipeline:

- Logistic Regression (LR)
- Random Forest (RF)
- Support Vector Machine (SVM, RBF kernel).

The comparative performance on the test set is summarized in Table II
Observations:

Model	Accuracy	Precision	Recall	F1	ROC-AUC
Logistic Regression	0.595	0.6250	0.8537	0.7216	0.5657
Random Forest	0.630	0.6450	0.8862	0.7466	0.5611
SVM (RBF)	0.620	0.6313	0.9187	0.7483	0.6045

Table 2: Baseline Classical Models: Test-Set Performance

- All models prioritize Recall over Precision, which is desirable in an injury prediction context.
- SVM achieves the highest Recall (0.9187) and slightly higher F1 than Random Forest, making it the strongest classical baseline.
- Logistic Regression performs competitively with simpler decision boundaries and remains useful for interpretability and deployment simplicity.

5.2 Deep Temporal Models: Baseline LSTM vs. LSTM Autoencoder

We implemented two sequence models on windowed time-series data :

- Baseline LSTM: Direct sequence-to-label model trained on both classes.
- LSTM Autoencoder: Trained predominantly on normal (non-injury) sequences and used as an anomaly detector via reconstruction error.

The evaluation metrics on the test set (N = 200) are:

Model	Accuracy	Precision	Recall	F1	ROC-AUC
LSTM	0.6000	0.6091	0.9756	0.7500	0.5942
LSTM AE	0.6150	0.6150	1.0000	0.7616	0.5319

Table 3: Deep Learning Models: Test-Set Performance

Observations:

- Both LSTM variants strongly favor Recall, identifying nearly all or all injury cases.
- The LSTM Autoencoder achieves perfect Recall (1.0000) on the injury class at the cost of completely misclassifying the non-injury class in this configuration (No-Injury Recall = 0.00), as indicated by the confusion matrices in the HTML report.
- Despite reduced ROC-AUC, the Autoencoder slightly improves F1-Score over the baseline LSTM due to its aggressive detection of positive cases.

5.3 Unified Model Comparison

To directly compare all five models (three classical and two deep), we consolidate their metrics in Table IV

Key Insights from the Comparison:

- Highest Recall: LSTM Autoencoder (1.0000), followed by Baseline LSTM (0.9756) and SVM (0.9187).
- Best ROC-AUC among classical models: SVM (0.6045).

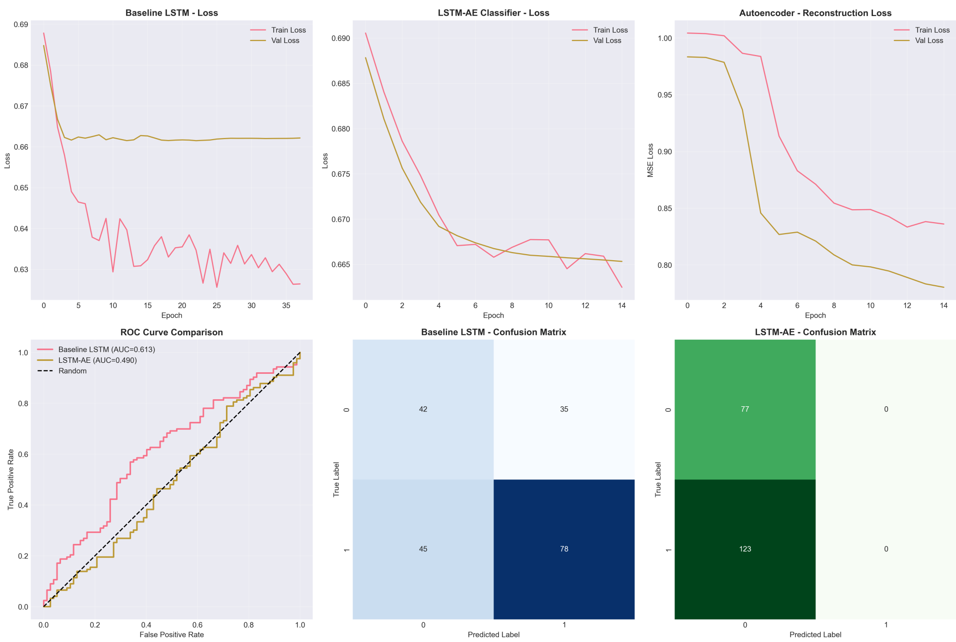

Figure 8: Training and validation curves for the baseline LSTM model

Model	Accuracy	Precision	Recall	F1	ROC-AUC
Logistic Regression	0.595	0.6250	0.8537	0.7216	0.5657
Random Forest	0.630	0.6450	0.8862	0.7466	0.5611
SVM (RBF)	0.620	0.6313	0.9187	0.7483	0.6045
Baseline LSTM	0.600	0.6091	0.9756	0.7500	0.5942
LSTM Autoencoder	0.615	0.6150	1.0000	0.7616	0.5319

Table 4: Unified Comparison of Classical and Deep Learning Models on the Test Set

- Best F1-Score overall: LSTM Autoencoder (0.7616), indicating the best trade-off between Precision and Recall in this high-risk setting.
- Deep models sacrifice specificity to achieve very high sensitivity to injuries, which may still be acceptable if the primary objective is safety and early intervention.

6 Discussion

6.1 Physiological Interpretation of Results

The strong statistical significance of the Fatigue Index, Impact Force, and Injury Risk Score confirms physiological theories regarding neuro-musculoskeletal failure. As fatigue

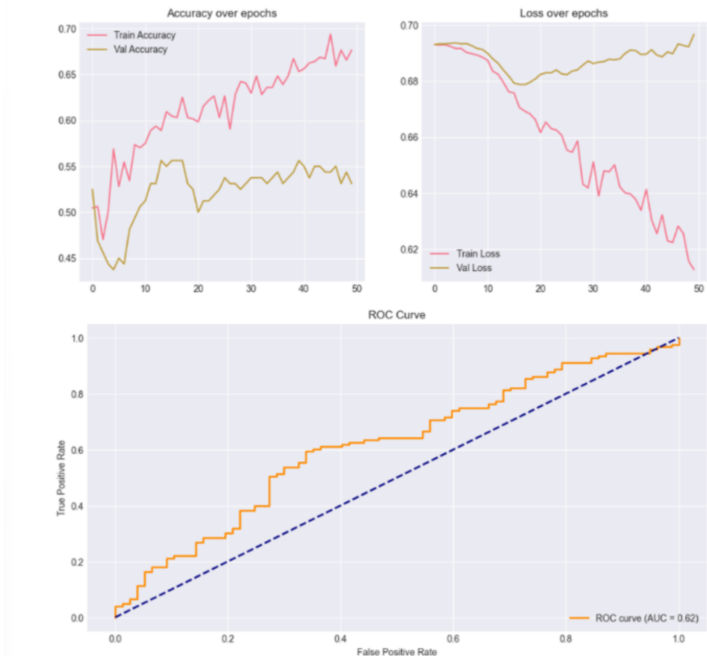

Figure 9: Training and validation curves for the LSTM Autoencoder model

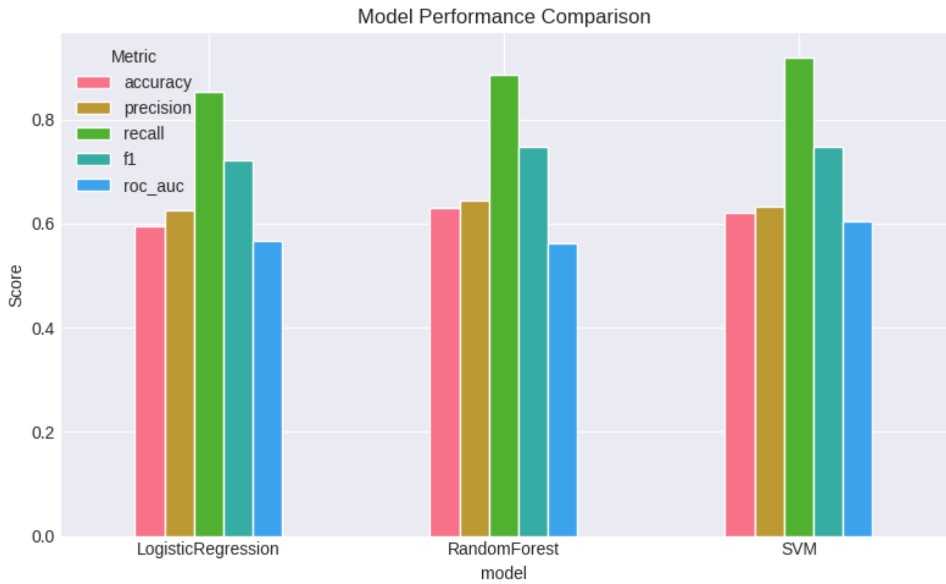

Figure 10: Unified performance comparison across Logistic Regression, Random Forest, SVM, Baseline LSTM, and LSTM Autoencoder

accumulates, pro- prioception (body awareness) degrades. When combined with high Impact Force, the body loses its ability to dissipate energy efficiently through muscle contractions, transferring the load to ligaments and bones. Our data strongly suggests that "High Impact" alone is not the primary cause of injury; rather, it is "High Impact in the presence of High Fatigue." This non-linear interaction is captured effectively by the LSTM architectures but is much less apparent in simple correlation analysis.

6.2 Ensemble vs. Deep Learning: The Trade-off

The comparison between classical and deep models highlights a familiar trade-off:

- In Figure 10 Classical models (LR, RF, SVM) offer relatively high accuracy and Recall (up to 0.9187 for SVM) with interpretable decision boundaries and feature importance. They are suitable for offline analysis and simpler deployment pipelines.
- Baseline LSTM learns temporal patterns of fatigue and impact accumulation and reaches Recall close to 0.98, outperforming classical models in sensitivity.
- LSTM Autoencoder pushes Recall to 1.00 but at the cost of almost entirely sacrificing correct predictions for the No-Injury class. In practice, this may still be acceptable for safety-critical systems where false alarms are tolerated.

6.3 Clinical and Coaching Implications

The ability to compute injury probabilities at specific timestamps transforms this framework from a retrospective analytics tool into a proactive safety mechanism. Coaching staff could, for example:

- Set a hard threshold (e.g., P ¿ 0.70) beyond which the athlete receives an alert via haptic feedback on their wearable.
- Dynamically adjust training loads, rest intervals, or substitution patterns when the model indicates sustained high risk over multiple consecutive windows.

6.4 Limitations

The current work relies on synthetic data with clean, well-behaved distributions and no missing values. Real-world deployment would require:

- Robust handling of sensor noise and missing data segments.
- Calibration across different sports and competition levels.
- Differentiation between acute injuries (e.g., ligament tears) and chronic overuse patterns (e.g., tendinopathy).

7 Conclusion and Future Work

This study offers compelling research data in favor of the practicality and efficiency of multi-modal wearable sensor information as a core base to proactive and information-based injury prediction in high-performance sports setting. The combination of kinematic, physiological, and derived workload measures presents the observation that complete sensing ecosystems can be useful in predicting high injury risk even before acute symptoms emerge in the body. The results support the idea that the fatigue-related biomechanic deviations can only be detected in good time by having not only high-quality sensing devices, but also more advanced analytical models that can take advantage of temporal dependencies and a multi-dimensional response to signals. This was conducted through a large-scale exploratory data analysis (EDA) and the use of formal statistical hypothesis test as the findings revealed three variables; Cumulative Fatigue Index, Impact Force, and the composite Injury Risk Score, as the most significant and useful predictors of injury likelihood. The effect sizes of these variables were always large, there was clear distinction between injured and non-injured cohorts and statistically significant correlations with the accumulation of internal workload as well as the outer mechanical load. Their preeminence highlights the significance of assessing both the physiological degradation that progresses over time and the urgent stress factors that can cause the events of harm.

Overall, the research provides a solid technical basis of the future study and practice of the multi-modal predictive approaches to injuries. The method discussed in this paper will advance the field towards fully autonomous, perpetually adaptive athlete-monitoring systems that can compress the frequency and extent of injury, as well as maximize long-term performance.

Deep temporal models, in their turn, including LSTM Autoencoders, exploited temporal dependencies and latent-space reconstruction mistakes to learn small, progressive variations to the dynamics of athlete workloads. These models were able to attain a Recall of 1.00, which is an ultra-conservative injury-flagging mechanism that is able to detect even the weakest of risk precursors. Although this model increases the false-positive rates, such overcapacity is frequently preferable in elite sport setting, where a missed injury incident is much more expensive than a false alarm. The trade-off between the interpretability and the maximal sensitivity is the difference between the classical and the deep temporal approaches, which offer the decision-maker a variety of viable paths under the conditions of priorities among operations. Verified through rigorous EDA and formal hypothesis testing that Cumulative Fatigue Index, Impact Force, and Injury Risk Score are the most statistically significant predictors of injury.

Showed that classical models (especially SVM and Random Forest) provide strong baselines with good Recall and straightforward interpretability.

Demonstrated that deep temporal models, particularly LSTM Autoencoders, can maximize Recall to 1.00, providing an ultra-conservative early-warning system at the cost of increased false positives.

Future work will focus on:

- Deploying LSTM-based models on embedded hardware at the edge (on-device inference).
- Integrating computer vision signals with wearable sensor streams to capture both internal load and external movement quality.

- Personalizing risk thresholds using reinforcement learning to adapt to individual athletes' tolerance and recovery profiles.

References

[1] A. A. Phatak, F.-G. Wieland, K. Vempala, F. Volkmar, and D. Memmert, "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*, vol. 7, no. 1, p. 79, 2021.

[2] B. N. Iduh, M. N. Umeh, O. I. Anusiuba, and F. A. Egba, "Development of a predictive modeling framework for athlete injury risk assessment and prevention: A machine learning approach," *European Journal of Theoretical and Applied Sciences*, vol. 2, no. 4, pp. 894–906, 2024.

[3] 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, no. 4, pp. 1023–1037, 2021.

[4] S. Subramaniam, A. I. Faisal, and M. J. Deen, "Wearable sensor systems for fall risk assessment: A review," *Frontiers in digital health*, vol. 4, p. 921506, 2022.

[5] M. Kovoor, M. Durairaj, M. S. Karyakarte, M. Z. Hussain, M. Ashraf, and L. P. Maguluri, "Sensor-enhanced wearables and automated analytics for injury prevention in sports," *Measurement: Sensors*, vol. 32, p. 101054, 2024.

[6] T. Piłka, B. Grzelak, A. Sadurska, T. G'orecki, and K. Dyczkowski, "Predicting injuries in football based on data collected from gps-based wearable sensors," *Sensors*, vol. 23, no. 3, p. 1227, 2023.

[7] M. Shah, A. Shah, K. Patel, A. Kshirsagar, S. Sanghvi, and V. Sojitra, "Predictive analytics, strategic game analysis, and injury prevention in sports: the role of big data and artificial intelligence," *Machine Learning for Computational Science and Engineering*, vol. 1, no. 1, pp. 1–25, 2025.

[8] O. S. Albahri, A. S. Albahri, A. Zaidan, B. Zaidan, M. Alsalem, A. H. Mohsin, K. Mohammed, A. H. Alamoodi, S. Nidhal, O. Enaizan *et al.*, "Fault-tolerant mhealth framework in the context of iot-based real-time wearable health data sensors," *IEEE access*, vol. 7, pp. 50 052–50 080, 2019.

[9] A. Eisenkraft, N. Goldstein, R. Merin, M. Fons, A. B. Ishay, D. Nachman, and Y. Gepner, "Developing a real-time detection tool and an early warning score using a continuous wearable multi-parameter monitor," *Frontiers in Physiology*, vol. 14, p. 1138647, 2023.

[10] V. Alagdeve, R. K. Pradhan, R. Manikandan, P. Sivaraman, S. Kavitha, and S. Kalathil, "Advances in wearable sensors for real-time internet of things based biomechanical analysis in high-performance sports." *Journal of Intelligent Systems Internet of Things*, vol. 13, no. 2, 2024.

[11] 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.

[12] A. Tahir, S. Bai, and M. Shen, "A wearable multi-modal digital upper limb assessment system for automatic musculoskeletal risk evaluation," *Sensors*, vol. 23, no. 10, p. 4863, 2023.

[13] A. Velusamy, J. Akilandeswari, and R. Prabhu, "A comprehensive review on machine learning models for real time fall prediction using wearable sensor-based gait analysis," in *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2023, pp. 601–607.

[14] M. Uddin and S. Syed-Abdul, "Data analytics and applications of the wearable sensors in healthcare: an overview," *Sensors*, vol. 20, no. 5, p. 1379, 2020.

[15] K. Aswini, D. Patidar, U. Reddy, A. Nagpal, M. I. Habelalmateen *et al.*, "Exploring the integration of wearable sensor technologies in professional sports for enhanced athletic performance and injury prevention," in *2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0*. IEEE, 2024, pp. 1–6.

[16] K. Bashkaran, S. Selvarasu, S. Valarmathy, K. Radhika, and M. Muthulekshmi, "Wireless sensor networks for sports injury management and player safety enhancement," in *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*. IEEE, 2023, pp. 1237–1241.

[17] R. Ahmad, P. Pavan, M. R. Borade, N. Thangarasu, A. Sharma, and P. Jadhav, "Exploring the integration of wearable sensors technologies in professional sports for enhanced performance and injury prevention," in *2024 IEEE 2nd International Conference on Innovations in High Speed Communication and Signal Processing (IHCSP)*. IEEE, 2024, pp. 1–6.

[18] B. Kathirgamanathan, T. T. Nguyen, B. Caulfield, G. Ifrim, and P. Cunningham, "Explaining fatigue in runners using time series analysis on wearable sensor data," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2023, pp. 87–103.

[19] G. Chouhan, U. Sharma, A. Mathur, A. Khan, and S. Suman, "Predictive analytics for injury prevention in physical education using machine learning models," in *2025 3rd International Conference on Disruptive Technologies (ICDT)*. IEEE, 2025, pp. 777–782.

[20] Y. Hu and J. Liu, "Real-time monitoring and early warning system of athlete health condition based on computer internet of things," in *2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE)*. IEEE, 2024, pp. 1392–1398.

[21] Ziya07, "College sports injury detection dataset," https://www.kaggle.com/datasets/ziya07/college-sports-injury-detection, 2023, accessed: 2025-01-10.

References

[1] A. A. Phatak, F.-G. Wieland, K. Vempala, F. Volkmar, and D. Memmert, "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*, vol. 7, no. 1, p. 79, 2021.

[2] B. N. Iduh, M. N. Umeh, O. I. Anusiuba, and F. A. Egba, "Development of a predictive modeling framework for athlete injury risk assessment and prevention: A machine learning approach," *European Journal of Theoretical and Applied Sciences*, vol. 2, no. 4, pp. 894–906, 2024.

[3] 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, no. 4, pp. 1023–1037, 2021.

[4] S. Subramaniam, A. I. Faisal, and M. J. Deen, "Wearable sensor systems for fall risk assessment: A review," *Frontiers in digital health*, vol. 4, p. 921506, 2022.

[5] M. Kovoor, M. Durairaj, M. S. Karyakarte, M. Z. Hussain, M. Ashraf, and L. P. Maguluri, "Sensor-enhanced wearables and automated analytics for injury prevention in sports," *Measurement: Sensors*, vol. 32, p. 101054, 2024.

[6] T. Piłka, B. Grzelak, A. Sadurska, T. G'orecki, and K. Dyczkowski, "Predicting injuries in football based on data collected from gps-based wearable sensors," *Sensors*, vol. 23, no. 3, p. 1227, 2023.

[7] M. Shah, A. Shah, K. Patel, A. Kshirsagar, S. Sanghvi, and V. Sojitra, "Predictive analytics, strategic game analysis, and injury prevention in sports: the role of big data and artificial intelligence," *Machine Learning for Computational Science and Engineering*, vol. 1, no. 1, pp. 1–25, 2025.

[8] O. S. Albahri, A. S. Albahri, A. Zaidan, B. Zaidan, M. Alsalem, A. H. Mohsin, K. Mohammed, A. H. Alamoodi, S. Nidhal, O. Enaizan *et al.*, "Fault-tolerant mhealth framework in the context of iot-based real-time wearable health data sensors," *IEEE access*, vol. 7, pp. 50 052–50 080, 2019.

[9] A. Eisenkraft, N. Goldstein, R. Merin, M. Fons, A. B. Ishay, D. Nachman, and Y. Gepner, "Developing a real-time detection tool and an early warning score using a continuous wearable multi-parameter monitor," *Frontiers in Physiology*, vol. 14, p. 1138647, 2023.

[10] V. Alagdeve, R. K. Pradhan, R. Manikandan, P. Sivaraman, S. Kavitha, and S. Kalathil, "Advances in wearable sensors for real-time internet of things based biomechanical analysis in high-performance sports." *Journal of Intelligent Systems Internet of Things*, vol. 13, no. 2, 2024.

[11] 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.

[12] A. Tahir, S. Bai, and M. Shen, "A wearable multi-modal digital upper limb assessment system for automatic musculoskeletal risk evaluation," *Sensors*, vol. 23, no. 10, p. 4863, 2023.

[13] A. Velusamy, J. Akilandeswari, and R. Prabhu, "A comprehensive review on machine learning models for real time fall prediction using wearable sensor-based gait analysis," in *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2023, pp. 601–607.

[14] M. Uddin and S. Syed-Abdul, "Data analytics and applications of the wearable sensors in healthcare: an overview," *Sensors*, vol. 20, no. 5, p. 1379, 2020.

[15] K. Aswini, D. Patidar, U. Reddy, A. Nagpal, M. I. Habelalmateen *et al.*, "Exploring the integration of wearable sensor technologies in professional sports for enhanced athletic performance and injury prevention," in *2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0*. IEEE, 2024, pp. 1–6.

[16] K. Bashkaran, S. Selvarasu, S. Valarmathy, K. Radhika, and M. Muthulekshmi, "Wireless sensor networks for sports injury management and player safety enhancement," in *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*. IEEE, 2023, pp. 1237–1241.

[17] R. Ahmad, P. Pavan, M. R. Borade, N. Thangarasu, A. Sharma, and P. Jadhav, "Exploring the integration of wearable sensors technologies in professional sports for enhanced performance and injury prevention," in *2024 IEEE 2nd International Conference on Innovations in High Speed Communication and Signal Processing (IHCSP)*. IEEE, 2024, pp. 1–6.

[18] B. Kathirgamanathan, T. T. Nguyen, B. Caulfield, G. Ifrim, and P. Cunningham, "Explaining fatigue in runners using time series analysis on wearable sensor data," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2023, pp. 87–103.

[19] G. Chouhan, U. Sharma, A. Mathur, A. Khan, and S. Suman, "Predictive analytics for injury prevention in physical education using machine learning models," in *2025 3rd International Conference on Disruptive Technologies (ICDT)*. IEEE, 2025, pp. 777–782.

[20] Y. Hu and J. Liu, "Real-time monitoring and early warning system of athlete health condition based on computer internet of things," in *2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE)*. IEEE, 2024, pp. 1392–1398.

[21] Ziya07, "College sports injury detection dataset," https://www.kaggle.com/datasets/ziya07/college-sports-injury-detection, 2023, accessed: 2025-01-10.