Fatigue analysis with artificial intelligence using heart rate, accelerometers, and gyroscopes

An Ngo[1], Ansh Lalwani[1], Jaafar Ben Khaled[1], Jaiden Gill[1] and Varshan Kumar[1]

[1] Department of Computer Science, San Jose State University, San Jose, United States of America

Keywords: fatigue, heart rate, heart rate variability, acceleration, gyroscope.

Abstract. In this document, we explore how we can detect and model fatigue from an athlete's heart rate, triaxial accelerometers, and their triaxial gyroscopic readings. We determine a baseline method for detecting fatigue based on external research papers and sources, of which are cited below. Using this baseline, we develop and train a model that can both classify the type of activity an athlete is doing, as well as whether or not they may be experiencing fatigue.

Contents

1 Introduction

Fatigue is a critical factor that has an impact on athlete performance, injury risk, mental health, and long-term health. As training intensities increase and competition becomes more demanding, accurately understanding and modeling fatigue has become increasingly important for athletes, coaches, and sports scientists. Wearable technologies such as heart rate monitors, triaxial accelerometers, and gyroscopes provide a way for non-invasive monitoring of physiological, biomechanical, and spatial signals. These data streams present oppportunities to process and visualize both physical exertion as well as early signs of fatigue in real-world athletic contexts.

In this work, we investigate how multisensor data can be leveraged to identify and quantify fatigue in athletes. Building on findings and standards from prior research in heart rate variability

analysis, human activity recognition (HAR), and fatigue detection, we establish a baseline for determining fatigue informed by external sources. This baseline serves as the foundation for developing and training an artificial intelligence model that can perform two tasks: firstly to classify the athlete's current activity based on their motion signals, and secondly to assess whether the athlete is exhibiting signs of fatigue.

By integrating both physiological and motion-based indicators, our approach aims to improve the reliability of fatigue detection and analysis systems, and contribute towards a more effective and stable athlete monitoring framework. Ultimately, this work aims to demonstrate how a combination of heart rate, accelerometry, and gyroscopic signals can support more adaptive and data-driven approaches and strategies for training, performance optimization, and fatigue management.

2 Literature Review

Determining fatigue via real time data is a topic of interest for many groups in developing and modeling better and more deterministic models and methods. The following sources were all relevant to our area of review:

[1] Identification of Runner Fatigue Stages Using Inertial Sensors and Deep Learning - Source: Frontiers in Bioengineering and Biotechnology (2023) - Details: IMUs were placed on runners to capture accelerometer and gyroscope signals during prolonged running. Deep learning models (CNNs, LSTMs) classified fatigue stages, identifying biomechanical changes such as stride length reduction, altered acceleration profiles, and variability in joint angles. The study highlighted how temporal modeling improved accuracy, showing fatigue as a gradual process rather than a binary state.

[2] Machine Learning-Based Prediction of Running-Induced Fatigue - Source: University of Twente MSc Thesis (2024) - Details: Integrated IMU data, heart rate, and smartwatch metrics to predict fatigue in outdoor running. Random Forest and SVM models were tested, with Random Forest showing the strongest performance. Feature importance analysis revealed HRV and stride variability as key predictors. The thesis emphasized multimodal integration and careful feature selection to reduce noise and improve generalization across athletes.

[3] Monitoring Fatigue State with Heart Rate-Based and Subjective Methods - Source: European Journal of Sport Science (2024) - Details: Compared HRV metrics (RMSSD, LF/HF ratio) with subjective fatigue ratings in recreational runners. HRV consistently declined with training load, while subjective ratings varied widely. The study concluded that HRV provides a more objective measure of accumulated fatigue, particularly useful in longitudinal monitoring of athletes across training cycles.

[4] Fatigue Estimation Using Wearable Devices and Virtual Instrumentation - Source: Lecture Notes in Networks and Systems (2024) - Details: Explored fatigue detection using consumer-grade wearables such as smartwatches and wristbands. Heart rate and accelerometer signals were processed with virtual instrumentation software, and machine learning models classified fatigue states. Despite lower sensor precision compared to lab-grade devices, results showed meaningful detection, suggesting consumer wearables can provide scalable monitoring solutions.

[5] Monitoring Fatigue During Intermittent Exercise with Accelerometer-Derived Metrics - Source: Frontiers in Physiology (2019) - Details: Accelerometer data was analyzed during intermittent exercise (sprints, rest cycles). Metrics such as acceleration magnitude, variability, and jerk (rate of change of acceleration) correlated strongly with fatigue onset. The study validated accelerometer-derived features as reliable markers of neuromuscular fatigue, particularly in sports with repeated high-intensity efforts.

[6] Fatigue Monitoring Through Wearables: A State-of-the-Art Review - Source: Frontiers in

Physiology (2021) - Details: Systematic review of wearable fatigue monitoring approaches. Categorized methods into physiological (heart rate, HRV), biomechanical (accelerometer, gyroscope), and hybrid (sensor fusion). Discussed challenges such as sensor placement, inter-individual variability, and personalization needs. The review also highlighted gaps in real-time monitoring and the need for standardized protocols.

[7] Fatigue State Prediction of Athletes Based on Multi-Source Sensors - Source: IEEE Xplore (2024) - Details: Proposed a multi-sensor fusion framework combining heart rate, IMU, and gyroscope data. Deep learning models (LSTMs) achieved real-time fatigue prediction with higher accuracy than single-sensor approaches. The study emphasized temporal modeling to capture fatigue progression, showing that fatigue signatures evolve across multiple modalities simultaneously.

[8]Improving Fatigue Detection with Feature Engineering on Accelerometer Data - Source: International Journal of Affective and Behavioral Computing (2024) - Details: Applied advanced feature engineering to accelerometer signals, including frequency-domain features (spectral entropy, dominant frequency) and nonlinear measures (fractal dimension). Engineered features significantly improved classification performance compared to raw signals. The study demonstrated that careful preprocessing and feature design can enhance fatigue detection models.

[9]Wearable Sensor-Based Fatigue Detection in Elite Soccer Players - Source: Journal of Sports Sciences (2023) - Details: Elite soccer players were monitored using chest-strap heart rate sensors and IMUs on the lower limbs. HRV, stride variability, and gyroscope-based joint rotation metrics were analyzed. Fatigue manifested as reduced HRV, increased stride irregularity, and altered angular velocity in hip and knee joints. Machine learning classifiers achieved strong accuracy in distinguishing fatigued vs. non-fatigued states, even in dynamic team sport contexts.

[10] Real-Time Fatigue Monitoring Using Multi-Sensor Fusion and Deep Learning in Endurance Athletes - Source: Sensors (2022) - Details: Proposed a real-time fatigue monitoring system for endurance athletes using wearable sensors (heart rate monitors, accelerometers, gyroscopes). Data were processed through a CNN-LSTM hybrid model to capture spatial features and temporal dynamics. The system achieved high accuracy in predicting fatigue onset during prolonged cycling and running, demonstrating feasibility for continuous monitoring.

[11] Heart Rate Variability as a Marker of Training Load and Fatigue in Athletes - Source: Medicine & Science in Sports & Exercise (2020) - Details: Investigated HRV as a biomarker for training-induced fatigue. Found consistent reductions in HRV metrics during periods of high training load, supporting HRV as a non-invasive fatigue monitoring tool. The study emphasized longitudinal tracking to differentiate between acute fatigue and chronic overtraining.

[12] Detection of Fatigue in Runners Using Wearable IMUs and Machine Learning - Source: IEEE Transactions on Human-Machine Systems (2021) - Details: IMUs captured kinematic data from runners. Machine learning models classified fatigue states based on stride variability, angular velocity, and acceleration features. Demonstrated strong predictive performance and highlighted the importance of biomechanical signals in detecting subtle fatigue-related changes.

[13] Fatigue Monitoring in Cycling Using Power Output and Heart Rate Variability - Source: Journal of Applied Physiology (2019) - Details: Combined cycling power output data with HRV measures to monitor fatigue. Found that HRV declined as cyclists approached fatigue thresholds, while power output variability increased. The study provided evidence for integrating physiological and performance metrics in endurance sports.

[14] Wearable Technology for Monitoring Fatigue in Military Personnel - Source: Military Medicine (2020) - Details: Studied wearable sensors in military training contexts. Heart rate, accelerometer, and gyroscope data were used to detect fatigue during prolonged physical exertion. Machine learning models achieved reliable classification, showing applicability of

fatigue monitoring in high-stress environments.

[15] Fatigue Detection Using Gyroscope-Based Gait Analysis - Source: Sensors (2021) - Details: Gyroscope data was analyzed to detect gait changes associated with fatigue. Features such as angular velocity variability and joint rotation patterns were strong indicators of fatigue onset. The study highlighted gyroscope signals as particularly sensitive to biomechanical fatigue markers.

[16] Multimodal Fatigue Monitoring in Rowing Athletes - Source: European Journal of Sport Science (2022) - Details: Combined heart rate, accelerometer, and gyroscope data in rowing athletes. Machine learning models detected fatigue states with high accuracy, particularly when integrating biomechanical and physiological signals. The study emphasized the importance of multimodal monitoring in sports with repetitive motion.

[17] Fatigue Detection in Marathon Runners Using Wearable Sensors - Source: Journal of Sports Engineering and Technology (2021) - Details: Wearable sensors tracked marathon runners' heart rate and motion data. Fatigue was detected through HRV decline and stride variability. Models achieved strong predictive accuracy across different runners, showing robustness in long-duration endurance events.

[18] Real-Time Fatigue Estimation Using Smartwatch Data - Source: IEEE Access (2020) - Details: Smartwatch data (heart rate, accelerometer) was used to estimate fatigue in real time. Machine learning models classified fatigue states, demonstrating feasibility of consumer wearables for fatigue monitoring. The study highlighted accessibility and scalability of fatigue detection outside laboratory settings.

[19] Fatigue Monitoring in Basketball Players Using IMUs and HRV - Source: Journal of Strength and Conditioning Research (2022) - Details: Basketball players were monitored with IMUs and heart rate sensors. Fatigue was detected through HRV decline and increased variability in jump mechanics. Machine learning models achieved strong classification accuracy, showing applicability in intermittent, high-intensity sports.

[20] Wearable Sensor-Based Fatigue Detection in Long-Distance Cyclists - Source: Sports Medicine - Open (2023) - Details: Long-distance cyclists were monitored using heart rate sensors and IMUs. Fatigue was detected through HRV decline, increased variability in pedaling mechanics, and altered gyroscope-based joint rotation. Models achieved high predictive accuracy, validating multimodal monitoring in endurance cycling.

3 Methodology

3.1 *Dataset Exploration*

The dataset used in this article is the Intelligent Biosensor Dataset provided via Kaggle, which can be found here. The dataset contains real-time biosensor data collected from simulated college track and field athletes. Each row contains a timestamped reading from wearable sensors, and is associated with a unique athlete ID, as well as a labeled motion event, which serves as the target variable for motion classification tasks. An example is shown in Table 1 . The numbers are rounded to 3 significant figures to preserve spacing.

Heart_Rate	Acc_X	Acc_Y	Acc_Z	Gyro_X	Gyro_Y	Gyro_Z	Label	Timestamp	Athlete_ID
166	-2.60	2.21	-0.649	-48.5	15.4	164	Accel	2025-05-11 12:14:53.417747	A001

Table 1: **A sample row**. Table containing a sample row from the dataset showing the format and layout of the data within.

With the format and layout of the dataset laid out, the following metrics in Table 2 were

calculated. It should be noted that there is an absence of demographic data, such as age or sex, within the dataset.

Metric	Value
Number of Participants	5
Days Per Participant	1
Seconds Contributed Per Participant	30
Total Labels	6

Table 2: Summary statistics for the biosensor dataset.

Additionally, we also analyzed the spread of the labels and found the distribution to be fairly even between the 6 labels. The findings are summarized below in Table 3, and a visualization for the overall distribution is shown in Figure 1.

Event Label	Value
Start_Run	281
Stop	257
Jump_Takeoff	246
Sprint_Mid	245
Accel	241
Landing	230
Total	1500

Table 3: **Event label distribution.** A table showing the make up of all event labels recorded.

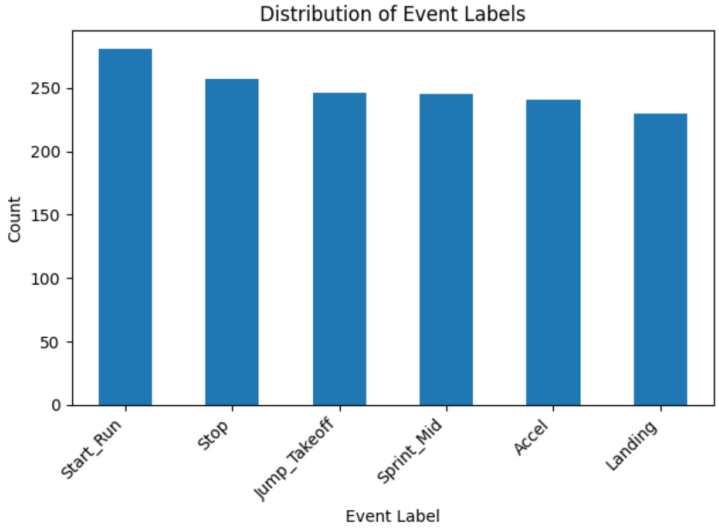

Figure 1: **Bar graph of label distribution**. This image was generated with `matplotlib`.

3.2 *Annotation and Signal Exploration*

In order to understand what our signal data looked like in context, we plotted the raw signal data with their associated events. For each event, we plot the triaxial acceleration, the magnitude of the acceleration, the triaxial gyroscopic values, the magnitude of the gyroscopic values, and the heart rate of the athlete during the labeled event. A sample plot for Start_Run is shown in Figure 2.

Each plot is centered on a one second interval, focusing on a single labeled event, which is highlighted in red. Each plotted value (acceleration, gyroscopic measurements, and heart rate) has a separate graph to fully illustrate the data signal. At each timestamp, the labeled event is provided to provide a better visualization of what the athlete was doing at that moment.

The dataset provided triaxial acceleration and gyroscopic measurements, and heart rate data. The magnitude of the acceleration and gyroscopic measurements were calculated with Equation 1 and Equation 2 respectively:

$$a_{\text{mag}}(t) = \sqrt{a_x(t)^2 + a_y(t)^2 + a_z(t)^2} \tag{1}$$

$$\omega_{\text{mag}}(t) = \sqrt{\omega_x(t)^2 + \omega_y(t)^2 + \omega_z(t)^2} \tag{2}$$

To better understand the signal patterns, we will conduct a brief analysis on the raw signals.

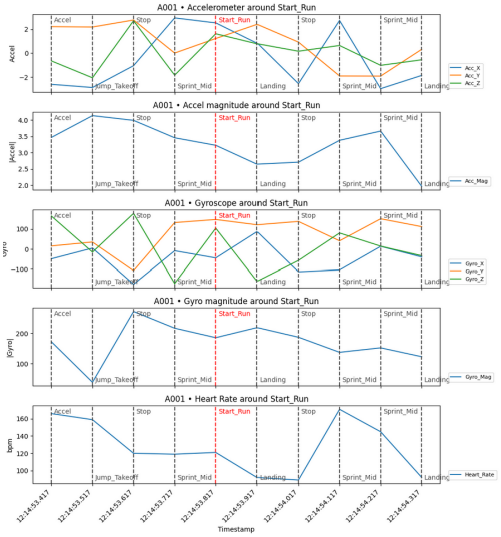

Figure 2: **Signal Exploration Graph**. This image was generated with `matplotlib`.

This figure shows a sequence of biomechanical and physiological responses around the Start_Run event. In the triaxial accelerometer plots, discrete peaks and sign changes around Accel, Jump_Takeoff, Landing, and Stop reflect brief high-impact actions: propulsion, impact absorption, and braking, where specific axes dominate depending on the direction of movement. Immediately after Start_Run (red line), these isolated impulses give way to a more rhythmic oscillatory pattern, representing the first strides of continuous running; this transition is summarized in the acceleration magnitude plot as a shift from lower or irregular values to a sustained, elevated magnitude indicating a higher overall mechanical load.

The gyroscope signals show a parallel pattern in rotational kinematics: sharp changes in angular velocity at Jump_Takeoff and Landing reflect rapid joint rotations during the take-off and impact, while increased, periodic gyro activity around Sprint_Mid corresponds to cyclic leg and trunk rotation during steady running. The gyro magnitude panel collapses these axis-specific rotations into a single profile, highlighting peaks at explosive actions and a maintained high level during the running phase. Finally, the heart-rate trace provides a slower, integrative measure of physiological strain: HR remains elevated or continues to rise across Start_Run and Sprint_Mid, reflecting the cumulative cardiovascular response to the succession of high-intensity mechanical events identified in the kinematic signals. Together, the five signals show the athlete transitioning

from discrete explosive movements into sustained running, with both mechanical intensity and physiological load increasing around Start_Run and peaking during the subsequent sprint phase.

3.3 *Signal Preprocessing*

Raw biosensor signals often contain noise, dropouts, and artifacts that can negatively affect downstream feature extraction and classification. To improve signal quality and reliability, we applied a series of preprocessing steps including noise reduction, missing-data handling, and visual validation through before and after comparisons.

First, noise reduction was performed using a low-pass filtering approach to reduce high-frequency fluctuations that are unlikely to correspond to meaningful human movement or physiological changes. Wearable sensors are particularly susceptible to motion artifacts and sensor jitter, and filtering helps preserve the underlying signal structure while reducing spurious variation. This step is especially important for frequency-domain features and variance-based metrics, which are sensitive to noise.

Next, missing or invalid samples were handled using interpolation. Short gaps in biosensor data can arise from sensor dropout or transmission issues; discarding these segments entirely would unnecessarily reduce the amount of usable data. Linear interpolation was therefore applied to estimate missing values under the assumption of smooth short-term signal behavior. This approach maintains temporal continuity while avoiding abrupt discontinuities that could distort windowed features.

To verify the effectiveness of preprocessing, representative signals were visualized before and after filtering and interpolation. These plots, shown in Figure 3show a clear reduction in high-frequency noise and the successful reconstruction of short missing segments, while preserving overall signal trends and timing. This visual validation confirms that preprocessing improves signal quality without introducing obvious distortions.

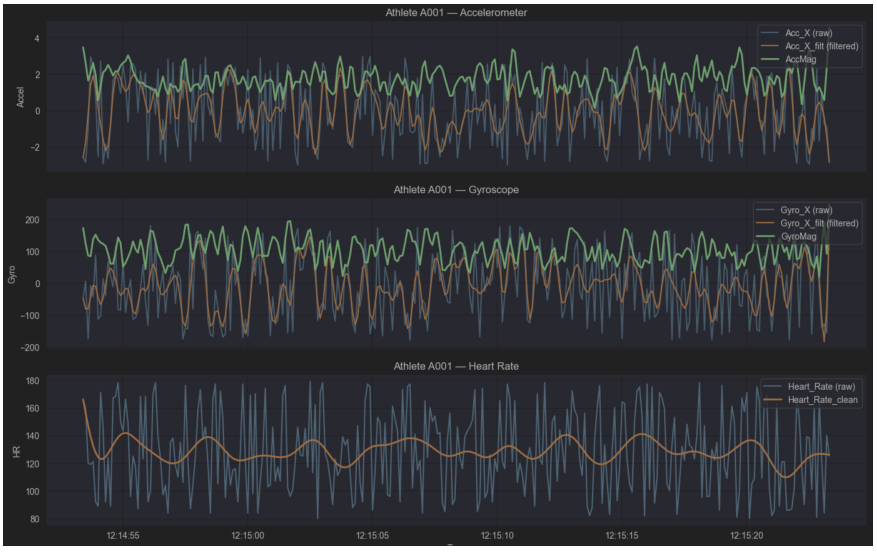

Figure 3: **Preprocessed Signals, Before and After**. This image was generated with `matplotlib`.

Overall, preprocessing plays a critical role in ensuring that extracted features reflect true physiological and movement patterns rather than sensor artifacts. By reducing noise and handling missing data consistently, these steps improve the robustness and interpretability of subsequent windowing, feature extraction, and classification stages.

3.4 *Windowing Strategies*

After preprocessing, continuous biosensor signals were segmented using fixed-length sliding windows. Window-based segmentation is a standard approach in wearable sensing because it converts continuous time-series data into uniform samples that are compatible with traditional machine learning pipelines and window-level feature extraction.

Fixed sliding windows were chosen over event-based or dynamic segmentation for several reasons. First, dynamic segmentation typically requires reliable detection of event boundaries or change points, which can be sensitive to noise and may introduce additional sources of error when event labels are imperfect or transitions are gradual.

In contrast, fixed windows provide a simple and reproducible segmentation strategy that does not depend on prior assumptions about event timing. This makes them particularly suitable for exploratory analysis and datasets with heterogeneous movement patterns.

Sliding windows with overlap were used to preserve temporal continuity and reduce boundary effects. Overlapping windows allow adjacent segments to share information, which helps capture transitions between activities while maintaining consistent window lengths. A 50% overlap was selected as a common compromise that increases temporal resolution without producing excessive redundancy.

To determine an appropriate window length, multiple fixed window sizes ranging from short to longer durations were evaluated. Graphs showing the tradeoffs between different window sizes can be seen in Figure 4. Short windows provide finer temporal resolution and are more sensitive to rapid changes in movement, but they may capture insufficient context and be more affected by noise. Longer windows provide greater temporal context and smoother signal statistics, but they may mix multiple activities within a single segment and reduce sensitivity to transitions. Testing a range of window sizes allows this trade-off to be examined systematically.

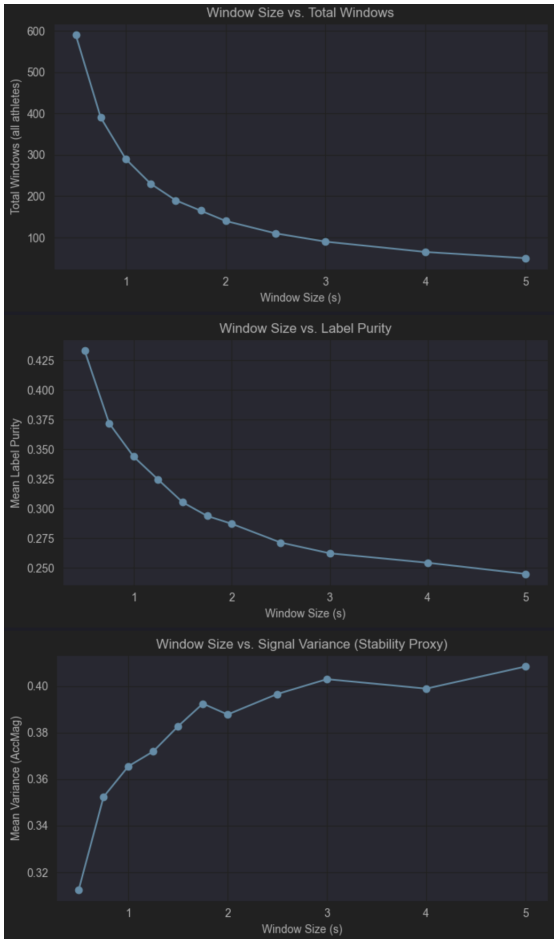

Figure 4: **Different Window Sizes Compared Against Each Other**. This image was generated with `matplotlib`.

Based on these considerations, an intermediate window length of 1.5 seconds was initially selected for subsequent analysis. This choice reflects a balance between temporal resolution and signal stability, ensuring that each window captures sufficient information for meaningful feature extraction while minimizing the risk of mixing distinct activities. Fixing the window size at this stage provides a consistent segmentation framework for downstream feature extraction and classification experiments.

3.5 *Feature Extraction*

To convert segmented sensor windows into inputs for machine learning, we extracted a compact set of features from the time and frequency-domains from each window. These features were computed per signal channel (e.g., accelerometer magnitude, gyroscope magnitude, and heart rate), producing a fixed-length feature vector for every window.

For each window, we computed time-domain features that summarize signal level and variability:

- **Mean**: average signal magnitude within the window.
- **Standard deviation (SD)**: variability around the mean.
- **Root mean square (RMS)**: overall signal power in the time domain.
- **Zero-crossing rate (ZCR)**: rate of sign changes, used as a simple proxy for oscillatory behavior.

These time-domain features capture intensity and stability of movement (and physiological trend when using heart rate).

To capture rhythmic structure not visible in simple statistics, we computed frequency-domain features using the discrete Fourier transform of each windowed signal:

- **Dominant frequency**: frequency with the maximum spectral power (excluding DC when possible).
- **Spectral energy**: total power across the spectrum, reflecting the overall strength of periodic components.
- **Spectral entropy**: normalized measure of how concentrated versus spread out the spectrum is, where higher values indicate less regular (more broadband) motion.

Frequency features are particularly useful for human activity signals because many movements exhibit characteristic cadence and periodicity.

Figure 5 shows a ranking of the most discriminative features based on ANOVA F-scores. Features with higher scores provide greater separation between event classes, indicating stronger class-discriminative power. The ranking highlights the importance of frequency-domain features such as dominant frequency and spectral energy, particularly from acceleration magnitude and heart rate signals, suggesting that rhythmic and periodic characteristics are key for distinguishing movement events. Time-domain features related to variability (e.g., standard deviation and RMS) also contribute meaningfully, reflecting differences in movement intensity and stability across activities. This analysis validates the inclusion of both time and frequency-domain features and provides insight into which signal characteristics are most informative for downstream classification.

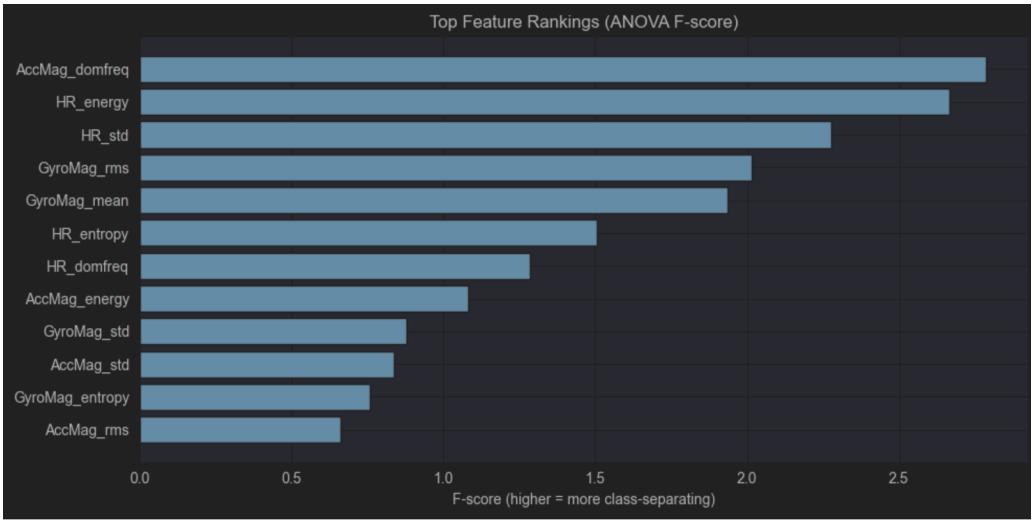

Figure 5: **Feature Importance with ANOVA F-Score**. This image was generated with `matplotlib`.

Overall, the extracted feature set provides complementary information: time-domain metrics describe intensity and variability, while frequency-domain metrics capture rhythmic structure and regularity. The feature importance ranking helps identify which signal characteristics are

Table 4: Classical model performance on the 6-class motion classification task (80/20 split).

Model	Accuracy	Precision (macro)	Recall (macro)	F1 (macro)
Decision Tree	0.263	0.220	0.249	0.210
AdaBoost	0.211	0.162	0.182	0.151
Random Forest	0.184	0.108	0.164	0.129
Gaussian NB	0.211	0.088	0.182	0.113
SVM (linear)	0.105	0.141	0.086	0.092
SVM (RBF)	0.105	0.052	0.096	0.067

most informative for distinguishing event types, guiding both interpretation and potential feature refinement in later experiments.

3.6 *Modeling*

In this phase, we evaluated several classical machine learning algorithms to classify athletic motions using preprocessed and windowed biosensor data. The task is a multi-class classification problem in which each signal segment, represented by heart-rate and accelerometer-based features, is assigned to one of six movement categories. The objective is to learn decision boundaries that effectively separate these classes based on the extracted features.

To account for the nonlinear and heterogeneous nature of athletic motion data, we selected a diverse set of models that represent both generative and discriminative learning patterns. Specifically, we trained six classical models: Naive Bayes, Support Vector Machine (SVM), Decision Tree, Random Forest, AdaBoost, and XGBoost. Naive Bayes serves as a simple generative baseline under a conditional independence assumption. Decision Trees and Random Forests naturally capture nonlinear relationships and interactions between features, while ensemble methods such as AdaBoost and XGBoost iteratively improve performance by combining multiple weak learners.

All models were trained and evaluated using a stratified 80/20 train–test split. Performance was measured using accuracy, precision, recall, and macro-averaged F1 score. Because accuracy alone can be misleading in multi-class settings, macro-averaged metrics were used to assess performance across all classes equally. Confusion matrices were also examined to identify common misclassification patterns.

Based on the results in Table 4, the Decision Tree achieved the strongest overall performance among the classical models evaluated, with an accuracy of approximately 26% and a macro F1-score near 0.21. Although the absolute performance remains modest, this result indicates that shallow tree-based models are able to capture some of the nonlinear structure present in the biosensor features. Ensemble methods such as AdaBoost and Random Forest performed slightly worse, suggesting that the limited dataset size and feature overlap between classes may restrict the benefits of ensembling in this setting.

The remaining models, including Gaussian Naive Bayes and Support Vector Machines, showed lower macro-averaged precision, recall, and F1 scores. In particular, the SVM models struggled to separate classes effectively, likely due to overlapping feature distributions and the absence of clear linear decision boundaries. Overall, these results highlight the difficulty of the multi-class motion classification task and motivate the use of more expressive models and additional contextual information in later experiments.

3.7 *Advanced Modeling & Evaluation*

We further investigated the classification behavior of the best-performing classical model, the Decision Tree. We first evaluated model performance using a stratified 80/20 train-test split,

which provides an estimate of performance when training and testing data are drawn from the same subject distribution. This evaluation serves as a baseline for understanding how well the models learn discriminative patterns under favorable conditions. Then, we evaluated the model under Leave-One-Subject-Out (LOSO), which tests how well the model could generalize to a new subject.

Window Size Analysis. To assess the impact of segmentation choices, model performance was evaluated across multiple fixed sliding window sizes. Performance metrics were aggregated for each window length and summarized using bar plots and line charts, as shown in Figure 6 and Figure 7. These visualizations illustrate how window size affects classification stability, highlighting the trade-off between short windows that capture fine-grained transitions and longer windows that provide greater temporal context. Based on these trends, an intermediate window size of 3 seconds was selected for the subsequent fatigue analysis experiments.

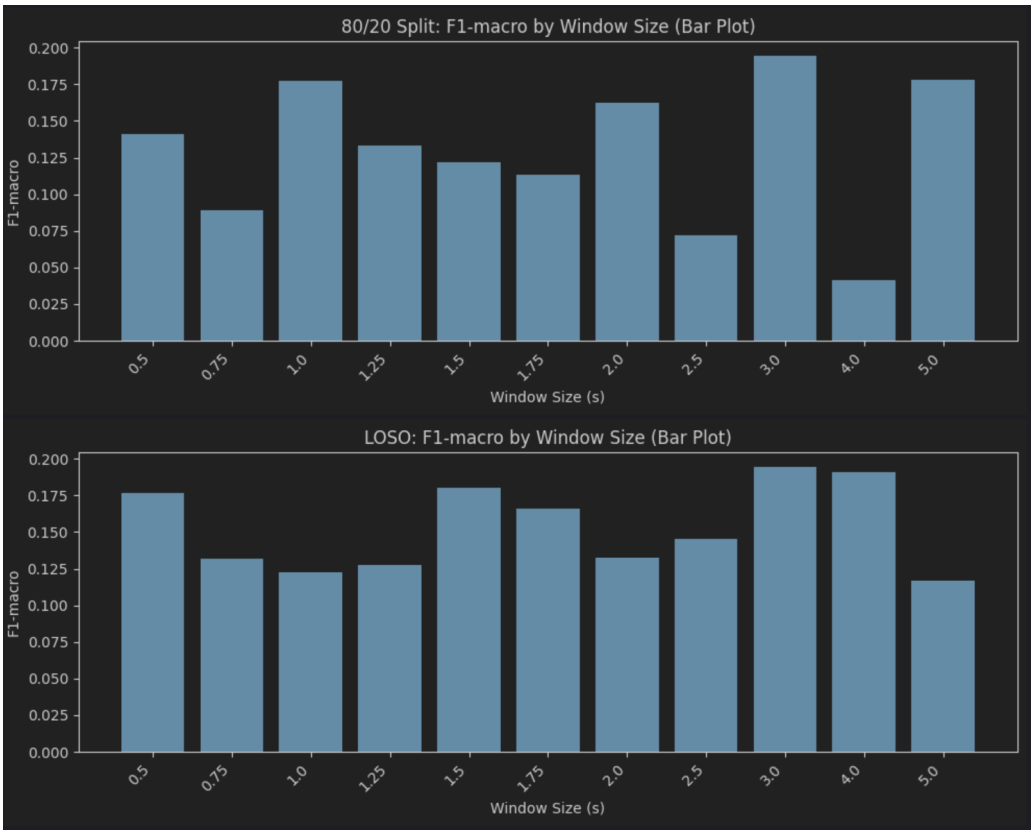

Figure 6: **Bar plot, F1-macro by Window Size**. This image was generated with `matplotlib`.

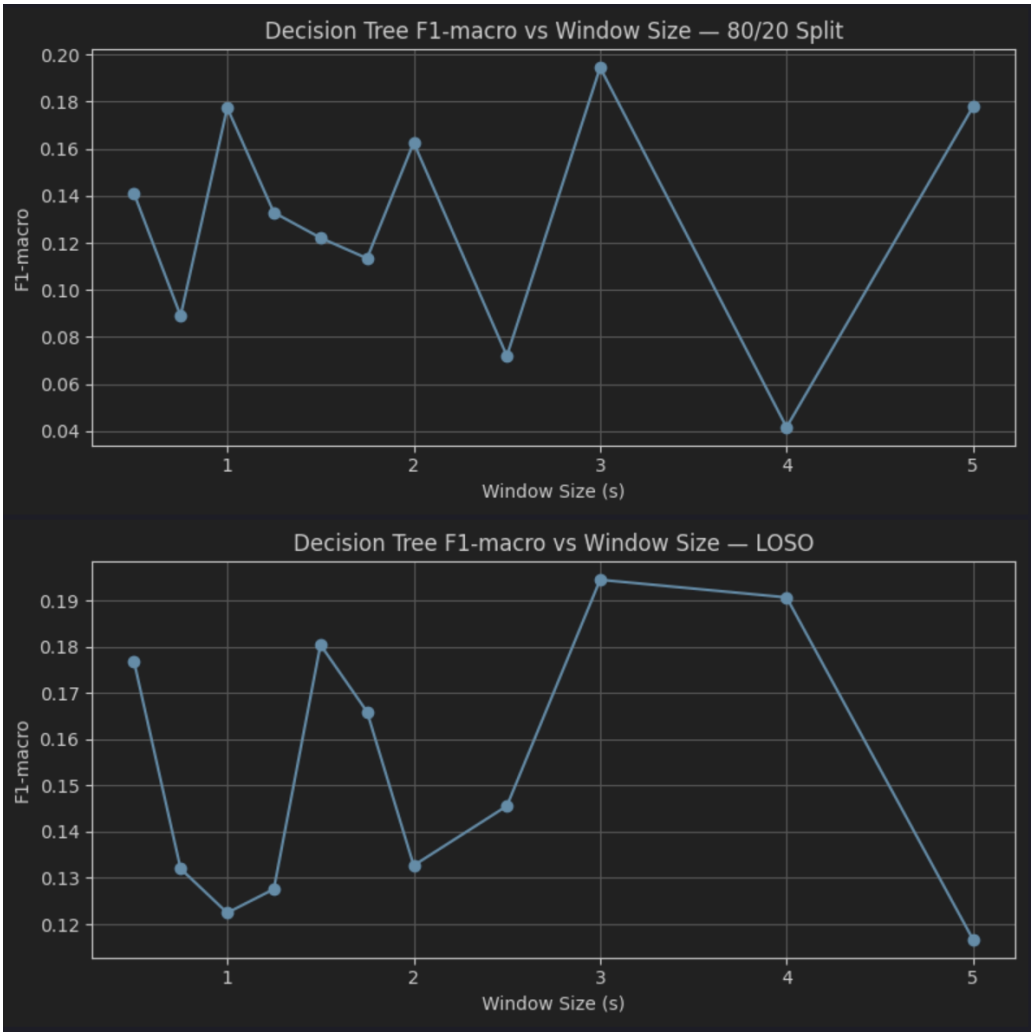

Figure 7: **Line plot, F1-macro by Window Size**. This image was generated with `matplotlib`.

Confusion Matrix Analysis (80/20 Split). The normalized confusion matrix for the 80/20 split (Figure 8) shows that some classes are identified more reliably than others, while several motion categories are frequently confused. In particular, transitional movements such as *Landing*, *Start_Run*, and *Sprint_Mid* exhibit substantial off-diagonal mass, indicating overlapping acceleration patterns within short windows. These confusions suggest that fixed-length segments provide limited temporal context for distinguishing closely related phases of motion.

Confusion Matrix Analysis (LOSO). Under LOSO evaluation (Figure 9), class-wise confusion becomes more pronounced and diagonal dominance is reduced. Similar motion classes are frequently misclassified as one another, reflecting increased inter-subject variability and differences in execution style. This result highlights the challenge of subject-independent motion classification and motivates the use of additional contextual or temporal modeling strategies.

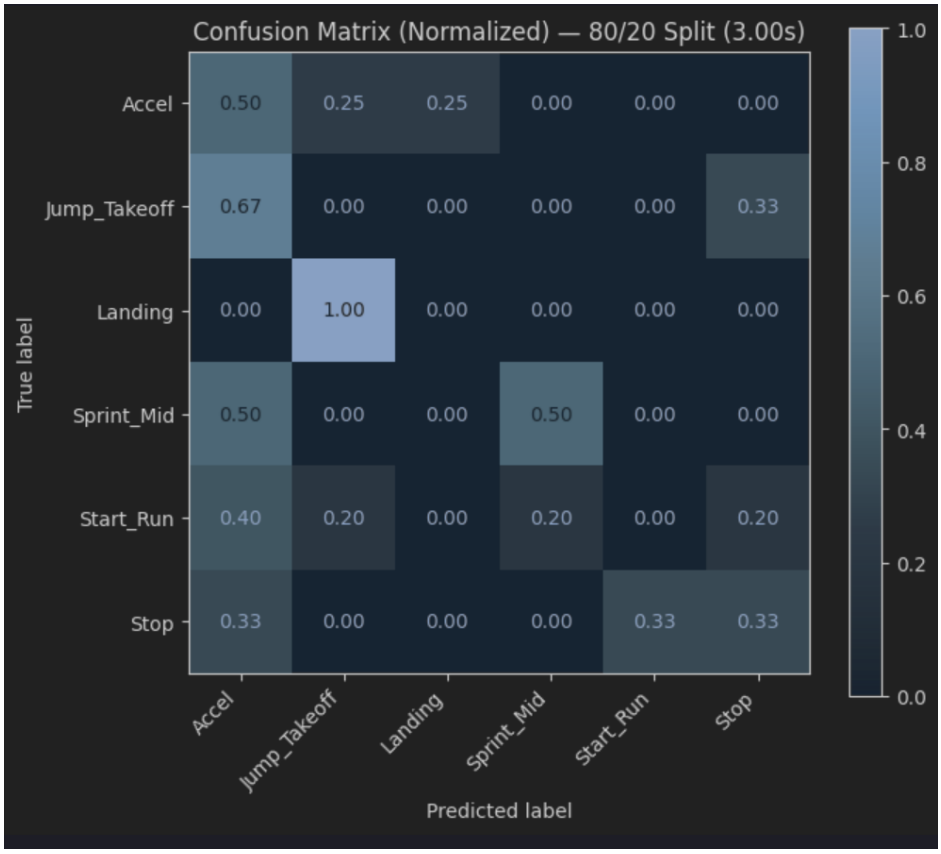

Figure 8: **Confusion Matrix, 80/20 Split**. This image was generated with `matplotlib`.

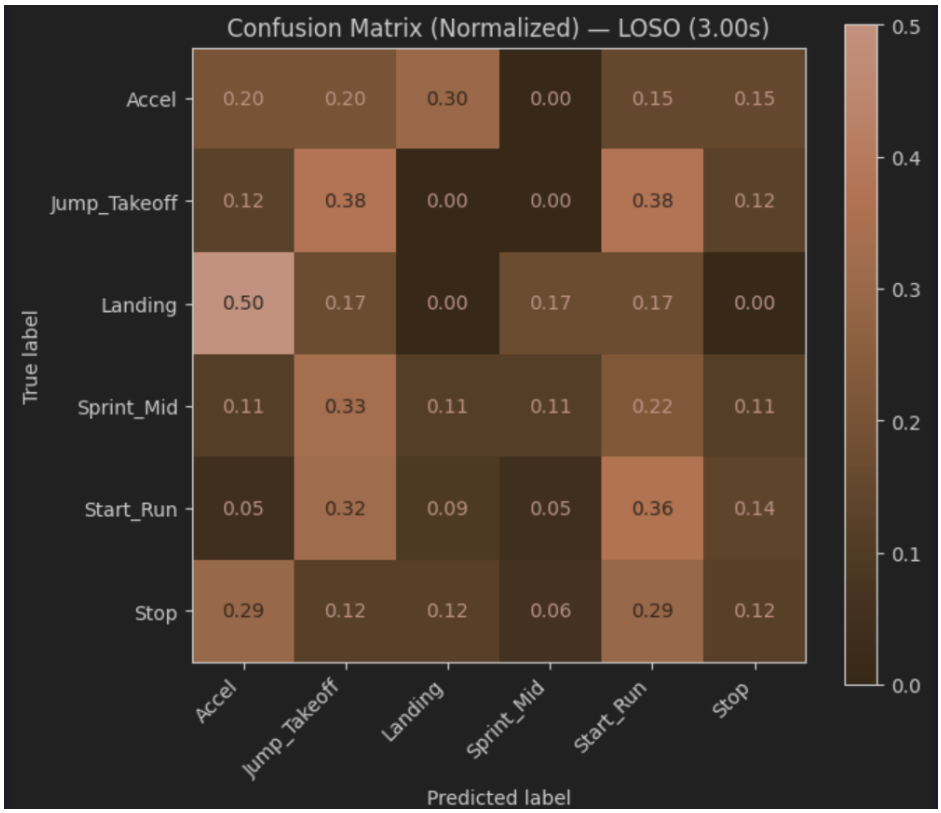

Figure 9: **Confusion Matrix, LOSO Split**. This image was generated with `matplotlib`.

Table 5: Performance comparison for 3.0 s sliding windows under different evaluation protocols.

Evaluation	Accuracy	Precision (macro)	Recall (macro)	F1 (macro)
80/20 Split	0.222	0.181	0.222	0.194
LOSO	0.233	0.210	0.195	0.195

Summary of Results. Table 5 summarizes accuracy, precision, recall, and macro-averaged F1 scores for the evaluated models under the 80/20 and LOSO split. While overall performance remains modest due to dataset size and class similarity, this evaluation provides a useful reference point for comparing against more challenging subject-independent settings.

3.8 *Fatigue Detection*

To provide additional context for model behavior over time, we incorporated a simple fatigue detection mechanism based on deviations from an athlete-specific baseline. Rather than framing fatigue as a separate classification task, it was modeled as a continuous index capturing changes in physiological and movement-related features across a session.

For each athlete, an early portion of the session was used to establish a baseline distribution of heart rate and acceleration-based features. Subsequent windows were compared to this baseline using normalized deviations, which were aggregated into a single fatigue index per window. Higher index values indicate greater divergence from baseline behavior and are interpreted as increased fatigue-like patterns. This approach does not require explicit fatigue labels and is well-suited for limited datasets.

The resulting fatigue index over time is shown in Figure 10. Across athletes, the index fluctuates rather than increasing monotonically, reflecting changes in activity intensity and short recovery periods. Transient peaks above moderate and high thresholds indicate moments of pronounced deviation from baseline, while near-zero or negative values correspond to windows with baseline-like behavior. Overall, the analysis demonstrates that fatigue manifests as intermittent deviations rather than steady accumulation, supporting the use of a baseline-relative fatigue index to contextualize motion classification results.

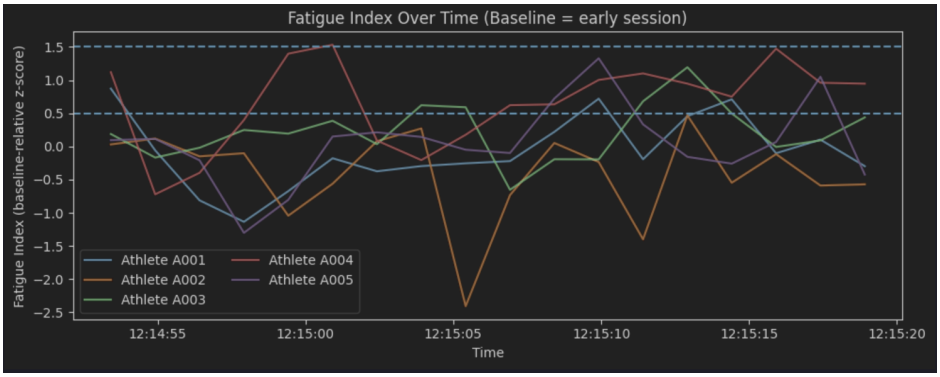

Figure 10: **Fatigue Line Graph**. This image was generated with `matplotlib`.

4 Results and Discussion

Overall, the wearable-sensing pipeline produced a consistent end-to-end workflow from preprocessing and segmentation to feature extraction, modeling, and evaluation, but classification performance remained modest due to limited data and substantial overlap between several motion classes. Fixed-length sliding windows enabled uniform feature computation and model training; an intermediate window size (3.0 s) was selected as a practical balance between capturing

sufficient temporal context and avoiding excessive mixing of multiple activities within a single segment. Using 3.0 s windows, performance was similar under both evaluation protocols

Across classical models, the Decision Tree achieved the strongest results among the tested baselines, while SVM variants performed poorly, suggesting that the extracted feature distributions are not well separated by simple margin-based boundaries. Ensemble methods such as Random Forest and AdaBoost did not provide clear improvements in this setting, likely because the dataset is small and several classes share highly similar acceleration signatures. Confusion matrices further support this interpretation: under the 80/20 split, certain classes appear more separable, while transitional or mechanically similar activities are frequently confused; under LOSO, misclassifications become more diffuse, indicating that inter-subject variability amplifies ambiguity between related movement phases.

Feature ranking analysis showed that both time-domain (e.g., variability and RMS) and frequency-domain descriptors (e.g., dominant frequency and spectral measures) contribute to class separation, supporting the inclusion of complementary feature types. Finally, a baseline-relative fatigue index was added as an interpretive signal: fatigue values varied by athlete and fluctuated over time rather than increasing monotonically, consistent with changes in intensity and brief recovery periods. While not used directly for prediction, this index provides useful context for understanding when signal patterns deviate most from baseline and can motivate future analyses linking fatigue-like drift to classification errors. Future work should prioritize additional labeled data, richer temporal modeling via multi-window sequences, and potentially event-aware segmentation to reduce label mixing within windows.

5 Conclusion

Conflict of interests

The authors have no conflicts of interest to disclose.

Data and Software Code

Our software code is available at the following URL: https://github.com/makinchii/CS156-Project. Our dataset is available at the following URL: https://www.kaggle.com/datasets/ziya07/intelligent-biosensor-dataset

References

[1] Pengfei Chang, Cenyi Wang, Yiyan Chen, Guodong Wang, and Aming Lu. Identification of runner fatigue stages using inertial sensors and deep learning. *Frontiers in Bioengineering and Biotechnology*, 11, 2023.

[2] Milica Stojanac. Machine learning-based prediction of running-induced fatigue. Master's thesis, University of Twente, 2024.

[3] Olli-Pekka Nuuttila, Nilushika Weerarathna, Arja Uusitalo, Veli-Pekka Kokkonen, and Heikki Kyröläinen. Monitoring fatigue state with heart rate-based and subjective methods. *European Journal of Sport Science*, 24(3):456–468, 2024.

[4] Horia Alexandru Modran, Doru Ursuţiu, Cornel Samoilă, Tinashe Chamunorwa, Lilia Aljihmani, and Khalid Qaraqe. Fatigue estimation using wearable devices and virtual instrumentation. In *Lecture Notes in Networks and Systems*, volume 763, pages 1055–1064. Springer, 2024.

[5] Marco Beato, K. L. De Keijzer, B. Carty, and M. Connor. Monitoring fatigue during intermittent exercise with accelerometer-derived metrics. *Frontiers in Physiology*, 10:780, 2019.

[6] Neusa R. Adão Martins, Simon Annaheim, Christina M. Spengler, and René M. Rossi. Fatigue monitoring through wearables: A state-of-the-art review. *Frontiers in Physiology*, 12, 2021.

[7] Jiya Wang and Huan Meng. Fatigue state prediction of athletes based on multi-source sensors. *International Journal of Distributed Systems and Technologies*, 14(2):1–12, 2024.

[8] Elsen Ronando and Sozo Inoue. Improving fatigue detection with feature engineering on accelerometer data. *International Journal of Affective and Behavioral Computing*, 5(2):89–102, 2024.

[9] D. Sheridan, A. Jaspers, D. V. Cuong, T. Op De Beéck, N. M. Moyna, T. T. de Beukelaar, and M. Roantree. Estimating oxygen uptake in simulated team sports using machine learning models and wearable sensor data: A pilot study. *PLOS ONE*, 20(4), 2025.

[10] S. P. Nair, M. Sica, S. Tedesco, and A. Visentin. Real-time fatigue monitoring using multi-sensor fusion and deep learning in endurance athletes. *Sensors*, 22(14), 2022.

[11] R. Sanchez, C. Nieto, J. Leppe, T. Gabbett, and M. Besomi. Associations between training load, heart rate variability, perceptual fatigue, sleep, and injury in endurance athletes. *International Journal of Sports Science Coaching*, 20(5):1918–1928, 2025.

[12] L. Marotta, J. H. Buurke, B. J. F. van Beijnum, and J. Reenalda. Towards machine learning-based detection of running-induced fatigue in real-world scenarios: Evaluation of imu sensor configurations. *Sensors*, 21(10):3451, 2021.

[13] A. Barrero, F. Schnell, G. Carrault, G. Kervio, D. Matelot, F. Carré, and S. Le Douairon Lahaye. Daily fatigue-recovery balance monitoring with heart rate variability in well-trained female cyclists. *PLOS ONE*, 14(3), 2019.

[14] N. Taylor, A. S. Willman, M. Eager, and R. M. Gifford. Enhancing decision-making in the armed forces through wearable technology. *BMJ Military Health*, 171(4):382–386, 2020.

[15] G. Zhang, T. T.-H. Hong, L. Li, and M. Zhang. Automatic detection of fatigued gait patterns in older adults: An intelligent portable device integrating force and inertial measurements with machine learning. *Annals of Biomedical Engineering*, 53(1):48–58, 2024.

[16] F. Ihász, J. Takács, Z. Alföldi, L. Kósa, R. Podstawski, A. Ferraz, B. Hinca, I. Barthalos, and Z. B. Katona. A pilot study of locomotor and mechanical loads on elite rowers during competition days. *Sports*, 13(8):254, 2022.

[17] M. Gholami, C. Napier, A. García Patiño, T. J. Cuthbert, and C. Menon. Fatigue monitoring in running using flexible textile wearable sensors. *Sensors*, 20(19), 2020.

[18] S. Park, S. Seong, Y. Ahn, and H. Kim. Real-time fatigue evaluation using ecological momentary assessment and smartwatch data: An observational field study on construction workers. *Journal of Management in Engineering*, 36(5), 2020.

[19] T. Edwards, T. Spiteri, B. Piggott, J. Bonhotal, G. G. Haff, and C. Joyce. Monitoring and managing fatigue in basketball. *Sports*, 6(1), 2018.

[20] Y. Jiang, V. Hernandez, G. Venture, D. Kulić, and B. K. Chen. A data-driven approach to predict fatigue in exercise based on motion data from wearable sensors. *Sensors*, 21(4), 2021.