Machine learning-based prediction of running-induced fatigue, during outdoor recreational running using IMUs, heart rate, and smartwatch data

Biomedical Signals and Systems
Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS)
University of Twente, Enschede, Netherlands

November, 2024

Student:
Milica Stojanac, Medical Devices Design MSc, Faculty of Science and Technology (TNW)

Committee members:
dr. Jasper Reenalda MSc, Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS)
dr. E.H.F. Edwin van Asseldonk, Faculty of Engineering Technology, Biomechanical Engineering (ETBE)
ir. B.L. Scheltinga BSc, Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS)

Table of Contents:

Machine learning-based prediction of running-induced fatigue, during outdoor recreational running using IMUs, heart rate, and smartwatch data

Biomedical Signals and Systems, Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), University of Twente, Enschede, Netherlands

Abstract

Background: Running offers numerous health benefits but unfortunately carries a high risk of running-related injuries (RRIs), particularly resulting from overuse. Fatigue monitoring methods, such as cardiopulmonary exercise testing (CPET) and lactate concentration measuring, are effective but impractical for real-world running conditions. Wearable sensors combined with novel machine learning (ML) algorithms offer a promising alternative for continuous, real-time fatigue monitoring in realistic, outdoor settings. **Methods**: Nineteen recreational runners participated in this study – fourteen in the first and five in the second experimental part. They completed three distinct outdoor running sessions: endurance, interval, and a 5 km run. Participants were equipped with seven Inertial Measurement Units (IMUs) placed on both tibias, thighs, pelvis, sternum, and wrist, along with a heart rate monitor and a smartwatch to collect kinematic and physiological data. During the second experimental part, fatigue was measured using the Borg Rating of Perceived Exertion (RPE) scale (0 to 10) at specific points during each run, while no such feedback was collected in the first experimental part. A Random Forest regression algorithm was trained on the processed labeled data from the second experimental part to predict RPE at intervals of 1 second. The model was developed using a nested Leave-One-Subject-Out (LOSO) cross-validation framework, with hyperparameter tuning conducted via RandomizedSearchCV. This machine learning framework was applied to selected IMU sensor combinations to optimize practicality and reduce sensor setup. The best-performing model across these sensor configurations was further visually validated on an unlabeled dataset from the first experimental part. **Results**: The single-sensor configuration (wrist) achieved the best performance in RPE prediction, with an average Mean Square Error (MSE) of 1.89. The two-sensor setup (thighs) had an MSE of 2.26, while the three-plus sensor setup (tibias, thighs, and pelvis) recorded the highest MSE of 2.44. The whole-body configuration, with an MSE of 2.16, did not outperform the wrist sensor. Across all sensor configurations, performance was highest in the endurance trial, followed by the interval and 5 km trials, with the 5 km trial showing the least accurate predictions. **Conclusion**: The wrist single-sensor configuration achieved the best performance, outperforming more complex multi-sensor setups. These findings suggest that more sensors do not necessarily improve prediction accuracy, particularly in steady-paced endurance runs. Future research should focus on expanding the sample size, integrating more biometric data, and validating this system against gold-standard fatigue assessment methods, such as electromyography (EMG) and VO2 max.

1. Introduction

Running is a widely popular sport, appreciated for its numerous health benefits including weight loss, cardiorespiratory fitness, mental health, and increased lifespan [61][2]. Due to its low cost and minimal equipment, running is worldwide accessible. Consequently, its popularity and the number of runners have grown substantially over the past 50 years [3].

Despite being an excellent exercise, it does not come without challenges, carrying a high risk of running-related injuries (RRIs). Studies report 7.7 running-related injuries (RRIs) for recreational runners and 17.8 for novice runners per 1000 hours of running [4]. More specifically, 50% of runners encounter an injury on a yearly basis that forces them to take a break from training, whereas 25% of runners are injured at any given moment [5]. Approximately 70% to 80% of RRIs come from overuse injuries, primarily affecting areas such as the knee, ankle/foot, and shank [6][7]. Regardless of their type, they require a temporary or permanent break from exercise due to recovery. Unfortunately, RRIs are one of the major determinants of future injuries and the reason why many runners decide to quit the sport [8][9].

These injuries not only reduce pleasure but also result in substantial financial implications. Considering the direct and indirect costs (e.g., healthcare, absenteeism from work) RRIs pose a significant economic burden [10] which is estimated to vary between €83 and €174 per RRI and €13 and €105 per participant training for an event [11]. Additionally, it is estimated that for every 1000 hours of running, the total cost of RRI is around €1849 [10].

Given the high prevalence and significant impact of running-related injuries, one can conclude that RRI represents an important public health issue. A good understanding of RRIs, and their underlying cause is essential for their prevention, optimizing training strategies, and achieving peak sports performance [12]. Moreover, managing training load and exercise-induced fatigue could be a method to reduce the risk of running-related injuries [12][13].

Most RRIs are overuse-related [14], manifest at the tissue level, and result from a mechanical fatigue phenomenon [15]. During running, the body repeatedly endures high-impact forces from foot-ground collisions, which become difficult to attenuate as fatigue accumulates, contributing to the risk of overuse injuries [14]. Moreover, repetitive loading causes tissue damage accumulation and progressive loss of strength, which then triggers body adaptation and remodeling. However, if loading continues without the incorporation of adequate rest and tissue repair, it may lead to failure [15]. For example, it is a common practice in running training programs to increase load according to the overload principle, meaning that workload demand on the body should be greater than the one to which it is accustomed [16]. While on the one hand, this could be beneficial for athletic performance, on the other, it could lead to fatigue buildup which as a result increases the risk of injuries if not timely detected and properly managed [17].

With the help of an effective fatigue feedback system, the risk of RRIs could be lowered by enabling athletes and coaches to make more informed decisions regarding training load and rest days. Not only these fatigue monitoring devices should provide immediate feedback during exercise [18], but they should also be non-invasive, time-efficient, and minimize any additional loading of the athlete [19]. This represents a serious challenge, especially due to the lack of single metrics that can accurately detect fatigue progression [13].

Over the past decade, researchers have explored various methods for fatigue monitoring [20]. In practice, fatigue measurements are task-specific, and one should carefully consider how to define fatigue for a given population and type of activity [21]. Nowadays, the most used methods are widely accessible tests of direct physiological means such as heart rate, blood lactate concentration, or self-reports and perception scales such as Borg's Rating of Perceived Exertion (RPE) [22][23]. Although it represents a subjective fatigue estimate, the RPE scale shows a strong correlation with heart rate ($r = 0.74$) and blood lactate ($r = 0.83$) concentration and is applicable in both controlled laboratory settings and real-world conditions [24]. Consequently, the RPE has gained increasing popularity in running related studies owing to its contribution to assessing an individual's training load and subsequent - injury risk [25]. Nevertheless, there is still a need for more objective measures, that could complement subjective assessments of fatigue [22][25]. Cardiopulmonary exercise testing (CPET) represents the golden standard for objective fatigue assessment but requires expensive equipment, laboratory settings, and trained personnel [24][25]. Since fatigue induces changes in running gait kinematics [14] 3D motion capture systems could be used to record those changes, but while accurate, they are costly and cumbersome [27]. Additionally, these methods are limited to indoor, laboratory testing, which is not suitable for fatigue assessment in real-life running scenarios. More recently, wearable sensors and machine learning (ML) algorithms have emerged as effective alternatives, as they offer a promising solution by providing ongoing, long-term monitoring of physiological signals in a comfortable and non-intrusive manner [20]. Inertial Measurement Units (IMUs), which use accelerometers, gyroscopes, and magnetometers, offer continuous, objective data collection on movement and can potentially bridge the gap between laboratory and everyday fatigue assessment [28]. Finally, heart rate monitors and GPS-enabled smartwatches can provide heart rate (HR) and global positioning information that could be valuable for fatigue monitoring.

This research focuses on predicting running-induced fatigue using a supervised ML algorithm that analyzes data from seven IMUs, a heart rate monitor, and a smartwatch during outdoor recreational running sessions, supplemented with general subject information (e.g., age, weight, height) and questionnaires. In this research, running-induced fatigue is quantified by the Borg Rating of Perceived Exertion (RPE) 0 to 10 scale, and the aim is to predict running-induced fatigue in the 1-second time intervals, which is defined as a multivariate time series regression problem. The study is unique in its scope due to its longitudinal design, aiming to predict fatigue across three distinct running scenarios – endurance, interval, and a 5 km run – in uncontrollable outdoor conditions, on various outdoor surfaces (e.g., athletic track, road). To our knowledge, this is the first study to employ this specific setup for fatigue forecasting. Furthermore, the study explores the feasibility of a minimal IMU sensor setup, contributing to wearable medical technology. The research questions guiding this study are:

1. Can machine learning algorithms be effectively developed for predicting running-induced fatigue in outdoor recreational running during different training sessions, utilizing data from IMUs, heart rate monitors, and smartwatches?
2. Additionally, what minimal sensor setup is optimal for predicting running-induced fatigue in outdoor recreational running?

2. Related work

Supervised machine learning has been widely used for fatigue detection, where models learn to predict outcomes from labeled input (e.g., physiological and motion data) and corresponding output (e.g., fatigue level) pairs provided during training. While these models have demonstrated promising results in physical fatigue monitoring in workplace settings [29], their application in sports research remains limited, primarily focusing on movement classification and exercise detection [30]-[32]. The use of IMUs for predicting running-induced fatigue is even less explored, with few studies [34] specifically addressing fatigue detection in outdoor running conditions.

Wang et al. developed a Random Forest model with a fatigue classification accuracy of 91.10% using a combination of tibia and thigh IMUs, while in the single sensor configuration, right tibial IMU data performed the best with an accuracy of 87.21% in running on the athletic track [33]. Marotta et al. demonstrated that using a Random Forest classification algorithm trained on IMU-derived biomechanical features, fatigue could be classified with accuracies up to 90.5% during running on an athletic track, with the left tibia being the most informative sensor location [34]. Buckley et al. predicted subject-dependent and subject-independent binary fatigue levels using data from a single IMU and Random Forest model. The study showed that the right shank IMU performed the best overall in subject-dependent fatigue estimation, and IMU on the lumbar spine in subject-independent model. The subject-dependent classifier achieved a higher accuracy of 100% compared to the subject-independent classifier of 75% accuracy [35]. Op De Beéck et al. demonstrated that gradient-boosted regression subject-independent trees performed best in predicting the RPE in outdoor runners on the athletic track, with wrist-worn IMU sensors providing the most accurate predictions in the single sensor configuration of mean absolute error of 1.89, while a fusion of sensors on tibia, wrist, and arm showed only minimal performance improvements of 1.84 [36].

Among various machine learning techniques, Random Forest models are particularly popular in fatigue research most likely due to their ease of use, robustness, and versatility across diverse tasks [20]. They excel at handling noisy data and reducing overfitting by averaging multiple decision trees and are reliable for both classification and regression tasks. Furthermore, they are a good option for problems that exploit non-linear, complex relationships between features, and the ability to provide feature importance makes them useful for understanding key variables in a dataset. Random Forests also perform well in high-dimensional spaces and are resilient to multicollinearity, meaning they remain effective even when features are highly correlated, although this may affect the interpretability of feature importance [37],[38].

Despite promising results, the mentioned studies highlight the limitations of current methods. Most existing fatigue research focuses on the binary or three-state classification of fatigue [33], which can oversimplify the complexity of fatigue potentially limiting opportunities for early intervention. Furthermore, this method requires reference measure thresholding which can alter models' performance (e.g., maximizing the distance between fatigued and non-fatigued stated improves model performance) [39]. Continuous monitoring through regression models could track the accumulation of fatigue providing real-time feedback, however, such approaches remain underexplored in the current literature [20][36]. Additionally, integrating personal information, such as age, weight, and height, with physiological and motion data has been shown to enhance model accuracy [39], yet only a few studies have explored this approach [20].

A significant issue in fatigue research is the lack of standardization in research methodology, making result comparisons across studies difficult. The relevance of fatigue predictors depends on the performed task [20] and many studies focus primarily on the lower extremities [33][35], often overlooking the potential insights from upper limb sensor placements. Overall, many of the conducted studies were short-term and performed in controlled environments. Research has shown that fatigue-induced kinematic changes significantly differ between laboratory treadmill and outdoor running [42], therefore, data acquisition should be done longitudinally and under realistic, uncontrollable conditions, such as diverse weather, running surfaces, and training sessions, to ensure the model's generalization to real-world settings.

3. Materials and methods

To address the research questions this study follows the proposed machine learning workflow summarized in Figure 1. The workflow involves 3 key steps:

Step 1: Data collection and processing, including segmentation based on gait cycles and feature extraction. This step ensures capturing biomechanical and physiological data relevant to fatigue, thus addressing the first research question.

Step 2: Development and fine-tuning of the fatigue prediction regressor for 15 different IMU sensor combinations, addressing the feasibility of a minimal sensor setup (second research question).

Step 3: Application of the best model from step 2 on an unlabeled dataset for further visual validation, by observing predicted RPE trends against collected HR data.

Each of these steps is explained in more detail in the following sections.

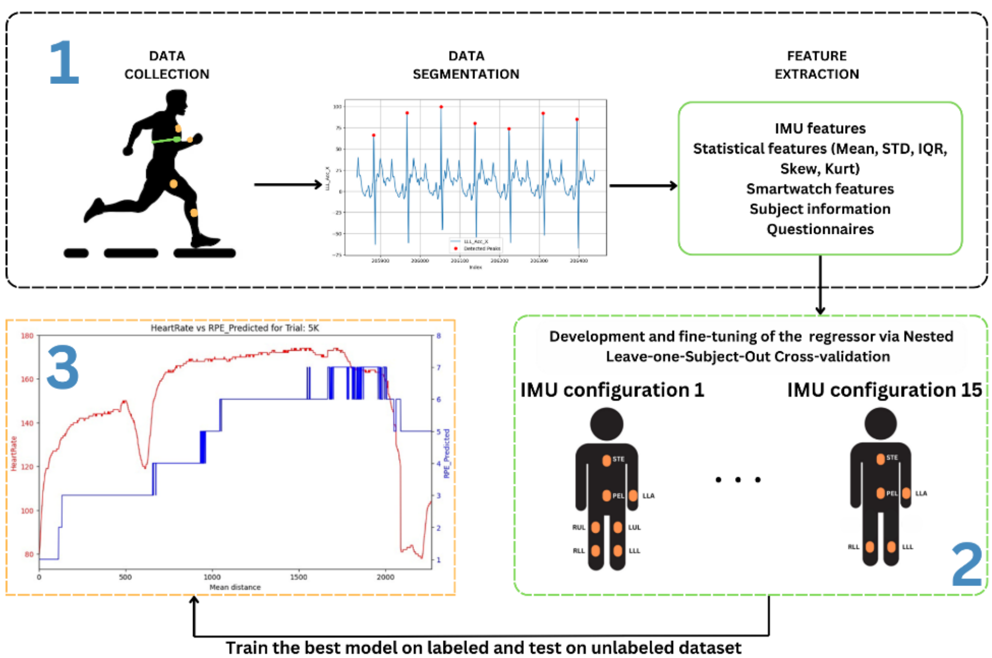

Figure 1 Machine learning algorithm workflow.

3.1. Experimental design

Nineteen recreational runners participated in the research after signing informed consent. Fourteen runners participated in the first experimental part (seven males, seven females), and five in the second (three males, two females). The only distinction between these two experimental parts is that, in the second set, participants were asked to report their RPE during the run, whereas no such feedback was collected in the first set. Two participants (S11, S12) were excluded from the research due to injury. Runners' characteristics are shown in Table 1. Participants were included in the research if they met the following criteria: 1) ran at least 20 km per week in the last 3 months, 2) ran 2 times per week or more on average in the last 3 months, 3) did not have any major running-related injuries in the lower extremities in the past 6 months, 4) are not pregnant. Participants were recruited through local athletics and triathlon associations. The protocol was approved by the University of Twente's ethics committee.

Table 1 Runners' characteristics
[1]STD – standard deviation

ID	Age	Gender	Weight (kg)	Height (cm)	5 km pace	Foot strike pattern	HR rest	HR max	Experience (years)	Km/ week
First experimental part										
S01	29	Male	74	193	03:54	non-heel	40	200	10	25
S02	26	Male	77	195	03:54	Heel	45	180	10	30
S03	25	Male	74	191	04:07	non-heel	47	188	10	35
S04	33	Male	81,7	195	03:36	non-heel	46	205	15	80
S05	23	Female	58,8	173	04:45	Heel	55	206	7	20
S06	20	Male	86,2	194	04:45	Heel	38	200	1	20
S07	22	Female	72,2	182	04:50	Heel	55	200	4	25
S08	44	Female	66,2	176	04:45	Heel	55	185	25	35
S09	42	Female	55,8	157	04:45	Heel	42	178	20	65
S10	21	Female	63,3	173		Heel	55	209	10	25
S11	27	Female	68,2	173		Heel	42	198		
S12	53	Female	67,7	176		Heel	50			
S13	23	Male	74,4	180	03:40	Heel	50	202	4,5	50
S14	27	Male	81,2	194	04:09	non-heel	42	183	5	55
Mean± STD[1]:	29,6± 9,6		71,5± 8,4	182,3± 11,2	04:17± 0:02	71,5± 8,4	47,3± 5,9	194,9± 10,2	10,1± 6,6	38,7± 18,6
Second experimental part										
S15	53	Male	89	194	04:10	Heel	53	200	1	55
S16	30	Female	66,5	173	05:00	non-heel	54	205	7	33
S17	50	Male	87,3	192	03:55	non-heel	40	176	12	55
S18	30	Male	73,9	184	03:50	non-heel	34	185	12	40
S19	27	Female	57,7	168	04:24	non-heel	44	195	10	23
Mean± STD:	38,0± 11,1		74,9± 12,0	182,2± 10,24	04:15± 0:02		45± 7,6	192,2± 10,5	8,4± 4,1	41,2± 12,5

3.2. Measurement setup

During each visit, participants were equipped with seven IMU sensors (Xsens DOT, Xsens Technologies B.V., Enschede, The Netherlands) with a sampling frequency of 120Hz, 3D accelerometer range of 16 g, 3D angular velocity range of 2000 °/s, along with a smartwatch (Garmin Forerunner 55, Olathe, KS. USA) with a recording sampling rate of 1 Hz, and a strap-based heart rate monitor (Garmin HRM-Dual, Olathe, KS, USA). Participants could use their own Garmin smartwatch if it was no older than 2018, and any compatible heart rate monitor.

The IMU sensors were placed on both tibias, thighs, pelvis, sternum, and wrist under the smartwatch, which was worn on either the left or right lower arm, depending on the participant's preference (see Figure 2). Double-sided tape and additional covering tape were used to secure the sensors. Magnetic field calibration and temporal synchronization of the sensors were performed following the manufacturer's instructions.

The warming-up and running pace were calculated during their first visit based on their latest representative results (5 km or 10 km). If that time was unknown, it was estimated with Riegel's rule from another race result.

Riegel's rule: $T_2 = T_1\,(\frac{D_2}{D_1})^{1.06}$

where T1 and D1 represent the time and distance of the known result and T2 and D2 the time and distance of the calculated result.

Next, the time for the sub-maximal 5 km was estimated by taking the 5 km time and adding 10%. This time was discussed with the participant and could be changed if he/she desired, making sure the speed offered a challenging intensity run while allowing participants to complete the sessions without excessive fatigue, thus enabling consistent data collection. The determined speeds were recorded in a measurement form (Appendix B) along with general participant information. Body mass was measured after sensor placement using a calibrated scale. Before and after each run, as well as the day after the run, subjects filled in a questionnaire, with questions regarding participants' perceived levels of muscle, tendon, and joint stiffness or pain, as well as overall fatigue and readiness for physical activity (Appendix C). Before each run was performed, Borg's Rate of Perceived Exertion scale was explained to the participant as described in Table 2. During the warming-up participant's feet were video recorded with a smartphone to determine their foot strike pattern (heel or non-heel striker).

Figure 2 Measurement setup. Sensor placement points: both tibias (1), thighs (2), pelvis (3), sternum (4) and wrist (5), and the sensor coordinate system.

3.3. Running protocol

Each participant ran an endurance, interval, and 5k submaximal running protocol in 3 separate visits, with at least one rest day in between. After the placement of the sensors, the recording was started, followed by the sensor calibration movements:

- 3 good morning movements
- 3 squats
- 3 knee flexes (right and left)
- 3 elbow flexes (arm with a sensor only)
- N-pose for 5 seconds: ensuring the participant stood straight with feet shoulder-width apart and knees at a 180-degree angle.

Once the calibration was complete, the participant performed a jump, starting the smartwatch as they landed, and immediately began their run.

1) The endurance run (45 minutes of running), consisted of 4 laps (4x400 meters) warming-up pace on the athletic track, directly followed by running on the predefined route on the University of Twente campus (see Figure 3). The endurance run pace was the same as their warming-up pace.

2) The interval run consisted of 4 laps (4x400 meters) warming-up pace on the athletic track, directly followed by 5 intervals of 1000 meters on the athletic track, with 2-minute breaks in between intervals. During the break, participants were free to either jog or walk, ensuring they maintained the same activity during each break. After the break, they either returned to where they had finished the previous interval (thus walk/jog forward and then come back to the interval finishing line) or started at the place where they had begun the 1000 m (thus walk/jog 200 m forward to the interval starting line). The interval run pace was the same as the 5 km sub-maximal pace.
3) 5 km sub-maximal run consisted of 4 laps (4x400 m) warming-up pace on the athletic track, followed by a 5 km run (2 laps) on the predefined route on campus (see Figure 3). After the warming-up, they could use a 2-minute break. They run their 5 km sub-maximal pace.

During the second set of experiments, participants were asked to rate their fatigue level from 0 to 10, based on the Borg's Rate of Perceived Exertion scale (Table 2) each lap during a warm-up and,

- 2 times during each lap of the 5 km and endurance session (0 km and 1,4 km points, see Figure 3), and
- each lap of the interval run.

This was considered an appropriate interval for capturing changes in fatigue without negatively impacting the runner's performance, such as by introducing distractions from more frequent RPE assessments.

Figure 3 Running route, with marked places in red where RPE was collected.

Table 2 RPE scale

RPE	Example
0	No effort
1	Barely any effort
2	Very light effort
3	Easy
4	Comfortable
5	Somewhat difficult
6	Difficult
7	Hard
8	Very hard
9	Extremely hard
10	Maximal exhaustion

3.4. Data processing

Python software (Wilmington, Delaware, US) was used for data processing. The dataset was cleaned by identifying extreme values in the IMU sensor data and replacing them with interpolated values, to reduce noise and ensure more accurate model predictions. Specifically, values exceeding predefined thresholds for quaternions (1.1), accelerometer readings (200 m/s²), and gyroscope readings (3000 °/s) were flagged as outliers. Running gait segmentation was performed by detecting peaks in the vertical tibial accelerometer data for both legs and each subject and trial. Peaks were identified based on a minimum peak height of 40 m/s², and a minimum distance calculated from a typical running cadence (105 steps per minute). The detected peaks were used to define the boundaries of individual gait cycles. Then the IMU data was decimated from 120 Hz to 1 Hz by averaging every 120 samples, to reduce the ML model's complexity and eliminate potential noise that could lead to overfitting. In practice, providing feedback at 1-second intervals is more realistic for real-time fatigue monitoring, as runners do not require updates more frequently.

The number of possible sensor combinations is 127 ($2^n - 1, n = 7$). Because it would be too computationally demanding to analyze all of them in this research, four different sensor configuration categories were defined (see Table 3). This was done based on the assumption that some IMU sensors are more valuable than others [34].

Data was labeled using the collected RPE data at a given time point; thus, the target variable in the dataset is continuous, ranging from 0 to 10. Samples between known time points were labeled using a forward-fill method.

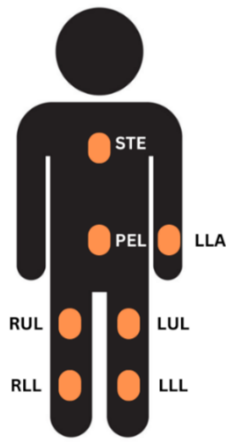

Figure 4 IMU codes: STE-sternum, PEL-pelvis, LLA-left lower arm, RUL-right upper leg (right tight), LUL-left upper leg (left tight), RLL-right lower leg (right tibia), LLL-left lower leg (left tibia).

Table 3 Different sensor configurations

1 sensor	2 sensors	3+sensors	Whole body
LLA	RLL+LLL	PEL+LLL+RLL	STE+PEL+LLA+LLL+RLL+RUL+LUL
LLL	RUL+LUL	PEL+LUL+RUL	
RLL		STE+ PEL+LLL+RLL	
LUL		LLA+ STE+ PEL+LLL+RLL	
RUL		PEL+LUL+RUL+LLL+RLL	
PEL			
STE			

3.5. Feature extraction

A total of 221 features (Table 4) were extracted for each subject, including 91 features from IMU sensors across 7 body segments (left and right tibia, left and right thigh, pelvis, sternum, and wrist), 105 statistical features from acceleration data calculated from each gait cycle, 6 features from a smartwatch, 4 from questionnaires, 10 from general subject information, and 5 other (e.g., time, RPE). The final labeled dataset included 30628 samples, while the final unlabeled dataset had 70493 samples.

Right lower arm features were renamed and merged with left lower arm features to reduce dataset complexity. Non-numerical features were converted into numerical ones. In questionnaire data: 'Don't know' was changed to ‘0’, 'Strongly disagree' to ‘1’, 'Disagree' to ‘2’, 'Neutral' to ‘3’, 'Agree' to ‘4’, 'Strongly agree' to ‘5’. Trial numbers were defined: endurance as ‘1’, interval as ‘2’, and 5 km as ‘3’, to capture the influence of different running scenarios on ‘RPE’. Moreover, 'non-heel' was defined as ‘0’, 'Heel' as ‘1’, 'Male' as ‘0’, and 'Female' as ‘1’.

The IMU sensor on the left tibia failed to record during the interval run of the S16 participant, and half of the endurance run of participant S19. For this reason, missing values in the dataset were replaced by zeros.

Table 4 Extracted features from all data sources in this study.
[2] STD = standard deviation; IQR = inter-quartile range; Skew = skewness; Kurt = kurtosis

Body segments						
STE	**PEL**	**LLA**	**RUL**	**LUL**	**RLL**	**LLL**
IMU features						
Quaternions X,Y,Z	Quaternions X,Y,Z	Quaternions X,Y,Z	Quaternions X,Y,Z	Quaternions X,Y,Z	Quaternions X,Y,Z	Quaternions X,Y,Z
Gyroscope X,Y,Z	Gyroscope X,Y,Z	Gyroscope X,Y,Z	Gyroscope X,Y,Z	Gyroscope X,Y,Z	Gyroscope X,Y,Z	Gyroscope X,Y,Z
Acceleration X,Y,Z	Acceleration X,Y,Z	Acceleration X,Y,Z	Acceleration X,Y,Z	Acceleration X,Y,Z	Acceleration X,Y,Z	Acceleration X,Y,Z
Magnetometer X,Y,Z	Magnetometer X,Y,Z	Magnetometer X,Y,Z	Magnetometer X,Y,Z	Magnetometer X,Y,Z	Magnetometer X,Y,Z	Magnetometer X,Y,Z
Statistical features						
Mean acceleration	Mean acceleration	Mean acceleration	Mean acceleration	Mean acceleration	Mean acceleration	Mean acceleration
STD[2] acceleration	STD acceleration	STD acceleration	STD acceleration	STD acceleration	STD acceleration	STD acceleration
IQR acceleration	IQR acceleration	IQR acceleration	IQR acceleration	IQR acceleration	IQR acceleration	IQR acceleration
Skew. acceleration	Skew. acceleration	Skew. acceleration	Skew. acceleration	Skew. acceleration	Skew. acceleration	Skew. acceleration
Kurt. acceleration	Kurt. acceleration	Kurt. acceleration	Kurt. acceleration	Kurt. acceleration	Kurt. acceleration	Kurt. acceleration
Smartwatch features						
Latitude, Longitude, Altitude Meters, Distance Meters, Speed, Heart Rate						

General subject information
Age, Gender, Weight (kg), Height (cm), Foot strike pattern, HR resting HR max, Experience (years), Km they run per week
Questionnaire features
Before readiness level, before tiredness level, before stiffness level, before muscle soreness
Other features
Left-right stride difference, RPE, Subject number, Time, type of run

3.6. Machine learning pipeline

To optimize and evaluate the Random Forest Regressor, this study used a nested Leave-One-Subject-Out (LOSO) cross-validation framework (Figure 5). The outer LOSO loop assessed the model's generalization performance, while an inner LOSO loop within each outer fold handled hyperparameter tuning and feature importance extraction.

In each iteration of the outer loop, one subject was excluded as the test set, while the remaining subjects were split in the inner loop, with one subject excluded for validation and the rest used for training. RandomizedSearchCV with 10-fold cross-validation was then employed to perform hyperparameter optimization. In other words, testing 10 different hyperparameter combinations for each inner fold split. The hyperparameters tested included 'n_estimators', 'min_samples_split', 'min_samples_leaf', 'max_features', and 'max_depth'. The model with the lowest MSE in the inner loop was selected as optimal. Then that model was trained on the twenty most relevant features from the combined training and validation data and subsequently tested on the subject left out in the outer loop. This was repeated until all 5 subjects were used as a test set once. The Mean Squared Error (MSE) was calculated for each outer fold to assess model performance. These MSE values were then aggregated and averaged to obtain the overall model performance in predicting RPE. This nested LOSO approach ensures that hyperparameter tuning is performed independently of the test set, effectively preventing data leakage and estimating model performance.

The described machine learning pipeline was applied to all IMU sensor combinations. One of the five models that performed the best in the outer loop, across all IMU configuration groups, was then trained on the entire labeled dataset and tested on the unlabeled dataset. This allowed for visually evaluating the model's performance on completely unseen data.

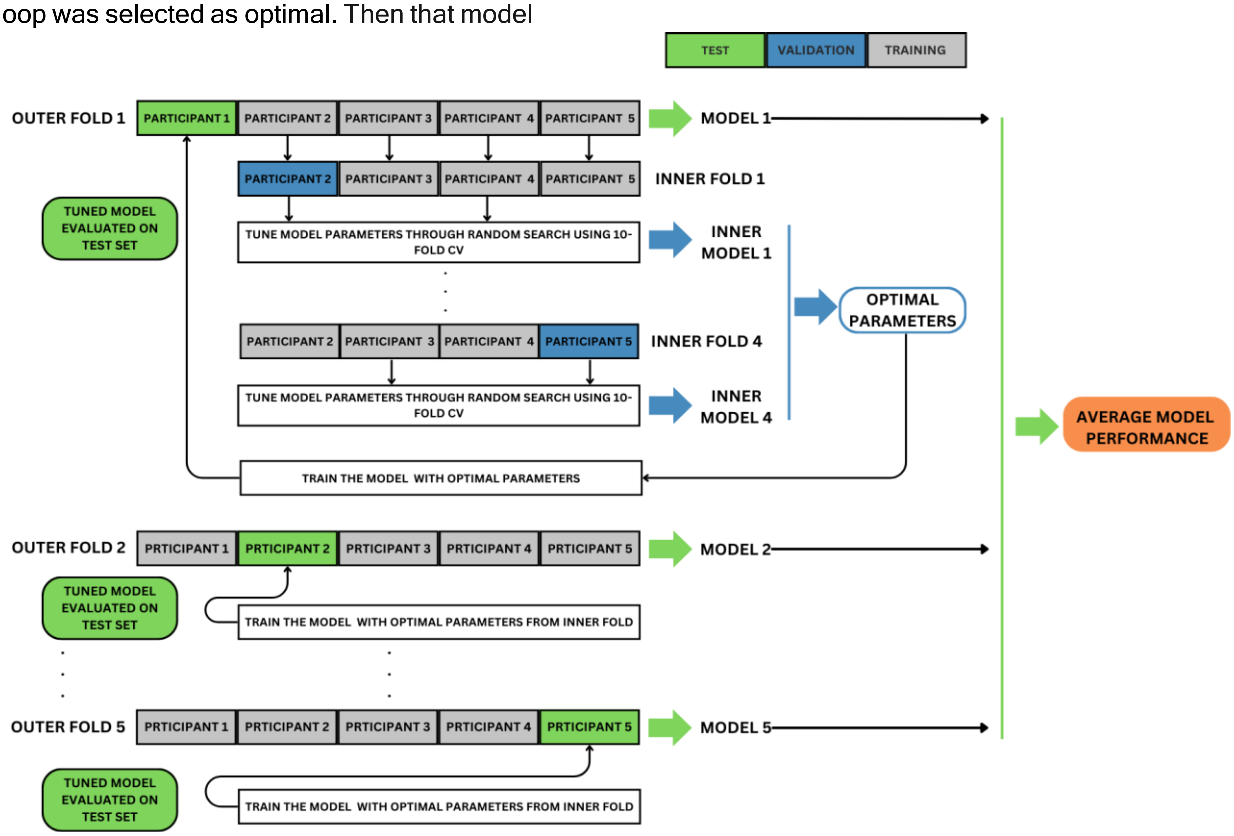

Figure 5 Nested Leave one Subject Out Cros Validation diagram.

4. Results

This chapter outlines the results of the experiments. The first section evaluates HR and RPE trends across different running scenarios and investigates correlations between RPE and various features to identify factors influencing fatigue perception. The second section presents machine learning model results of different IMU sensor configurations.

4.1. Exploratory data analysis

Mean RPE and HR increased across all three runs (Figure 6), with a slight decrease in both measures after the warmup in the 5 km run during a 2-minute rest, as well as during rest periods after running intervals in the interval run. The 5 km run demonstrated the highest increase in perceived exertion due to its high intensity, with an RPE range from 0 to 9 (median 4), making it the most challenging run of the three (Appendix A Figure 13). The interval run showed moderate RPE levels with a range from 0 to 8 (median 4), indicating more consistent exertion among participants. In contrast, the endurance run showed a more gradual increase in RPE, with a range from 0 to 5 (median 3), making it the least demanding trial overall. The lower variability in RPE during the endurance run suggests it was less taxing and more steady-paced compared to the other trials. These results align with expectations for each running scenario, demonstrating that the experimental setup and protocol were effective.

The correlations between RPE and various subject characteristics, as well as kinematic features derived from IMU sensor data, were calculated to better understand the factors influencing perceived exertion (Appendix A, Figures 15, 16, and 17). These correlations were obtained by applying Pearson's correlation coefficient to the data, where RPE was the target variable, and other features were predictors. Below are the key findings with Pearson's correlation coefficient indicated in brackets:

Subject characteristics and smartwatch features:

Distance (0.50) and Heart Rate (0.34) show a positive correlation with RPE, therefore as distance and heart rate increase, so does RPE. At the same time, pre-exercise factors like before tiredness level (0.36) and before stiffness level (0.27) similarly contribute to higher fatigue levels. The positive correlation with gender (0.35) suggests women generally reported higher exertion than men. In contrast, negative correlations with age (-0.48), weight (-0.46), and km run per week (-0.47) imply that older, lighter, and more trained participants perceive less exertion. Trial type (0.31) positively correlates with RPE, suggesting that different exercise formats influence fatigue levels. In other words, participants experienced higher fatigue levels during the 5 km run than endurance. Additionally, heel foot strike pattern (-0.26) and greater before readiness level (-0.26) correlate with reduced RPE, suggesting that non-heel strikers tend to experience higher RPE levels.

Kinematic features from IMU sensors:

The right tibia (RLL) shows notable associations with increased exertion, particularly through mean Z-axis acceleration (-0.32) and high variability across all axes (STD X: 0.30, Y: 0.32, Z: 0.25). Similarly, the left tibia (LLL) is relevant, with reduced mean acceleration in the X (-0.20) and Y (-0.25) axes, reflecting increased fatigue.

For the right tight (RUL), the model shows strong correlations with RPE through negative mean Z-axis acceleration (-0.41), high variability (IQR Z: 0.34, STD Z: 0.28), and asymmetry in the Y-axis (SKEW Y: -0.27). The left tight (LUL) follows a similar trend with negative mean Z-axis acceleration (-0.39), moderate variability (IQR Z: 0.29), and asymmetry in the Y-axis (SKEW Y: -0.24).

The left lower arm (LLA) demonstrates a positive correlation in the Y-axis mean acceleration (0.24), suggesting that lateral arm movement increases with exertion. The pelvis (PEL) and sternum (STE) also reveal significant findings, where the Z-axis variability (IQR PEL Z: 0.37, IQR STE Z: -0.31) indicates that stability in these areas correlates with lower perceived exertion. Additionally, Z-axis variability in the sternum (STD Z: -0.20) further supports the importance of upper body stability in managing exertion.

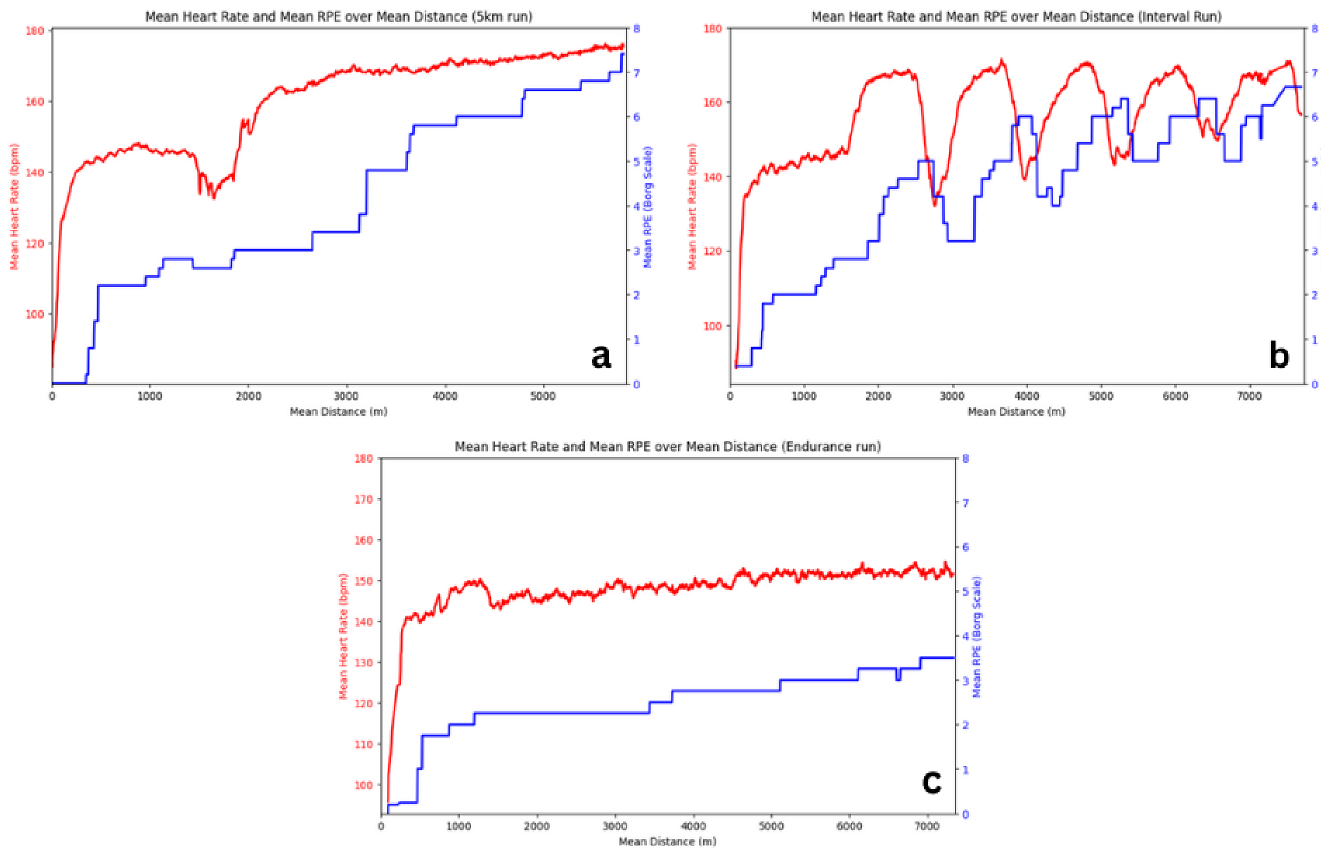

Figure 6 HR (red) and RPE (blue), over distance, averaged for all 5 subjects for a) 5 km sub-maximal, b) interval, and c) endurance runs.

4.2. Machine learning model results

The regression analysis in this chapter aimed to identify the best-performing sensor configurations for predicting RPE, comparing single, dual, and multi-sensor setups to determine the optimal approach for real-life scenario fatigue monitoring.

4.2.1. Model tuning

The nested LOSO cross-validation approach allowed for tailored model tuning across different sensor configurations, optimizing parameters for each sensor setup and outer fold. Table 5 summarizes the best hyperparameters for each configuration. Table 6 details fold-specific hyperparameters for the LLA configuration, reflecting how model tuning varied across outer folds to adapt to differences between subjects.

4.2.2. Feature Importance Across Configurations

To understand which features drive RPE prediction, Table 7 ranks the top 20 features of the best model of each configuration chosen by the Random Forest model. DistanceMeters consistently emerged as the most important feature, with HeartRate and Speed also ranking highly across all setups. In the whole-body setup, specific motion features like Mean_RLL_Acc_Z and Std_LLL_Acc_Z were particularly relevant. Other influential features included demographic and trial-related variables such as Age, Km/week, and Trial_numeric.

4.2.3. Model performance evaluation

Table 8 shows the average model performance metrics for the IMU sensor setup with the lowest average MSE within each sensor configuration category. The single-sensor configuration using the left lower arm (LLA) achieved the best overall performance, with an average MSE of 1.89. In contrast, the two-sensor combination (RUL+LUL) produced a slightly higher average MSE of 2.26, while the three-plus sensor setup (PEL+LUL+RUL+LLL+RLL) had the highest average MSE of 2.44. Interestingly, the whole-body configuration, while incorporating most sensors, improved performance slightly compared to other multi-sensor configurations with an MSE of 2.16, but did not outperform the LLA sensor alone. Across all configurations, the model yielded the best average MSE results for the endurance running session, and the worst for 5 km sub-maximal sessions.

Table 5 Hyperparameters for the IMU sensor setup model with the lowest overall MSE across outer folds for each sensor configuration category.

LLA	RUL+LUL	PEL+LUL+RUL+LLL+RLL	Whole body
'n_estimators':100, 'min_samples_split':10, 'min_samples_leaf': 4, 'max_features': 'sqrt', 'max_depth': 15	'n_estimators':150, 'min_samples_split':10, 'min_samples_leaf':2, 'max_features':'sqrt', 'max_depth': 10	'n_estimators': 150, 'min_samples_split': 15, 'min_samples_leaf': 4, 'max_features': 'sqrt', 'max_depth': 15	'n_estimators':100, 'min_samples_split':10, 'min_samples_leaf':4, 'max_features':'sqrt' 'max_depth': 15

Table 6 Hyperparameters for the LLA sensor configuration model with the lowest inner fold MSE, for each outer fold.

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
'n_estimators':100 'min_samples_split':10 'min_samples_leaf': 4 'max_features':'sqrt' 'max_depth': 15	'n_estimators':200 'min_samples_split':5 'min_samples_leaf':8 'max_features':0.5 'max_depth':15	'n_estimators':100 'min_samples_split':10 'min_samples_leaf':2 'max_features':0.5 'max_depth': 10	'n_estimators':150 'min_samples_split': 5 'min_samples_leaf':4 'max_features':'sqrt' 'max_depth':15	'n_estimators':100 'min_samples_split': 15 'min_samples_leaf':4 'max_features':0.5 'max_depth':None

Table 7 Twenty most important features per sensor configuration.

Feature rank	LLA	RUL+LUL	PEL+LUL+RUL+LLL+RLL	Whole body
1	DistanceMeters	DistanceMeters	DistanceMeters	DistanceMeters
2	Trial_numeric	Speed	HeartRate	Mean_RLL_Acc_Z
3	HeartRate	HeartRate	Speed	HeartRate
4	Age	Age	Std_RLL_Acc_X	Mean_LLL_Acc_Z
5	Km/week	Mean_RUL_acc_Z	Km/week	Mean_RUL_Acc_Z
6	Speed	Km/week	Age	Std_RLL_Acc_X
7	AltitudeMeters	Trial_numeric	Mean_RUL_Acc_Z	Speed
8	Std_LLA_Acc_Z	Mean_LUL_Acc_Z	Mean_RLL_Acc_Z	Std_LLL_Acc_Z
9	mean_LLA_Acc_Y	Skew_LUL_Acc_Y	Trial_numeric	Iqr_LLL_Acc_X
10	Iqr__LLA_Acc_Y	Mean_RUL_Acc_X	Mean_LLL_Acc_Z	Std_LLL_Acc_X
11	Mean_LLA_Acc_X	RUL_Acc_Z	Std_PEL_Acc_Y	Height (cm)
12	std _LLA_Acc_X	Mean_LUL_Acc_Y	Skew_LUL_Acc_Y	Std_PEL_Acc_Y
13	HR max	Iqr_RUL_Acc_X	Mean_PEL_Acc_Y	Std_LLA_Acc_Z
14	iqr _LLA_Acc_X	Iqr_LUL_Acc_X	Iqr_RUL_Acc_X	Weight (kg)
15	kurtosis _LLA_Acc_X	AltitudeMeters	Mean_LLL_Acc_Y	Mean_LLA_Acc_Y
16	kurtosis _LLA_Acc_Z	RUL_Quat_X	Mean_LUL_Acc_Z	Km/week
17	skew _LLA_Acc_Z	HR max	AltitudeMeters	Age
18	mean _LLA_Acc_Z	Latitude	Latitude	Skew_LUL_Acc_Y
19	iqr _LLA_Acc_Z	Before_muscle_sorness	HR max	Iqr_LLA_Acc_Z
20	skew _LLA_Acc_Y	Longitude	Before_muscle_soreness	Skew_LLA_Acc_Z

Table 8 Average model performance metrics for the IMU sensor setup with the lowest average MSE within each sensor configuration category.

Category	Best IMU combination	Average MSE	Average MSE per trial
1 sensor	LLA	**1.89±1.59**	5K: 2.07±1.98 END: **1.58±1.22** INT: 2.05±1.56
2 sensors	RUL+LUL	2.26±1.36	5K: 2.90±1.72 END: 1.75±0.94 INT: 2.08±1.41
3+sensors	PEL+LUL+RUL+LLL+RLL	2.44±1.36	5K: 3.25±1.66 END: 1.55 ±0.97 INT: 2.44±1.44
Whole body	STE+PEL+LLA+LLL+RLL+RUL+LUL	2.16±1.65	5K: 2.30 ±1.71 END: 1.68 ±1.53 INT: 2.39±1.71

Figure 7 shows the MSE distribution across outer folds for both validation (inner_fold_mse) and test (outer_fold_mse) sets, providing insights into model consistency across subjects. Furthermore, omitting additional subject-specific information (e.g., age, weight, height) and pre-exercise questionnaire responses resulted in notable performance decreases, as seen in Figure 8. This performance is expected given the feature importance of subject information that was previously discussed in Chapter 4.1. Furthermore, it can be observed that the differences in performance between LLA, RUL, and RUL+LUL sensor configurations are minimal. The RUL sensor configuration is particularly interesting, with consistent stability in its results, demonstrating the smallest MSE difference, with and without subject-specific information and questionnaire data.

To further investigate the dynamics captured by these sensors, Figure 9 displays the acceleration data and its rolling variance across the X and Y axes for the LLA, RUL, and LUL sensors during the 5 km trials, averaged across 5 subjects. Focusing on acceleration data is motivated by prior correlation analyses, which showed that acceleration metrics had the highest correlation with the RPE reported by participants. This analysis revealed distinct trends, especially during the 5 km trial. For the LLA sensor, a notable decreasing linear trend was observed in the X-axis acceleration. Likewise, the RUL sensor showed a consistent decrease in Z-axis acceleration, paralleling the trends observed in the LUL measurements, and supporting findings of diminishing acceleration as fatigue progressed. Additionally, variability trends increased across all sensors, particularly towards the end of the run, indicating more erratic movement patterns as fatigue set in.

Lastly, the predictive performance of the LLA configuration model trained on the labeled data from 5 subjects and tested on the unlabeled dataset of 12 subjects is visualized in Figure 10.

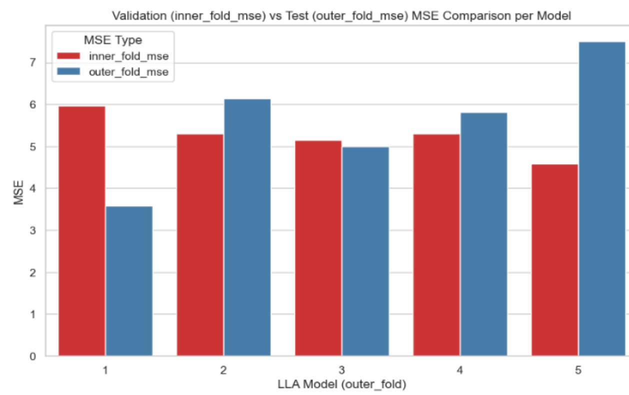

Figure 7 LLA configuration MSE results for each outer fold and the average MSE across all inner folds.

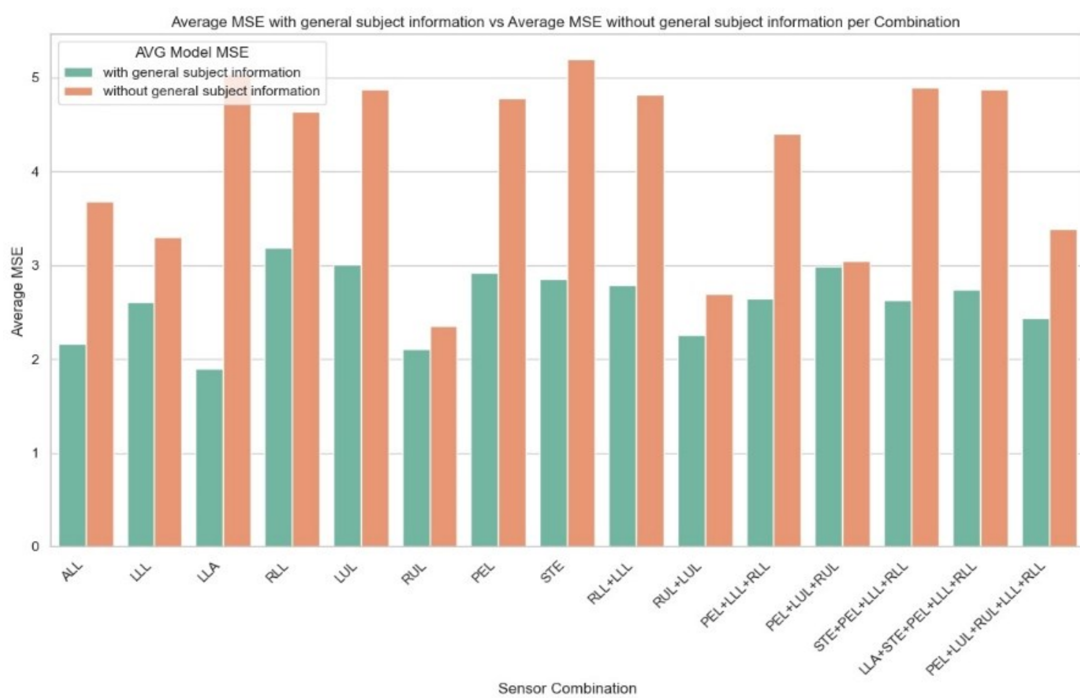

Figure 8 Average model MSE for all sensor combinations, with and without general subject information (e.g., age, weight, height, etc.), and before the run questionnaire.

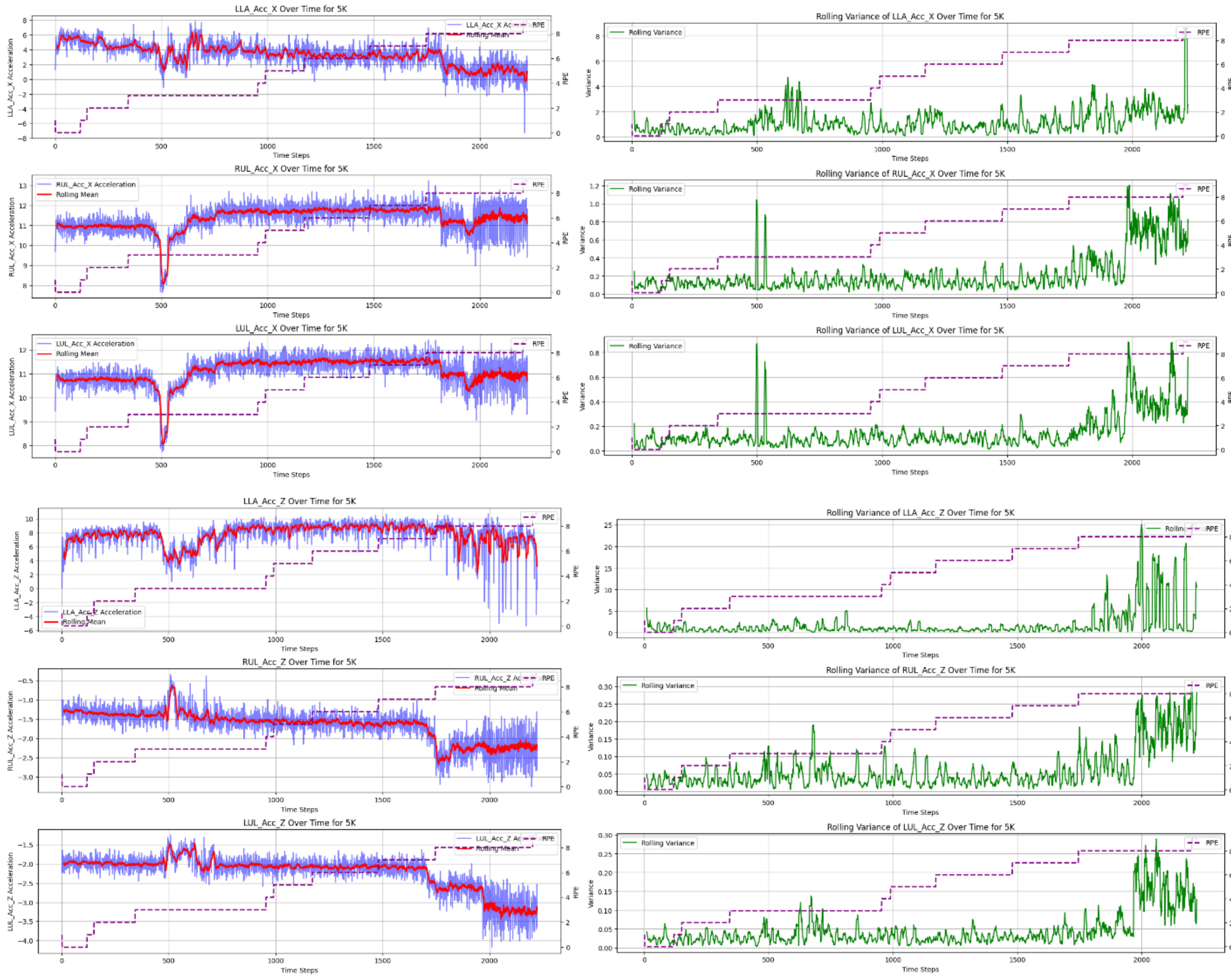

Figure 9 Acceleration data and rolling variance across the X and Z axes for the left lower arm (LLA), right upper leg (RUL), and left upper leg (LUL) sensors during the 5 km trials, averaged across 5 subjects.

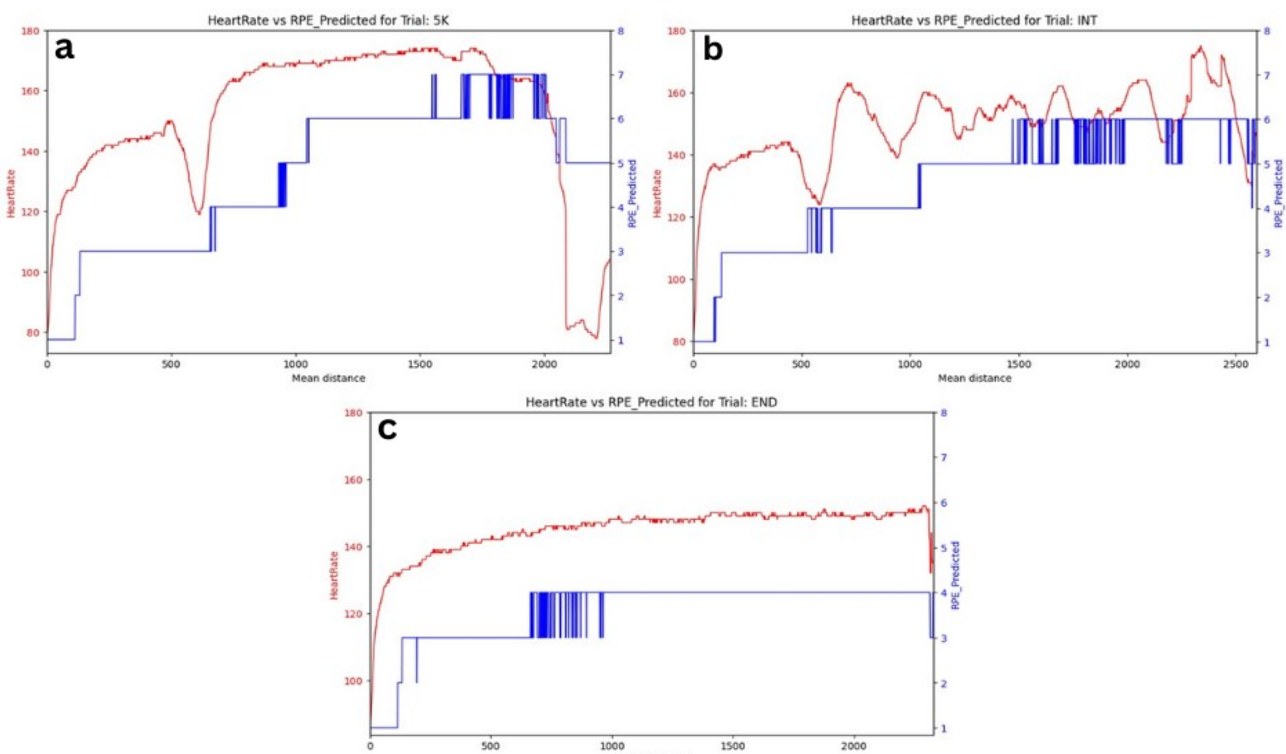

Figure 10 Mean RPE prediction and HR values of 12 subjects, and 3 trials: a) 5 km sub-maximal, b) interval, c) endurance, using the best LLA sensor configuration model.

5. Discussion

The primary goal of this study was to develop a machine learning-based system to predict running-induced fatigue within 1-second intervals during outdoor recreational running sessions, using data from wearable sensors such as IMUs, heart rate monitors, and smartwatches. Fatigue was quantified through the Borg Rating of Perceived Exertion (RPE) scale across endurance, interval, and 5 km running scenarios.

Current literature on outdoor running-induced fatigue monitoring is limited. Most studies are cross-sectional [33-35] rather than longitudinal and often rely on two- or three-state classifications of fatigue (e.g., fatigued vs. non-fatigued) [20,33-35], which oversimplifies its complex nature [20]. Furthermore, these studies are typically conducted on a single running surface, such as an athletic track, over shorter distances within controlled scenarios [33-36], and lack the variability of realistic training conditions. Individual characteristics such as age, fitness level, and training history are frequently overlooked, with a predominant focus on lower-body sensors and minimal attention to upper-limb data [33-35].

This study addresses these shortcomings by employing a broader range of commercially available, wearable sensors, for capturing data across different running scenarios (endurance, interval, and 5 km) and surfaces (e.g. athletic track, road), enabling continuous fatigue tracking beyond controlled environments. By integrating individual subject characteristics (e.g., age, fitness level) and pre-run questionnaire data assessing subjective readiness, this study captures both physiological and situational factors contributing to fatigue, enhancing model accuracy in real-world conditions. Unlike prior studies with simplified fatigue classifications, this study uses a 0-10 scale for a more detailed analysis of fatigue progression. Furthermore, it employs a nested Leave-One-Subject-Out Cross-Validation framework, a robust validation method often absent in similar studies [20]. Together, these features make this study a unique and valuable contribution to the field, addressing gaps in the current literature.

Direct comparisons across studies are challenging due to variations in sensor placement, sensor quantity, and different ML models used. The study most comparable to ours is De Beeck et al., which is also longitudinal, focuses on outdoor running, employs a 0-10 RPE regression scale, and incorporates a wrist-worn IMU. Consistent with this study's findings, they identified the wrist IMU as the most effective single-sensor configuration [36]. However, their study lacks multiple running scenarios and surfaces, subject-specific data (e.g., age, weight, height), and heart rate integration, all of which are included in this paper's approach.

This study shows the variability in model performance across different running trials – where the endurance run yielded the most accurate predictions, followed by interval and 5 km runs. The superior performance in endurance trials can be attributed to the steady pacing and rhythm maintained throughout the run. The more constant RPE values observed in the endurance trials provided a stable target for the model to learn from. In contrast, 5K trials introduce higher pacing variability, complicating the prediction model's accuracy. While interval trials are more structured with alternating effort and rest phases, they still involve significant variability compared to endurance runs.

The prominence of Distance across all sensor configurations suggests that the total distance covered during a run is a key factor in predicting fatigue. This aligns with the general understanding that physical and mental exhaustion also rise as running distance increases [43]. The significance of Heart Rate and speed further emphasizes the role of physiological load in fatigue development. Higher heart rates and running speeds are commonly associated with increased exertion [36], making these metrics valid indicators of fatigue. In the whole-body sensor setup, lower limb acceleration features, especially in the sagittal plane (Z-axis), were highly ranked, indicating that biomechanical changes are important for identifying fatigue. For example, the tibial IMU Z-axis in this setup corresponds to the sagittal plane, it captures changes in stride dynamics, such as propulsion and braking forces, which tend to vary as fatigue progresses, making the Z-axis sensitive to adjustments in running efficiency.

Moreover, the decrease in model performance after omitting the before-run questionnaire such as tiredness, readiness, and muscle stiffness levels, and general subject information, such as age, gender, and training history, suggests that these factors play an important role in enhancing prediction accuracy. These factors allowed the model to account for individual differences in physical capabilities, tailoring fatigue predictions to each subject's unique characteristics and significantly enhancing overall performance. While time of day could also influence fatigue levels, this feature was not implemented in the current study due to dataset imbalances. Most 5K runs were conducted in the morning, which could have skewed the results. Future research should explore a more balanced dataset by including data collected at different times of day (morning, afternoon, evening) in equal proportions to evaluate the potential impact of time of day on fatigue predictions.

In this study, each model in the Leave-One-Subject-Out Cross-Validation (LOSO CV) framework was tuned to different hyperparameters. This variation in hyperparameters occurred because each fold excluded a different subject, resulting in unique training data for each iteration. As a result, the model adapted its complexity and feature selection based on the distinct patterns, characteristics, and variability of the remaining subjects. This indicates that no single model was universally superior, but rather, the model is adjusting to the individual characteristics of subjects. The challenge of subject-independent fatigue detection comes from the variability in motion signals, which can differ both between individuals and within the same individual across different trials [20]. Furthermore, the outer and inner fold MSE varies across models, as shown in Figure 7, due to differences in the training and test data for each fold, primarily influenced by the distinct test subjects used. This variability is especially pronounced in the fifth model, where a notable discrepancy between the outer and inner fold MSE arises from data quality issues. In this case, the test subject's data was compromised by a fall during recording, introducing inaccuracies.

The findings indicate that the wrist sensor configuration alone achieved the highest prediction accuracy (MSE = 1.89), outperforming multi-sensor setups, suggesting that more sensors do not necessarily improve model performance, potentially due to an increase in noise and complexity. Given the minimal difference in error rates between the models, however, the value of this comparison is limited. Notably, as RPE increased, the wrist sensor's X-axis (forward-backward arm swing) acceleration linearly decreased while variance rose, suggesting reduced arm drive and more erratic movement with fatigue. This change likely provided identifiable patterns, making it easier for the model to capture fatigue progression. Next, it is crucial to note marginal differences in performance among the models utilizing LLA, RUL, and the combined RUL + LUL sensors. RUL sensor demonstrates the smallest error difference when used with and without subject information, highlighting the strong importance of its features, which further aligns with correlation findings in Chapter 4.1. Therefore, while the LLA sensor demonstrated competitive performance in terms of error measurement, the stability of results provided by the RUL sensor – its strong correlation with RPE and minimal error differences regardless of subject information – positions it as a potentially superior choice in this analysis. The observed trends in acceleration further support the argument that RUL may be the best sensor in this context. Nevertheless, these findings demonstrate that machine learning can effectively be used to predict running-induced fatigue in outdoor settings with varied training sessions (research question 1), indicating that single IMU setups (wrist, right tight) can yield low prediction errors, potentially reducing the need for multiple sensors (research question 2). Further research should focus on refining sensor selection and placement to optimize fatigue prediction and model reliability. Moreover, the required MSE threshold remains subjective, as accuracy expectations may vary. For example, experienced runners may demand higher precision, while others may find moderate accuracy sufficient. Additionally, it is worth questioning if 1-second sampling is overly frequent; reducing the sampling interval could likely reduce errors, potentially enhancing model stability without sacrificing useful insights.

Lastly, the predictive performance of the LLA configuration model, tested on the unlabeled dataset of 12 subjects, again shows that the endurance trial shows the most stable predictions, whereas the model struggled more with the interval and 5K trials, which involve greater variability in effort. Good validation is indicated by consistent patterns, such as predicted RPE increasing with elevated HR, and stable predictions without erratic fluctuations. Although the endurance trial yielded better estimations, fluctuations are still present. Overall, the results are visually reasonable, especially considering the limited training data, indicating the model's potential to generalize across subjects.

5.1. Challenges and limitations

Despite the promising results of this study, several limitations need to be considered:

Small sample size

The study's sample size of five subjects limits the generalizability of the findings, as the model may not fully capture the variability in running-induced fatigue across a broader population. Individual differences such as weight, height, fitness level, readiness, and training background introduce unique running styles that affect key biomechanical parameters like step length, step frequency, and arm movement [36]. With a larger sample, the model could better account for these inter-individual variations, improving its ability to generalize to diverse groups of runners. Additionally, a larger dataset would enhance model reliability by reducing overfitting, as it would provide more data points for training and validation. In such a scenario, the wrist (LLA) sensor configuration, which performed best in this study, might not remain the most effective option. Furthermore, a larger sample would likely yield more consistent hyperparameters across folds, allowing the model to generalize more effectively.

RPE scale

While the machine learning model used various objective and subjective input data, the target variable (RPE scale) represents a subjective fatigue estimate. First, individual differences in interpreting exertion mean that each runner's perception of fatigue can vary, making it difficult to capture subtle, gradual changes. Some participants struggled with understanding the RPE scale, finding it unintuitive; they were often unsure about which number to report and tended to select a value that seemed expected or logical rather than accurately reflecting their true exertion. Additionally, as participants became more familiar with the scale across tests, their ratings likely evolved, potentially introducing further variability. Another limitation is that RPE was recorded only at specific points during each running trial, while the model predicts fatigue every second. This discrepancy between the prediction frequency and the actual RPE data collection may have impacted the model's accuracy, as the target data lacked the granularity needed to align with the model's continuous predictions.

Lack of controlled fatiguing protocol

Unlike many fatigue studies, this research did not include a controlled fatiguing protocol, which is typically used to ensure that each participant reaches a comparable level of fatigue under standardized conditions. While the used approach reflects real-world conditions more accurately, it results in an imbalanced dataset, as high fatigue states are underrepresented relative to moderate or low fatigue levels. This imbalance can limit the model's ability to predict a full range of fatigue levels.

Technical limitations

A technical limitation occurred with the failure of the left tibial (LLL) sensor for one subject during the interval run and for another during half of the endurance run. Imputation of the missing data was necessary, which may have introduced bias into the analysis.

Random Forest Regressor

While Random Forests handle complex, non-linear relationships effectively, they are difficult to interpret due to their complexity, complicating efforts to explain the model's decision-making. Random Forests are also computationally demanding, which may limit usability in real-time applications like continuous fatigue monitoring. Overfitting, especially in small datasets, poses another challenge, as it may cause the model to capture noise instead of meaningful patterns, reducing its generalizability. Additionally, the algorithm may introduce bias in feature importance, favoring continuous variables like distance or heart rate, and may not perform well in cases where fatigue data is imbalanced (e.g., fewer instances of high-fatigue levels) [37][38], such as the case in this research.

Sensor combinations and placement variability

Lastly, only 15 out of 127 potential IMU sensor combinations were tested in this study. Furthermore, consistent sensor placement across participants and sessions is challenging. Variations in placement, both among different participants and across sessions for the same participant, may introduce inconsistencies in the data.

5.2. Applications and Future Research

The algorithm developed in this study shows potential for real-world applications of fatigue monitoring devices in outdoor recreational running. This allows athletes and coaches to receive real-time feedback on fatigue levels, helping to adjust training loads, prevent overuse injuries, and optimize recovery plans.

Recent advances in commercial fitness devices, such as Garmin's Performance Condition [44] and Apple Watch's Training Load [45], highlight the growing interest in fitness monitoring and fatigue detection. While these devices work based on metrics like heart rate, VO2 Max, age, height, weight, and GPS data, details of their algorithms remain proprietary. The presented study explores the algorithm's ability to function with a minimal sensor setup, for potentially enabling integration into commercial smartwatches for a wider audience of recreational runners.

In addition to real-time fatigue monitoring, integrating post-run and day-after questionnaires could further enhance the algorithm's predictive capabilities, and provide more personalized feedback. Incorporating this information could refine the RPE estimates, thus helping athletes make more informed decisions about their recovery and readiness for the next running sessions [46]. Additionally, integrating advanced biometric data, such as sleep patterns, and stress levels, could be beneficial for providing meaningful insights into an athlete's overall condition. Further research could also explore the long-term predictive capabilities of this system by incorporating historical data from multiple training sessions.

Future studies should also consider collecting RPE data at more frequent intervals to improve the model's responsiveness to real-time changes in fatigue. While this study's approach is inherently reactive, capturing fatigue after it has already impacted performance, an ideal model would predict performance-based fatigue indicators first, and then use those to estimate perceived fatigue levels. This proactive framework could be achieved by leveraging time-series forecasting models, like attention-based Transformers, to predict short-term fatigue trends in real-time [13]. Moreover, unsupervised learning techniques could identify fatigue patterns without explicit RPE labeling, supporting a non-intrusive, predictive solution.

Future research should explore additional sensor configurations to find an optimal balance between sensor count and predictive accuracy. For practical applications where tibial sensors might be excluded, alternative methods for gait cycle extraction should be considered.

Finally, it would be best to validate the effectiveness of IMU-based techniques by comparing them against established objective methods for physical fatigue assessment, such as electromyography (EMG) or maximal oxygen consumption (VO2 max).

6. Conclusion

This study demonstrates the potential of machine learning algorithms, combined with wearable sensors, for predicting running-induced fatigue in outdoor settings. With IMUs, heart rate monitors, and smartwatches, the system provides non-intrusive monitoring that could help prevent injuries and optimize recovery strategies for runners. Notably, the best-performing sensor setup, using a single left wrist (LLA) sensor, achieved an average MSE of 1.89, outperforming more complex multi-sensor configurations. This result highlights the feasibility of reducing the sensor setup and the potential for integration into consumer fitness devices like smartwatches. However, the study's small sample size and reliance on subjective fatigue measures, such as the Borg Rating of Perceived Exertion, indicate the need for further research with larger, more diverse populations. Future studies should incorporate additional biometric data and validate the system against gold-standard fatigue assessment methods, such as electromyography (EMG) or VO2 max, to improve the model's generalizability and precision.

7. Acknowledgements

I would like to express my gratitude to all the participants who generously dedicated their time and effort to this study, often running in the coldest, hottest, and rainiest weather conditions. Your commitment was truly commendable. I extend my sincere appreciation to my supervisors for their guidance, insightful feedback, and continuous support throughout the research process. Thank you for giving me a project that was a perfect fit. I am especially grateful to the University of Twente for providing the scholarship that made my studies possible. Lastly, I acknowledge the encouragement and patience of my friends and family. Thank you for your never-ending support.

8. References

[1] Quirk, H., Bullas, A., Haake, S. et al. Exploring the benefits of participation in community-based running and walking events: a cross-sectional survey of parkrun participants. BMC Public Health 21, 1978 (2021). https://doi.org/10.1186/s12889-021-11986-0

[2] Hespanhol Junior, L.C., Pillay, J.D., van Mechelen, W. et al. Meta-Analyses of the Effects of Habitual Running on Indices of Health in Physically Inactive Adults. Sports Med 45, 1455–1468 (2015). https://doi.org/10.1007/s40279-015-0359-y

[3] Vitti, A., Nikolaidis, P. T., Villiger, E., Onywera, V., & Knechtle, B. (2019). The "New York City Marathon": participation and performance trends of 1.2M runners during half-century. Research in Sports Medicine, 28(1), 121–137. https://doi.org/10.1080/15438627.2019.1586705

[4] Videbæk, S., Bueno, A. M., Nielsen, R. O., & Rasmussen, S. (2015). Incidence of Running-Related Injuries Per 1000 h of running in Different Types of Runners: A Systematic Review and Meta-Analysis. Sports medicine (Auckland, N.Z.), 45(7), 1017–1026. https://doi.org/10.1007/s40279-015-0333-8

[5] Fields, Karl B.1; Sykes, Jeannie C.2; Walker, Katherine M.3; Jackson, Jonathan C.4. Prevention of Running Injuries. Current Sports Medicine Reports 9(3):p 176-182, May 2010. | DOI: 10.1249/JSR.0b013e3181de7ec5

[6] Hreljac, A., 2004. Impact and overuse injuries in runners. Medicine & Science in Sports & Exercise, 36(5), pp.845-849.

[7] Arnold, M.J. and Moody, A.L., 2018. Common running injuries: evaluation and management. American family physician, 97(8), pp.510-516.

[8] Menheere D, Janssen M, Funk M, van der Spek E, Lallemand C, Vos S. Runner's perceptions of reasons to quit running: influence of gender, age and running-related characteristics. Int J Environ Res Public Health. (2020) 17:1–12. doi: 10.3390/ ijerph17176046

[9] Desai, P.; Jungmalm, J.; Borjesson, M.; Karlsson, J.; Grau, S. Recreational runners with a history of injury are twice as likely to sustain a running-related injury as runners with no history of injury: A 1-year prospective cohort study. J. Orthop. Sports Phys. Ther. 2021, 51, 144–150.

[10] Hespanhol Junior, L. C., van Mechelen, W., Postuma, E., & Verhagen, E. (2015). Health and economic burden of running-related injuries in runners training for an event: A prospective cohort study. Scandinavian Journal of Medicine & Science in Sports, 26(9), 1091–1099. https://doi.org/10.1111/sms.12541

[11] Sleeswijk Visser TSO, van Middelkoop M, Fokkema T, de Vos RJ. The socio-economic impact of running-related injuries: A large prospective cohort study. Scand J Med Sci Sports. 2021 Oct;31(10):2002-2009. doi: 10.1111/sms.14016. Epub 2021 Jul 11. PMID: 34228834; PMCID: PMC8518541.

[12] Kakouris N, Yener N, Fong DTP. A systematic review of running-related musculoskeletal injuries in runners. J Sport Health Sci. 2021 Sep;10(5):513-522. doi: 10.1016/j.jshs.2021.04.001. Epub 2021 Apr 20. PMID: 33862272; PMCID: PMC8500811.

[13] Jiang Y, Malliaras P, Chen B, Kulić D. Real-time forecasting of exercise-induced fatigue from wearable sensors. Comput Biol Med. 2022 Sep;148:105905. doi: 10.1016/j.compbiomed.2022.105905. Epub 2022 Jul 20. PMID: 35905661.

[14] Marit A. Zandbergen, Luca Marotta, Roos Bulthuis, Jaap H. Buurke, Peter H. Veltink, Jasper Reenalda, Effects of level running-induced fatigue on running kinematics: A systematic review and meta-analysis, Gait & Posture, Volume 99, 2023, Pages 60-75, ISSN 0966-6362, https://doi.org/10.1016/j.gaitpost.2022.09.089.

[15] Edwards, W. Brent. (2018). Modeling Overuse Injuries in Sport as a Mechanical Fatigue Phenomenon. Exercise and Sport Sciences Reviews. 46. 1. 10.1249/JES.0000000000000163.

[16] Bompa, T.O. and Haff, G.G., 2009. *Periodization: Theory and Methodology of Training*. 5th ed. Champaign, IL: Human Kinetics.

[17] Al-Mulla MR, Sepulveda F, Colley M. An Autonomous Wearable System for Predicting and Detecting Localised Muscle Fatigue. Sensors. 2011; 11(2):1542-1557. https://doi.org/10.3390/s110201542

[18] M.J. Pinto-Bernal, A. Aguirre, C.A. Cifuentes, M. Munera, Wearable sensors for monitoring exercise and fatigue estimation in rehabilitation, in: Internet of Medical Things, CRC Press, 2021, pp. 83–110.[4] H. Dong, I. Ugalde, N. Figueroa, A. El Saddik, T

[19] Thorpe, R. T., Atkinson, G., Drust, B., & Gregson, W. (2017). Monitoring Fatigue Status in Elite Team-Sport Athletes: Implications for Practice. International Journal of Sports Physiology and Performance, 12(s2), S2-27-S2-34. Retrieved Sep 4, 2024, from https://doi.org/10.1123/ijspp.2016-0434

[20] Adão Martins NR, Annaheim S, Spengler CM, Rossi RM. Fatigue Monitoring Through Wearables: A State-of-the-Art Review. Front Physiol. 2021 Dec 15;12:790292. doi: 10.3389/fphys.2021.790292. PMID: 34975541; PMCID: PMC8715033.

[21] Theofilidis G, Bogdanis GC, Koutedakis Y, Karatzaferi C. Monitoring Exercise-Induced Muscle Fatigue and Adaptations: Making Sense of Popular or Emerging Indices and Biomarkers. Sports (Basel). 2018 Nov 26;6(4):153. doi: 10.3390/sports6040153. PMID: 30486243; PMCID: PMC6315493.

[22] Bestwick-Stevenson T, Toone R, Neupert E, Edwards K, Kluzek S. Assessment of Fatigue and Recovery in Sport: Narrative Review. Int J Sports Med. 2022 Dec;43(14):1151-1162. doi: 10.1055/a-1834-7177. Epub 2022 Apr 25. PMID: 35468639.

[23] Cleveland Clinic. (n.d.). *Rated perceived exertion (RPE) scale*. Cleveland Clinic. https://my.clevelandclinic.org/health/articles/17450-rated-perceived-exertion-rpe-scale (accessed 14.09.2024.)

[24] Scherr, J., Wolfarth, B., Christle, J.W., Pressler, A., Wagenpfeil, S., and Halle, M., 2012. Associations between Borg's rating of perceived exertion and physiological measures of exercise intensity. Springer-Verlag, DOI 10.1007/s00421-012-2421-x

[25] Paquette, M.R.; Napier, P.C.; Willy, P.R.W.; Stellingwerff, T. Moving Beyond Weekly 'Distance': Optimizing Quantification of Training Load in Runners. *J. Orthop. Sports Phys. Ther.* 2020, 1–20.

[26] Richard V. Milani, Carl J. Lavie, Mandeep R. Mehra, Hector O. Ventura, Understanding the Basics of Cardiopulmonary Exercise Testing, Mayo Clinic Proceedings, Volume 81, Issue 12, 2006, Pages 1603-1611, ISSN 0025-6196, https://doi.org/10.4065/81.12.1603.
[27] Animost. (n.d.). Is motion capture expensive? Retrieved from https://animost.com/ideas-inspirations/is-motion-capture-expensive (accessed 25.07.2024.)
[28] Aroganam G, Manivannan N, Harrison D. Review on Wearable Technology Sensors Used in Consumer Sport Applications. Sensors (Basel). 2019 Apr 28;19(9):1983. doi: 10.3390/s19091983. PMID: 31035333; PMCID: PMC6540270.
[29] Ranavolo, A.; Draicchio, F.; Varrecchia, T.; Silvetti, A.; Iavicoli, S. Wearable Monitoring Devices for Biomechanical Risk Assessment at Work: Current Status and Future Challenges—A Systematic Review. *Int. J. Environ. Res. Public Health* 2018, *15*, 2001. https://doi.org/10.3390/ijerph15092001
[30] V. Camomilla, E. Bergamini, S. Fantozzi, and G. Vannozzi, "Trends Supporting the In-Field Use of Wearable Inertial Sensors for Sport Performance Evaluation: A Systematic Review," Sensors (Basel)., vol. 18, no. 3, Mar. 2018, doi: 10.3390/S18030873.
[31] M. O'Reilly, B. Caulfield, T. Ward, W. Johnston, and C. Doherty, "Wearable Inertial Sensor Systems for Lower Limb Exercise Detection and Evaluation: A Systematic Review," Sports Med., vol. 48, no. 5, pp. 1221-1246, May 2018, doi: 10.1007/540279-018-0878-4.
[32] Jiang Y, Hernandez V, Venture G, Kulić D, K Chen B. A Data-Driven Approach to Predict Fatigue in Exercise Based on Motion Data from Wearable Sensors or Force Plate. Sensors (Basel). 2021 Feb 22;21(4):1499. doi: 10.3390/s21041499. PMID: 33671497; PMCID: PMC7926834.
[33] G. Wang, X. Mao, Q. Zhang and A. Lu, "Fatigue Detection in Running with Inertial Measurement Unit and Machine Learning," *2022 10th International Conference on Bioinformatics and Computational Biology (ICBCB)*, Hangzhou, China, 2022, pp. 85-90, doi: 10.1109/ICBCB55259.2022.9802471.
[34] Marotta, L.; Buurke, J.H.; van Beijnum, B.-J.F.; Reenalda, J. Towards Machine Learning-Based Detection of Running-Induced Fatigue in Real-World Scenarios: Evaluation of IMU Sensor Configurations to Reduce Intrusiveness. *Sensors* 2021, *21*, 3451. https://doi.org/10.3390/s21103451
[35] C. Buckley *et al*., "Binary classification of running fatigue using a single inertial measurement unit," *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, Eindhoven, Netherlands, 2017, pp. 197-201, doi: 10.1109/BSN.2017.7936040.
[36] Tim Op De Beéck, Wannes Meert, Kurt Schütte, Benedicte Vanwanseele, and Jesse Davis. 2018. Fatigue Prediction in Outdoor Runners Via Machine Learning and Sensor Fusion. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18). Association for Computing Machinery, New York, NY, USA, 606-615. https://doi.org/10.1145/3219819.3219864
[37] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: with applications in R*. Springer.
[38] GeeksforGeeks. (n.d.). *What are the advantages and disadvantages of random forest?* GeeksforGeeks. https://www.geeksforgeeks.org/what-are-the-advantages-and-disadvantages-of-random-forest/ (accessed 14.09.2024.)
[39] Sedighi Maman Z, Alamdar Yazdi MA, Cavuoto LA, Megahed FM. A data-driven approach to modeling physical fatigue in the workplace using wearable sensors. Appl Ergon. 2017 Nov;65:515-529. doi: 10.1016/j.apergo.2017.02.001. Epub 2017 Mar 1. PMID: 28259238.
[40] Ashrant Aryal, Ali Ghahramani, Burcin Becerik-Gerber, Monitoring fatigue in construction workers using physiological measurements, Automation in Construction, Volume 82,2017, Pages 154-165, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2017.03.003.
[41] Zahra Sedighi Maman, Ying-Ju Chen, Amir Baghdadi, Seamus Lombardo, Lora A. Cavuoto, Fadel M. Megahed, A data analytic framework for physical fatigue management using wearable sensors, Expert Systems with Applications, Volume 155,2020,113405, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2020.113405.
[42] Strohrmann C., Harms H., Kappeler-Setz C., Troster G. Monitoring kinematic changes with fatigue in running using body-worn sensors. IEEE Trans. Inf. Technol. Biomed. 2012;16:983–990. doi: 10.1109/TITB.2012.2201950
[43] Hoda M. Abd-Elfattah, Faten H. Abdelazeim, Shorouk Elshennawy,Physical and cognitive consequences of fatigue: A review,Journal of Advanced Research,Volume 6, Issue 3,2015,Pages 351-358,ISSN 2090-1232,https://doi.org/10.1016/j.jare.2015.01.011.
[44] Garmin. (n.d.). *What is the Performance Condition feature on my Garmin fitness device?* Garmin Support. https://support.garmin.com/en-US/?faq=A28UA4k16v1qjjGuvSFgo8 (accessed 14.09.2024.)
[45] Apple Inc. (2024, June 5). *watchOS 11 brings powerful health and fitness insights*. Apple Newsroom. https://www.apple.com/newsroom/2024/06/watchos-11-brings-powerful-health-and-fitness-insights/(accessed 14.09.2024.)
[46] Karahanoğlu, A., Coskun, A., Postma, D., Scheltinga, B. L., Gouveia, R., Reidsma, D., & Reenalda, J. (2023). Is it just a score ? Understanding Training Load Management Practices Beyond Sports Tracking. In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 1116, 2024, Honolulu, HI, USA (Vol. 1, Issue 1). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642051
[47] Wan JJ, Qin Z, Wang PY, Sun Y, Liu X. Muscle fatigue: general understanding and treatment. Exp Mol Med. 2017 Oct 6;49(10):e384. doi: 10.1038/emm.2017.194. PMID: 28983090; PMCID: PMC5668469.
[48] Uchida, T.K. and Delp, S.L., 2015. Biomechanics of Movement: The Science of Sports, Robotics, and Rehabilitation.
[49] Meeusen R, Watson P, Hasegawa H, Roelands B, Piacentini MF. Central fatigue: the serotonin hypothesis and beyond. Sports Med. 2006;36(10):881-909. doi: 10.2165/00007256-200636100-00006. PMID: 17004850.

[50] Pope HG Jr, Wood RI, Rogol A, Nyberg F, Bowers L, Bhasin S. Adverse health consequences of performance-enhancing drugs: an Endocrine Society scientific statement. Endocr Rev. 2014 Jun;35(3):341-75. doi: 10.1210/er.2013-1058. Epub 2013 Dec 17. PMID: 24423981; PMCID: PMC4026349.
[51] Mayo Clinic. (2022). Performance-enhancing drugs: Know the risks. Mayo Foundation for Medical Education and Research. https://www.mayoclinic.org/healthy-lifestyle/fitness/in-depth/performance-enhancing-drugs/art-20046134 (accessed 10.09.2024.)
[52] Kluger BM, Krupp LB, Enoka RM. Fatigue and fatigability in neurologic illnesses: proposal for a unified taxonomy. Neurology. 2013 Jan 22;80(4):409-16. doi: 10.1212/WNL.0b013e31827f07be. PMID: 23339207; PMCID: PMC3589241.
[53] Mosso, A., Drummond, M., & Drummond, W.B., 1904. Fatigue. New York: G. P. Putnam's Sons; London: Swan Sonnenschein & Co.
[54] Enoka RM, Duchateau J. Translating Fatigue to Human Performance. Med Sci Sports Exerc. 2016 Nov;48(11):2228-2238. doi: 10.1249/MSS.0000000000000929. PMID: 27015386; PMCID: PMC5035715.
[55] Noakes TD. Fatigue is a Brain-Derived Emotion that Regulates the Exercise Behavior to Ensure the Protection of Whole Body Homeostasis. Front Physiol. 2012 Apr 11;3:82. doi: 10.3389/fphys.2012.00082. PMID: 22514538; PMCID: PMC3323922.
[56] Cleveland Clinic. (2023). *Heart rate monitor*. Cleveland Clinic. https://my.clevelandclinic.org/health/diagnostics/23429-heart-rate-monitor
[57] Seshadri, D.R., Li, R.T., Voos, J.E. *et al.* Wearable sensors for monitoring the internal and external workload of the athlete. *npj Digit. Med.* 2, 71 (2019). https://doi.org/10.1038/s41746-019-0149-2
[58] Joseph Mizrahi, Oleg Verbitsky, Eli Isakov, David Daily, Effect of fatigue on leg kinematics and impact acceleration in long distance running, Human Movement Science, Volume 19, Issue 2,2000, Pages 139-151, ISSN 0167-9457,https://doi.org/10.1016/S0167-9457(00)00013-0.
[59] Derrick TR, Dereu D, McLean SP. Impacts and kinematic adjustments during an exhaustive run. Med Sci Sports Exerc. 2002 Jun;34(6):998-1002. doi: 10.1097/00005768-200206000-00015. PMID: 12048328.
[60] Sheerin KR, Reid D, Besier TF. The measurement of tibial acceleration in runners-A review of the factors that can affect tibial acceleration during running and evidence-based guidelines for its use. Gait Posture. 2019 Jan;67:12-24. doi: 10.1016/j.gaitpost.2018.09.017. Epub 2018 Sep 14. PMID: 30248663.
[61] Inertial Labs. (2023). *Advantages and disadvantages of inertial measurement units*. Inertial Labs. https://inertiallabs.com/advantages-and-disadvantages-of-inertial-measurement-units/

9. Appendix

The appendix provides supplementary information that supports the findings and discussions presented in the main body of this research paper. It includes explanations of concepts such as fatigue and its physiological models, as well as additional data, figures, and forms used throughout the study. Sections 9.1 through 9.3 offer a deeper dive into the mechanisms and models of fatigue. Additionally, Appendices A, B, and C contain used participant measurement forms, and supporting figures related to the study's objectives.

9.1. Understanding fatigue

Fatigue is a complex phenomenon and a common non-specific symptom that is often characterized as an overwhelming feeling of exhaustion and translates to difficulty in performing voluntary tasks. If not resolved, over time can lead to overwork, chronic fatigue syndrome, endocrine and immunity dysfunction, a threat to overall human health, and in sports – overtraining syndrome and injuries [47]. This chapter describes some of the important factors that contribute to fatigue development.

9.2. Fatigue from a physiological perspective

According to its duration, fatigue can be divided into acute (can be quickly resolved by rest or lifestyle change) and chronic (lasting several months and is not resolved by rest). Furthermore, there is mental fatigue, which is a cognitive fatigue aspect, and physical fatigue which refers to the performance of the human motor system [47].

Muscle fatigue can be simply characterized as a decrease in maximal voluntary contraction. Based on its origin in the motor pathway, it is usually classified as central and peripheral. As the name suggests, the first one is produced in the central nervous system (CNS) and it decreases its ability to recruit motor units, while the latter is due to the biochemical changes in the working muscle, at or distal to the neuromuscular junction. The muscle contractile mechanism is responsible for muscle production, but any failure in nervous, ion, vascular, or energy systems leads to its decrease [47].

When voluntarily contracted, muscles usually fire at 50-60 Hz, and a decrease in these firing rates translates to the loss of force thus marking fatigue. From the neural perspective, impaired calcium ion (Ca^{2+}) release has been identified as a contributor to fatigue in isolated skeletal muscle fibers, since it plays a key role in cross-bridge cycling [47][48]. Moreover, the so-called central fatigue hypothesis states that changes in the concentration of certain neurotransmitters such as serotonin (5-HT), dopamine (DA), and norepinephrine (NA) are responsible for fatigue development during prolonged exercise. For instance, an increase in the serotonin-dopamine ratio is linked to feelings of lethargy and tiredness, marking the onset of fatigue, whereas a low ratio promotes motivation and arousal [49].

Another factor is blood flow which is essential for the removal of by-products of metabolic processes and brings oxygen to the working muscle that is necessary for aerobic adenosine triphosphate (ATP) production. As the muscle contracts, it increases the mean arterial blood pressure and decreases the net blood flow to the muscle, therefore leading to fatigue. Moreover, enriched oxygen uptake and ATP production are increased until the VO2max (maximal oxygen consumption) is reached. This means that if the athlete is exercising at a very high intensity, the VO2max is already reached, and there is a demand for more ATP that cannot be met by oxygen delivery, which eventually leads to fatigue [47].

To contract, muscles need energy – a ready supply of ATP, which is used for Ca^{2+} release, reuptake, and Na^+/K^+-pump function. To produce ATP, human bodies use glycogen, a carbohydrate fuel for muscle force production. Once its stores are depleted the exercise cannot continue. Lastly, during muscle contraction, the accumulation of metabolic factors such as hydrogen (H^+) ions, lactate, inorganic phosphate (Pi), reactive oxygen species (ROS), heat shock protein (HSP), and orosomucoid (ORM) plays a role in muscle fatigue production since they contribute to changes in cross-bridge cycling [47].

Even though it is easy to know when we are fatigued, it is hard to understand exactly which physiological processes led to such a condition. Consequently, there are no official recommendations for muscle fatigue treatment. Nonetheless, some nonspecific treatments are used, mainly in sports and the military, to enhance physical performance by manipulating an individual's physiological processes. They can be either synthetic (e.g. amphetamine, caffeine) or natural products (e.g. ginseng, garlic), but also nutritional supplements (e.g. vitamins, creatine, protein powder) [47]. However, it should be noted that the use of certain stimulants raises alarming ethical and safety concerns [50][51].

9.3. Fatigue models

Much research fails to define fatigue due to it being a complex, multi-factorial problem, but also due to the common assumption that the term *fatigue* is known to all, and the use of a wide range of definitions [52]. The challenge is to create a model that encompasses the most relevant factors that describe fatigue. This chapter explains some of the proposed fatigue models in the literature.

Italian physiologist Angelo Mosso is known as one of the pioneers in fatigue research for publishing his book “La Fatica” in the late nineteenth century. He developed the ergograph, the instrument he used to measure muscle fatigue by recording the muscle contractions during repetitive tasks. His statement “*fatigue of the brain reduces the strength of the muscles*” distinguishes central (e.g., the will, mental) and peripheral (muscular) fatigue from each other [53]. Since then, his two-domain concept of fatigue has served as an inspiration for further research, and his original fatigue scheme has been broadened.

Kluger et al. define fatigue as a neurological illness and divide it into two components. They argue the importance of the distinction between the terms *fatigue*, which refers to subjective sensations, and *fatigability* which refers to objective changes in performance. The first is influenced by homeostatic factors (e.g. depletion of glycogen and phosphocreatine, accumulation of lactate) and psychological factors (e.g. perceptions of effort, expectations, motivation, mood), whereas the latter is affected by peripheral (e.g. physiologic changes in muscle, the neuromuscular junction, and peripheral nerves) and central (e.g. disruptions in the CNS mechanisms) factors. According to this theory, the perception of fatigue and performance fatigability are not only distinct but also potentially independent. Moreover, they influence each other and are influenced by central and peripheral dysfunction or illness [52].

Similarly to Kluger et al., Enoka and Duchateau argue that fatigue is a disabling symptom consisting of two components: perceived fatigability and performance fatigability, and as such can only be measured by self-report (Figure 11). *Perceived fatigability* is affected by homeostasis (e.g. neurotransmitters, temperature, metabolites) and psychological state (e.g. mood, motivation, performance feedback), while *performance fatigability* is defined by contractile function and muscle activation [54].

T.D. Noakes argues that fatigue is no more than a brain-derived emotion that protects the body's homeostasis by regulating exercise behavior. He explains *The Central Governor Model of Exercise Regulation* scheme whose center is the brain which manages exercise performance by continuously adjusting the number of motor units activated in the working muscles. This regulation is influenced by both conscious and subconscious factors present before and during exercise, e.g. emotional state, sleep deprivation, level of motivation and experience, etc. The purpose of this control is to ensure that individuals always stop exercising before there is a risk of severe homeostatic disruption. As a result, every exercise is submaximal due to unused motor units in the working muscle. According to this model, the best performances are achieved by athletes who effectively manage these deceptive fatigue signals during exercise. He hypothesizes that in the case of a close finish, physiology does not decide who wins the race, but the athlete’s brain “decides” to win [55].

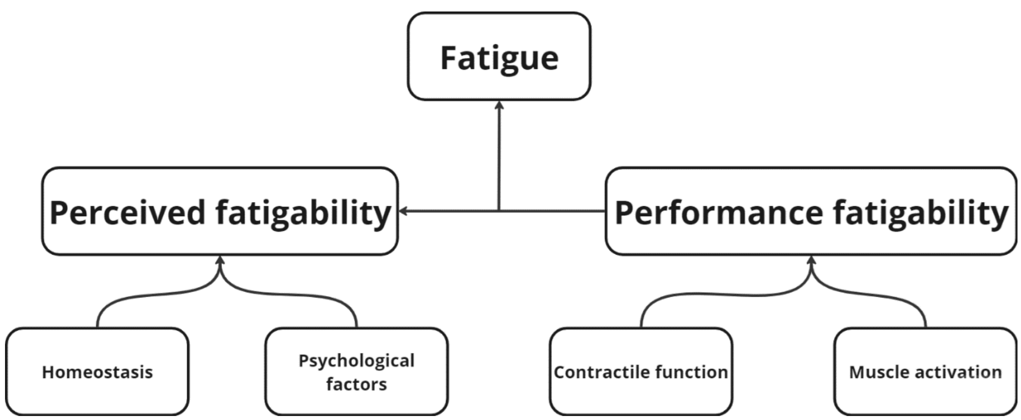

Figure 11 Fatigue model, adapted from Enoka and Duchateau [54].

9.4. Fatigue assessment methods

Difficulties in defining fatigue directly translate to challenges in its measuring and monitoring. In practice, fatigue measurements are task-specific, and one should carefully consider how to define fatigue for a given population and type of activity [21]. That said, fatigue in sports is commonly measured by widely accessible tests of direct physiological means such as heart rate, blood lactate concentration, or psychological questionnaires [22][23].

Mood changes are associated with training load and fatigue; therefore, self-report methods can be utilized as a straightforward and inexpensive way to capture the cognitive and emotional aspects of fatigue. Even though there are many validated questionnaires in the literature, sports organizations often choose their own, customized self-reports, due to existing questionnaires in the literature lacking sports specificity [22].In general sports research, the most used and practical tool for fatigue identification is the Borg Rating of Perceived Exertion (RPE) scale [21]. It represents a subjective fatigue indication but also shows a strong correlation with heart rate (r = 0.74) and blood lactate (r = 0.83) concentration and is applicable in both controlled laboratory settings and real-world conditions [23][24]. One study indicated that RPE serves as a more sensitive measure of acute stress levels, such as fatigue, compared to objective measures like blood lactate concentration and heart rate. This is because RPE captures both the psychological and physiological aspects of fatigue [22]. Consequently, the Rating of Perceived Exertion has gained increasing popularity in studies related to running injuries and clinical settings, owing to its contribution to accurately assessing an individual's training load and subsequent - injury risk [25]. It should be noted that subjective questionnaires are not a robust indicator of the athlete's performance, and conclusions drawn from self-report measures must be considered with caution, Finally, there is a need for more objective measures of fatigue [21].

One such example is cardiopulmonary exercise testing (CPET) which represents the golden standard for fatigue determination [23].It is a valuable tool in sports for measuring fatigue as it provides comprehensive data on an athlete's aerobic capacity and efficiency, ventilatory thresholds, and overall endurance. However, this method comes with a high cost and a complex laboratory setting [29]. In addition, physiological markers of fatigue and recovery provide insight into an athlete's response to workload. Unfortunately, only a limited number of these markers have strong scientific support for their use, and there is not one conclusive marker of fatigue. Moreover, many of the proposed markers such as biochemical (e.g. lactate, urea, creatine kinase), immunological (antibodies, cytokines, glutamine) and endocrine (e.g. stress hormone level) require laboratory analyses and cannot provide immediate fatigue status. On the other hand, heart rate, heart rate recovery, and variability are widely used markers in sports science since their alternations coincide with hormonal changes seen during training-induced fatigue [21]. During running heart rate can be monitored accurately, non-invasively, and continuously in real-time using heart-rate sensors [56]. It should be noted that, while HR monitoring provides valuable insights into the cardiovascular response during running, it may not fully capture the musculoskeletal fatigue that contributes to overuse injuries. At low to medium aerobic intensities, a runner's biomechanical loading can gradually accumulate, leading to movement compensations even when HR remains relatively stable. This indicates a potential mismatch between musculoskeletal and cardiovascular fatigue levels. Therefore, solely relying on HR is not sufficient for effective fatigue monitoring [36].

Wearable systems offer highly promising solutions for fatigue monitoring by allowing continuous, long-term tracking of biomedical signals in sports environments, ensuring the necessary comfort and non-intrusiveness [19]. It has been shown in research that motion analysis can serve as a valuable tool for injury prevention [18][19]. Motion capture systems, such as optical or inertial sensors can provide an accurate estimation of human motion, but they are expensive and not suited for outdoor applications [27]. As an alternative, Inertial Measurement Units could be employed for motion tracking and provide a better understanding of how running form and athlete's kinematic variables change with the fatigue progression. For example, some research found that peak tibial accelerations increase during running due to fatigue [57][58] and it is believed that higher peak accelerations indicate a higher load on the body and increase the injury risk, though this needs further research [59]. Although IMUs may be affected by noise, this can be managed through proper calibration and filtering techniques [60][61].

9.5. Appendix A

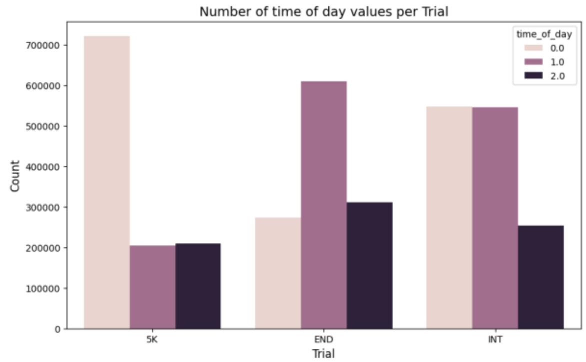

Figure 12 Distribution of time of the day samples per trial when the recording was done.

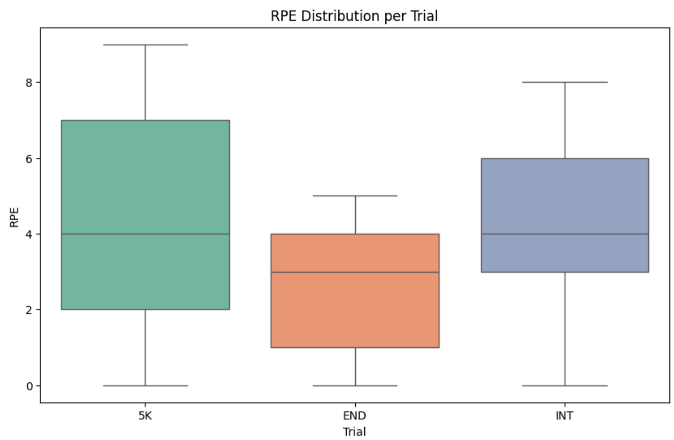

Figure 13 RPE scores distribution per trial.

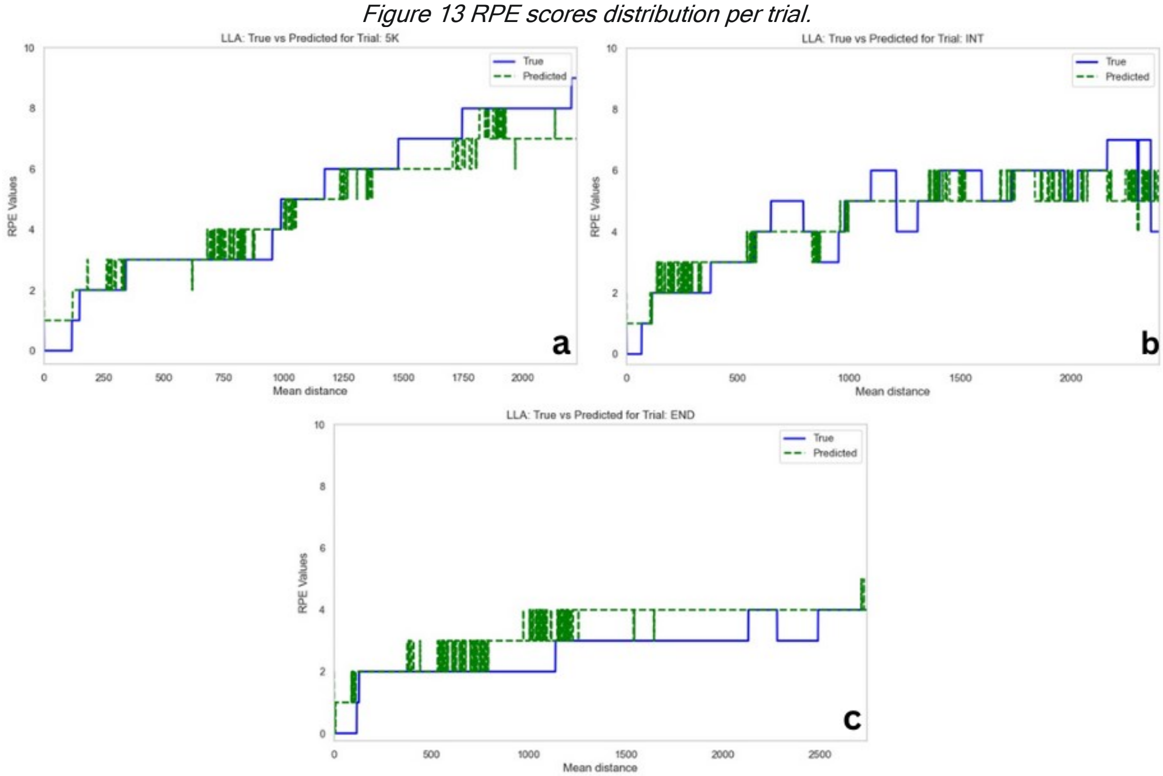

Figure 14 Prediction and true values averaged across all 5 subjects, and 3 trials, using the best LLA sensor configuration model.

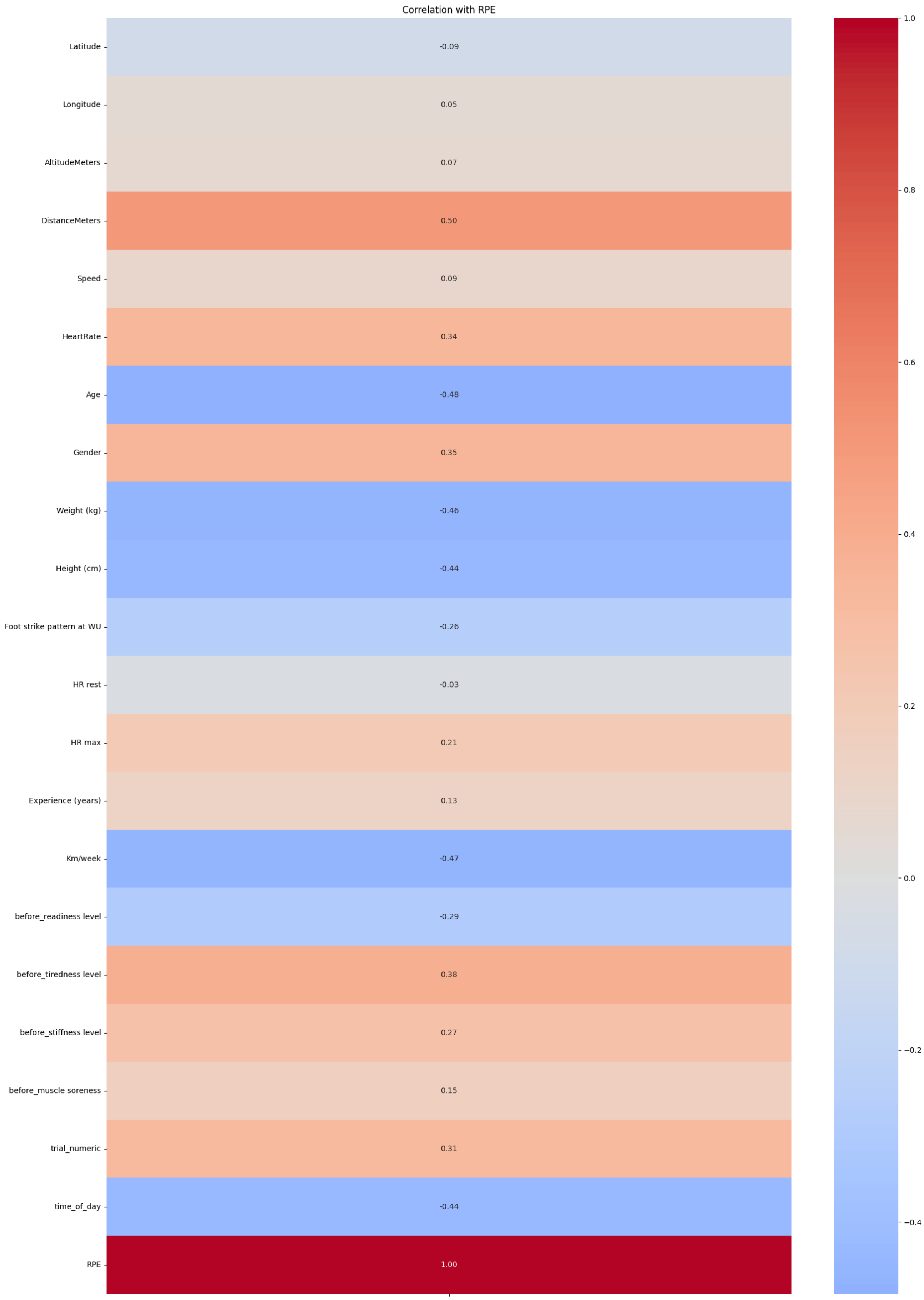

Figure 15 Correlation with subject characteristics with RPE.

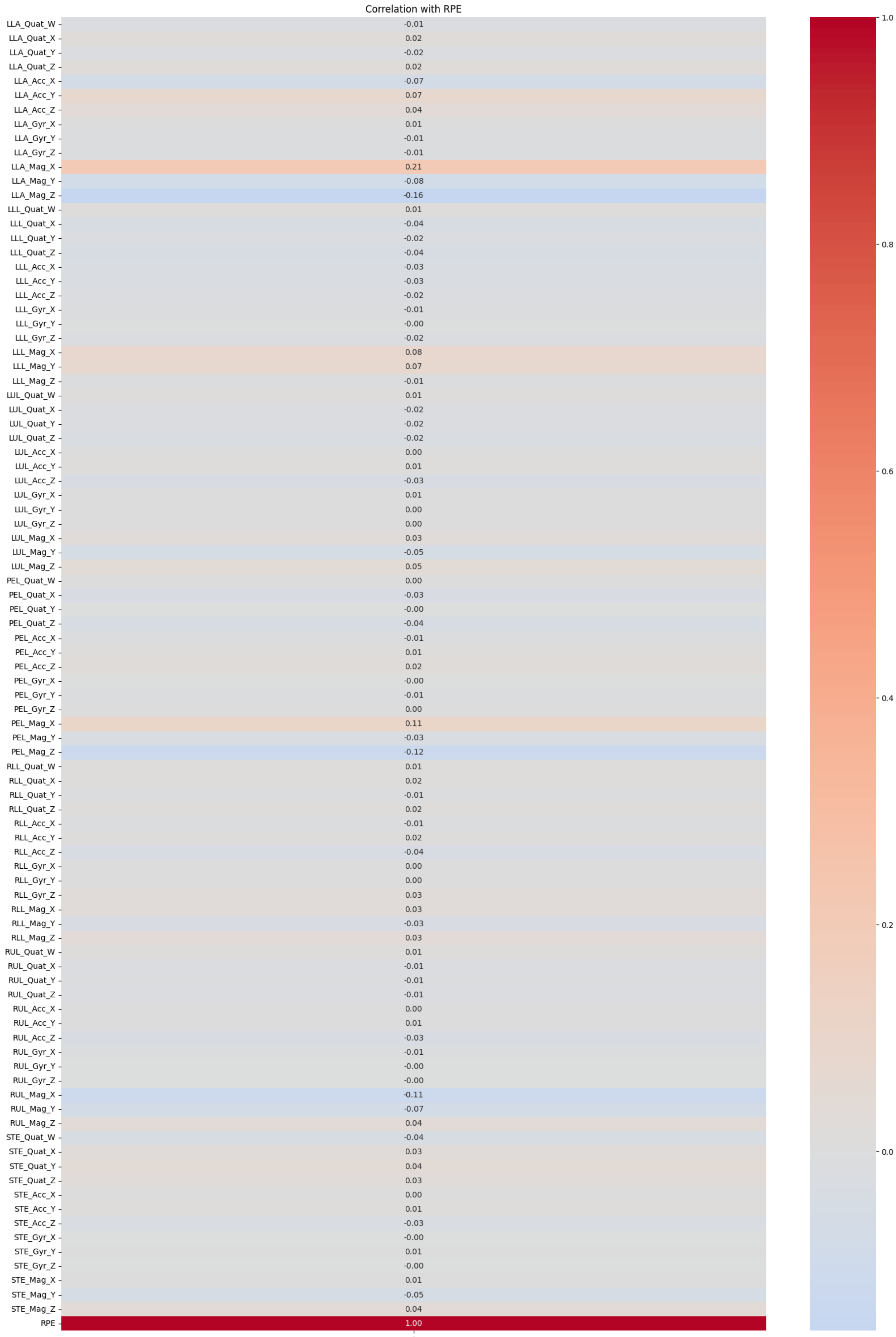

Figure 16 Correlation of IMU features with RPE.

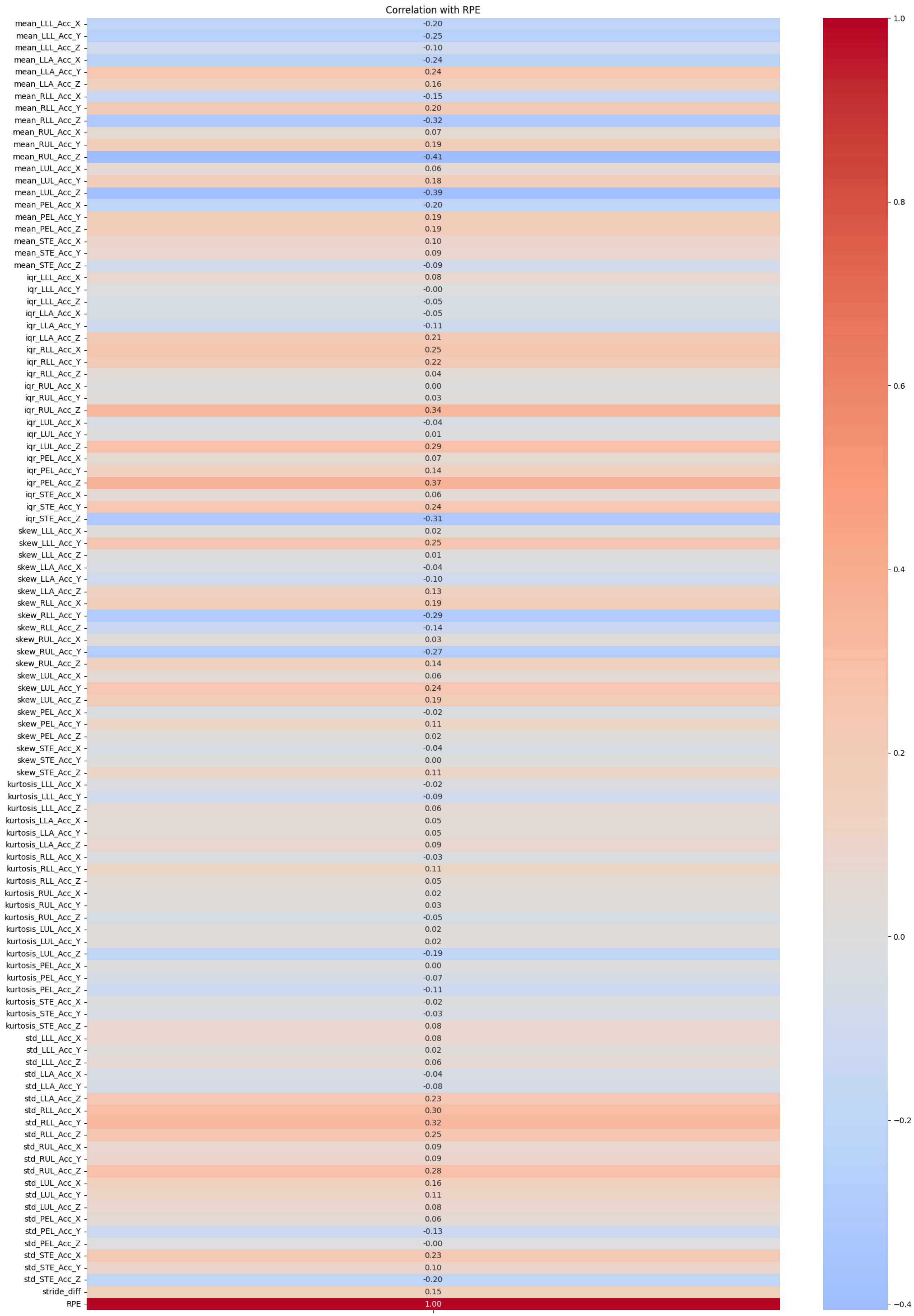

Figure 16 Correlation of statistical IMU features with RPE.

9.6. Appendix B

Measurement form

Personal information

Participant code:	
Date:	

Parameter	Value
Age	
Gender	M / V

Inclusion criteria

Criterium	Answer
Did you run the last 3 months at least 20km/week?	Yes / No
Did you run the last 3 months 2 times a week or more on average?	Yes / No
Did you had a major running injury in the lower extremities the past 6 months?	Yes / No
Are you pregnant?	Yes / No

General information

Question	Answer
Max heart rate (if known)	
Resting heart rate (if known)	

Question	Answer
How often do you run per week?	
How many kilometers do you run per week?	
How many years of running experience do you have?	
Dominant leg (kick a ball)	Left / Right
Shoe brand / type	
Did you had any running injuries, causing 4 or more weeks of no running? If yes, which?	
Do you do other sports at this moment?	
Representative 5 or 10km race result	
Endurance and warming-up pace	_____________ min/km
Target time and pace 5km sub-maximal	____________ _________ min/km

Question	Answer
Body height (cm)	
Body weight (kg), *with clothing and sensors*	

Measurement information

Participant code:	Date:
Session: END – INT – 5K	Location:

Sensor placement

Location	Sensor ID
LLL	
RLL	
LUL	
RUL	
PEL	
STE	
LLA / RLA	
Watch and heart rate monitor	

Additional remarks

Measurement information

Participant code:	Date:

Session: END – INT – 5K	Location:

Sensor placement

Location	Sensor ID
LLL	
RLL	
LUL	
RUL	
PEL	
STE	
LLA / RLA	
Watch and heart rate monitor	

Additional remarks

Measurement information

Participant code:	Date:
Session: END – INT – 5K	Location:

Sensor placement

Location	Sensor ID
LLL	
RLL	
LUL	
RUL	
PEL	
STE	
LLA / RLA	
Watch and heart rate monitor	

Additional remarks

9.7. Appendix C

Voor training vragenlijst

In deze vragenlijst worden vragen gesteld over hoe klaar u bent voor de komende training. Voorafgaande aan iedere training met sensoren zal deze worden ingevuld.

Wat is uw deelnemerscode die u van de onderzoek heeft gekregen? Dit is een S met twee getallen erachter, bv S99 *

Welke training staat er voor vandaag op de planning? *

- [ ] Duurloop
- [ ] Interval
- [ ] 5km (sub)maximaal

Vul in wat van toepassing is *

	Helemaal mee oneens	Oneens	Nuetraal	Eens	Helemaal mee eens	Weet ik niet
Ik ben klaar voor een zware training	○	○	○	○	○	○
Ik voel me vermoeid	○	○	○	○	○	○
Ik voel me stijf	○	○	○	○	○	○
Ik heb spierpijn	○	○	○	○	○	○

Heeft u nog verdere opmerkingen over uw gereedheid voor de training?

Voor training vragenlijst

In deze vragenlijst worden vragen gesteld over hoe klaar u bent voor de komende training. Voorafgaande aan iedere training met sensoren zal deze worden ingevuld.

Wat is uw deelnemerscode die u van de onderzoek heeft gekregen? Dit is een S met twee getallen erachter, bv S99 *

Welke training staat er voor vandaag op de planning? *

- [ ] Duurloop
- [ ] Interval
- [ ] 5km (sub)maximaal

Vul in wat van toepassing is *

	Helemaal mee oneens	Oneens	Nuetraal	Eens	Helemaal mee eens	Weet ik niet
Ik ben klaar voor een zware training	○	○	○	○	○	○
Ik voel me vermoeid	○	○	○	○	○	○
Ik voel me stijf	○	○	○	○	○	○
Ik heb spierpijn	○	○	○	○	○	○

Heeft u nog verdere opmerkingen over uw gereedheid voor de training?

Dag na training vragenlijst

In deze vragenlijst worden vragen gesteld over hoe u zich voelt de na de training met sensoren. De dag na iedere training met sensoren zal deze worden ingevuld.

Wat is uw deelnemerscode die u van de onderzoek heeft gekregen? Dit is een S met twee getallen erachter, bv S99 *

Welke training heeft u gisteren uitgevoerd? *

- ☐ Duurloop
- ☐ Interval
- ☐ 5km (sub)maximaal

Vul in wat van toepassing is *

	Helemaal mee oneens	Oneens	Neutraal	Eens	Helemaal mee eens	Weet ik niet
Ik ervaar meer stijfheid/spierpijn dan normaal in **spieren** vandaag	○	○	○	○	○	○
Ik ervaar meer stijfheid/pijn dan normaal in **pezen** vandaag	○	○	○	○	○	○
Ik ervaar meer stijfheid/pijn dan normaal in **gewrichten** vandaag	○	○	○	○	○	○
Ik ervaar meer vermoeidheid dan normaal vandaag	○	○	○	○	○	○

Heeft u nog verdere opmerkingen over hoe u zich nu voelt?