

## Article

# A Data-Driven Approach to Physical Fatigue Management Using Wearable Sensors to Classify Four Diagnostic Fatigue States

Maria J. Pinto-Bernal <sup>1</sup>, Carlos A. Cifuentes <sup>1,\*</sup>, Oscar Perdomo <sup>2</sup>, Monica Rincón-Roncancio <sup>3</sup>  
and Marcela Múnera <sup>1</sup>

<sup>1</sup> Department of Biomedical Engineering, Colombian School of Engineering Julio Garavito, Bogotá 111166, Colombia; maria.pinto@mail.escuelaing.edu.co (M.J.P.-B.); marcela.munera@escuelaing.edu.co (M.M.)

<sup>2</sup> School of Medicine and Health Sciences, Universidad del Rosario, Bogotá 111711, Colombia; oscarj.perdomo@urosario.edu.co

<sup>3</sup> Fundación Cardioinfantil-Instituto de Cardiología, Bogotá 110131, Colombia; mrinron@hotmail.com

\* Correspondence: carlos.cifuentes@escuelaing.edu.co

**Abstract:** Physical exercise contributes to the success of rehabilitation programs and rehabilitation processes assisted through social robots. However, the amount and intensity of exercise needed to obtain positive results are unknown. Several considerations must be kept in mind for its implementation in rehabilitation, as monitoring of patients' intensity, which is essential to avoid extreme fatigue conditions, may cause physical and physiological complications. The use of machine learning models has been implemented in fatigue management, but is limited in practice due to the lack of understanding of how an individual's performance deteriorates with fatigue; this can vary based on physical exercise, environment, and the individual's characteristics. As a first step, this paper lays the foundation for a data analytic approach to managing fatigue in walking tasks. The proposed framework establishes the criteria for a feature and machine learning algorithm selection for fatigue management, classifying four fatigue diagnoses states. Based on the proposed framework and the classifier implemented, the random forest model presented the best performance with an average accuracy of  $\geq 98\%$  and F-score of  $\geq 93\%$ . This model was comprised of  $\leq 16$  features. In addition, the prediction performance was analyzed by limiting the sensors used from four IMUs to two or even one IMU with an overall performance of  $\geq 88\%$ .

**Keywords:** fatigue diagnosis; classification models; inertial measurement units; EMG; walking rehabilitation; physical exercise



**Citation:** Pinto-Bernal, M.J.; Cifuentes, C.A.; Perdomo, O.; Rincón-Roncancio, M.; Múnera, M. A Data-Driven Approach to Physical Fatigue Management Using Wearable Sensors to Classify Four Diagnostic Fatigue States. *Sensors* **2021**, *21*, 6401. <https://doi.org/10.3390/s21196401>

Academic Editors: Angelo Maria Sabatini and Michael Hahn

Received: 12 August 2021

Accepted: 22 September 2021

Published: 25 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Exercise rehabilitation during or after medical treatment is considered an effective means of restoring physical and psychological function [1–3]. Recent reviews highlight the following benefits provided by physical exercise used as a therapeutic measure. Physical exercise contributes to achieving and maintaining therapeutic goals and improving quality of life, physical function, functional capacities, muscle strength, emotional well-being, and can even reduce depression and anxiety and increase self-esteem [4]. Additionally, it can lower the risk of heart disease, diabetes, cancer, stroke, reduce the risk of orthopedic problems, recover mobility of limbs, strengthen the immune defense (influenza), among others [4–7]. In this context, physical exercise has the potential to affect people's health conditions in many ways; hence, the American Health Association established physical exercise as one of the main components of improving people's health and decreasing morbidity and mortality levels [8,9].

Although these positive health-related outcomes of regular physical exercise are well documented, the amount and intensity of exercise needed to obtain these positive results

are unknown [10–12]. Several considerations must be kept in mind for its implementation in rehabilitation. For instance, elderly patients have more severe impairments and comorbidities than younger patients; therefore, the rehabilitation needs of the older are different from those of younger patients [1]. The same occurs with individuals with different diseases. Consequently, an essential question when prescribing exercise is what is the optimal therapeutic dose required to produce a specific health benefit according to the individuals' needs? Typically, when considering exercise dose about health outcomes, exercise is characterized by type, intensity, and volume (session duration and frequency) according to the age, weight, fitness level and pathologies of each patient [13]. Recent studies have shown that intensity is the most relevant feature in prescribing physical exercise [14,15] because it determines the amount of energy expenditure and can be seen as the “dose” of the prescription [16]. Since controlling exercise intensity avoids overtraining patients, which can affect their rehabilitation and make them suffer health consequences (i.e., physical, or physiological complications) [17–19]. In this context, it is essential to clarify the meaning of the various terms associated with physical activity and exercise for consistent interpretation of exercise intensity in the context of dose–response issues.

Intensity is the magnitude of the increase in energy expenditure necessary to perform the activity (aerobic or endurance exercise) or the force produced by the muscle contractions (resistance or strength exercise) [20,21]. Thus, the intensity has been used to classify the physical exercise into three groups: low-intensity exercises that are composed of soft activities which demand low energy cost (50% of the maximal HR ( $HR_{max}$ )), and are usually used for patients with extreme risk conditions; moderate-intensity exercises that contemplate non-stop activities with a long duration that require a low or moderate effort around 50% to 75% of the patient's  $HR_{max}$ ; and high-intensity exercises that are workouts that alternate hard-charging intervals with a short duration (15 s to 5 min) that increase the  $HR_{max}$  up-to 85% to 100% with a recovery period of equal or longer duration than the work interval [16,22]. Regarding prescribing exercise, moderate-intensity training is the most implemented in the rehabilitation process because it involves large muscle groups in dynamic activities that result in substantial increases in HR and energy expenditure [23,24]. The main challenge for the implementation of these exercises is the difficulties in managing the intensity, which makes their prescription a complex task [19,22,23].

Some successful rehabilitation programs have been explored to supervise the exercise intensity during therapies through monitoring of the patient's fatigue state [1,21,24–26]. Fatigue has generally been defined as a subjective state of tiredness or exhaustion and the reduction of capacity for regular activity [27]. Additionally, it is defined as the inability of the muscles to maintain the required level of strength during exercises. It can also result in the deterioration of health in the long term, including work-related musculoskeletal disorders [28], chronic fatigue syndrome [29] and compromised immune function [30]. Therefore, fatigue is a common concern among clinicians and individuals who participate in physical activities based on training or rehabilitation [27]. An important first step in managing fatigue is the rapid and accurate detection of its occurrences. However, because of the wide range of factors that can produce fatigue, there is no scientifically accepted method to identify it.

## 2. Related Works

Diverse fatigue detection techniques have been studied and used in rehabilitation that can be divided into qualitative and quantitative approaches. Qualitative methods are centered around the use of subjective scales of fatigue perception [31–33]. Some questionnaires consist of asking patients about their perceived tiredness level according to a pre-established ordinal numeric scale [34,35]. One of the most implemented in physical rehabilitation is the ten points Borg Rating of Perceived Exertion scale (Borg CR10) where a lower number represents a state of absence of fatigue, and a higher number represents a state of extreme fatigue [36]. However, the understanding of how an individual's perfor-

mance changes throughout rehabilitation is limited because the qualitative methods do not always represent the actual intensity has led to a decrease in reliability [32,37,38].

Quantitative approaches such as physiological parameters and exercise performance have been proposed for continuous monitoring of patients with chronic diseases during physical sessions [29,39–42]. Regarding physiological parameter measurement, one of their applications is the indirect estimation of fatigue [30,43–45]. The parameters most related to fatigue are: oxygen uptake ( $\text{VO}_2$ ) that represents the oxygen consumption that the body takes up and utilizes the exercising muscle [46]; heart rate (HR), which is one of the most used physiological parameters to control fatigue [47,48] due to its measurement facility and the linear relationship with  $\text{VO}_2$  [49]; and blood lactate, which is one of the most often measured parameters during clinical exercise testing as well as during performance testing of athletes [50]. It is essential for clinicians to understand the pathological response as well as the normal response to exertion [51]. However, it requires a blood sample and a specialized instrument, which are not easy to get during physical therapies [52]. Although the physiological parameters are considered accurate in measurement technique terms, these parameters present difficulties to monitor in real-time due to their acquisition processes. In addition, they may present different behaviors depending on the exercise type performed (e.g., moderate or high-intensity exercise), which makes it difficult to relate them to the fatigue level.

Considering that regardless of the exercise type performed the exercise performance has a directly proportional relationship with fatigue [53], several methods for monitoring fatigue through exercise performance have been implemented using ambulatory sensors (e.g., electromyography and inertial sensors) or non-ambulatory sensors (e.g., motion analysis system) to identify when an event exists outside the typical pattern, which supports the rehabilitation process and performance monitoring of activities [54].

Electromyography (EMG) is considered the gold standard to detect muscular fatigue considering that it directly assesses the bio-electrical muscle function [55,56]. However, EMG processing is a complex task to execute in real-time, since it requires power and frequency analysis [57] to identify fatigue progression. Therefore it inhibits their daily usage for real-time fatigue detection. Inertial measurement units (IMUs) are reliable and cheap sensors that are used to capture a person's acceleration and motion data in real scenarios without the use of external sources or devices [58–60], allowing to assess the activity performance through estimations of the kinematic and spatiotemporal parameters [61], and motion analysis [62]. Although it is possible to identify the person's fatigue level using these sensors, considering that the kinematic study in fatigue is still an early topic, the use of other physiological parameters like blood lactate [63], EMG [64], or even perceived level of fatigue [65] are widely used to corroborate the results. A motion analysis system is widely used in fatigue estimation due to its high accuracy and robustness in the measurement of the kinematic parameters. The motion analysis system is based on the use of infrared cameras to estimate the position of reflective markers to segment an object or an individual and measure variables such as position and orientation [66]. However, motion capture systems often require special setups which make them better suited for controlled environments [67].

As this article has already pointed out, humans' performance changes as a function of a person's individual characteristics (e.g., age, gender, fitness level, injury history, etc.), time (which can be manifested through detrimental performance due to fatigue and improved performance due to learning effects) and degree of exercise difficulty. Therefore, to enhance fatigue estimation the use of artificial intelligence systems in optimizing and transforming human performance has been implemented as a further alternative to monitor and understand how an individual's performance deteriorates with fatigue accumulation [38,68]. These models appear as a complementary alternative to collected human performance data from the diverse detection techniques (i.e., qualitative or quantitative approaches) that allow classifying fatigue levels. Table 1 summarizes fatigue modeling research using artificial intelligence systems. These studies differ between the physical activity movement

performed, the sensors used, and the fatigue identification. For instance, some studies attempted to identify high/extreme fatigue that results in an inability to generate muscle forces and consequently, an individual's performance decrease and inability to perform the physical activity [69–72]. The other studies focused on detecting fatigue without reaching exhaustion, where individuals are still able to perform their physical activity at a diminished level [38,73–75]. Since exhaustion in physical activity is often localized, the associated literature [69–73,75,76] focused on one physical activity only (e.g., walking or squats) primarily utilized 3D motion capture systems, EMG, accelerometers or IMUs, which were used to model the individual's performance, whereas only two studies developed models to focus on a more complex task [38,74], which utilized IMUs and HR monitors. Five studies used a subjective fatigue scale as a reference of the individuals' fatigue perception in contrast to their implemented fatigue detection method [38,70,73,74].

**Table 1.** A summary of fatigue modeling research.

Research	Physical Activity	Fatigue Detection Technique	Method
Maman et al. (2020) [38]	Manufacturing task	IMUs, HR, Borg Scale	SVM, RF, LR, PLR
Maman et al. (2017) [74]	Manufacturing task	IMUs, HR, Borg Scale	PLR
Zhang et al. (2014) [69]	Walking	3D optical tracking, IMUs	SVM
Karg et al. (2014) [70]	Squats	3D optical tracking, subjective scale	HMM, LR
Lee et al. (2009) [75]	Walking	3D optical tracking	LDA, Statistical test
Karg et al. (2008) [71]	Walking	3D optical tracking	LDA, SVM, kNN, NB
Helbostad et al. (2007) [76]	Walking	Accelerometers	Statistical test
Kavanagh et al. (2006) [72]	Walking	EMG	Statistical test
Yoshino et al. (2004) [73]	Walking	subjective scale, EMG, accelerometers	LR

SVM = support vector machines, RF = random forest, LR = logistic regression, PLR = penalized logistic regression, HMM = hidden markov models, LDA = linear discriminant analysis, kNN = k-nearest neighbors, NB = naive bayes.

These models presented a significant potential for clinical scenarios because they provide an objective indicator of the user's fatigue condition. However, they considered only two fatigue states, i.e., fatigued, or non-fatigued state, which were generally obtained in two separate steps; in the first step, the participants have no fatigue while in the second step fatigue is induced to the participants. The above limits the accurate monitoring of the user's exhaustion during therapy, restricts the possibility to determine the adequate "dose" (i.e., intensity) of the individuals to produce a specific health benefit according to their individuals' needs, and thus limits improvements to the user's performance during therapy. On the other hand, from a detailed literature review, we could not identify any article that discussed or established the best classifier for the identification and diagnosis of fatigue. This may be attributed to the lack of understanding of how an individual's performance deteriorates with fatigue accumulation, which can vary based on user conditions and physical activity.

As a first step, this paper proposes a framework for developing a fatigue identification model based on the individuals' exercise performance assessment to classify four fatigue diagnosis stages (low, medium, high, and very high). The fatigue diagnosis stages allow clinicians to pinpoint the hazard directly. They can then prescribe interventions from a

large number of options, including assigning rest breaks (which can reduce the level of fatigue before it reaches potentially dangerous levels) or redesigning the activity (which can eliminate the development of fatigue) [38]. Four fatigue stages allow accurate monitoring of patients' fatigue conditions during exercise to avoid any injuries or affect the rehabilitation process.

To this end, our framework takes benefits from the advances and widespread use of wearable sensors for data collection. These sensors offer an individualized insight into the individuals' performance and present a unified performance benchmark that does not depend on process cycle time (it is essential for real-time scenarios). Furthermore, we have chosen to focus on monitoring fatigue in walking tasks since (i) localized muscle fatigue is a potential risk factor for injury or falls as muscle fatigue adversely affects proprioception, movement coordination and muscle reaction times, leading to postural instability and gait alterations [65,69,73,76,77]; therefore, gait patterns associated with fatigue may help in the assessment of fatigue-related fall risks or injuries in various environments; and (ii) moderate-intensity training as walking exercises are one of the most used in physical rehabilitation process due to improving cardiovascular system and skeletal muscle function [20,22,23,78].

In this context, this framework allows to understand how an individual's performance deteriorates with fatigue accumulation in walking tasks. The code and data implemented in this work are offered as supplementary materials to encourage adoption in practice and further investigations by researchers.

Finally, it is significant to emphasize that social interaction has been shown to have a positive impact on general mental and physical well-being in physical exercise therapy [79,80]. Therefore, the use of these artificial intelligence systems has a great potential in physical rehabilitation scenarios application through the use of socially assistive robots (SAR) [81–83]. SAR is a system that employs interaction strategies, including the use of speech, facial expressions, and communicative gestures, to assist according to a specific healthcare environment [80]. Several studies have shown that SAR systems improve (i) the ability to influence the patient's intrinsic motivation to perform the task and (ii) the ability to personalize the social interaction to maintain user engagement in the task through permanent patient monitoring and feedback [79,81–86]. Therefore, with the support of the machine learning models, this technology has been used to provide better service and improve patient's experience and treatment outcomes [81,83]. An example of the above is the study conducted by Casas et al. [81], who implemented a social robot to act as a training assistant in a cardiac rehabilitation program in a clinical setting. The results suggested that SARs are well received by patients and have a positive impact on their willingness to perform prescribed rehabilitation exercises. In summary, the social robot achieved continuous monitoring and continuous feedback to patients on their performance, leading to an increase in their motivation during the rehabilitation session. Similar outcomes are reported by [79,83]. In this regard and the context of fatigue detection, a future better application would be to implement these machine learning models in the RAS. Because this would provide continuous monitoring and feedback on the patient's performance, allowing to control the intensity and the correct execution of the exercise. Similarly, alerts could be generated when the patient is presenting high levels of fatigue or there is a variation in exercise performance parameters. This is very promising for clinical scenarios and would be of great help to avoid overtraining in patients, which could affect their rehabilitation and even suffer physical or physiological complications.

### 3. Materials and Methods

The study has been divided into two parts: an experimental setup to obtain the corresponding dataset and a proposed framework for a fatigue classifier to develop and assess the fatigue estimation models. Specifically, six classification strategies have been proposed and implemented to automatically identify four fatigue diagnosis levels in walking exercises by monitoring 43 kinematic/temporal and biomechanical features.

### 3.1. Experimental Setup

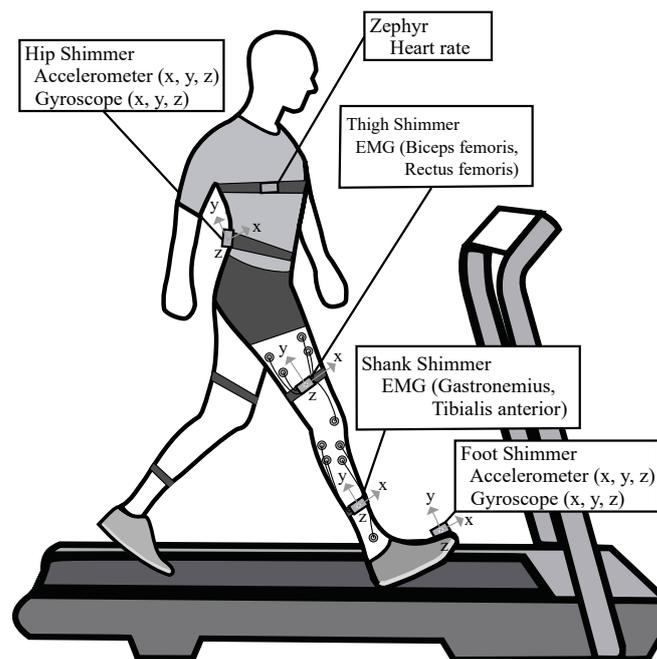
The experimental protocol was performed to quantify the detection success ratio of the fatigue classifiers. A total of 24 healthy subjects (14 males, 10 females,  $21.75 \pm 1.16$  years old,  $1.71 \pm 0.09$  m) performed the study, who perform regular physical exercise and who had no known physical or cognitive disability, injuries, pain, or any impediment to exercise (see Table 2 for further information). The participants enrolled in this study, must not present any fatigue state, i.e., they had to be in a non-fatigued state to avoid affecting their test performance. To this end, five different types of fatigue—general fatigue, physical fatigue, mental fatigue, reduced motivation, and reduced activity—were assessed using the “multidimensional fatigue inventory” [87] questionnaire. All subjects presented non-fatigued conditions, and were informed about the scope and purpose of the experiment. Written consent was obtained from each of them before the study. The Ethics Committee of the Colombian School of Engineering Julio Garavito, Bogota, Colombia approved the protocol.

**Table 2.** Summary of participants’ descriptive data (M  $\pm$  SD). BMI, body mass index.

Gender	Age [Years Old]	BMI [kg/m <sup>2</sup> ]	Walking Speed [m/s]
Male	$21.83 \pm 1.40$	$22.84 \pm 2.90$	$0.18 \pm 0.37$
Female	$21.64 \pm 0.74$	$22.25 \pm 3.09$	$0.18 \pm 0.35$

Volunteers were first instructed to perform three 10 m tests at a self-selected speed to determine their average overground speed, which was successively set on a treadmill (NIZA RX K153D-A-3, SportFitness, Bogota, Colombia). Participants were equipped with four Shimmer3 (Shimmer, Dublin, Ireland) IMU units, one located on their foot dominant instep and one located around the center of L5-S1 with a sample rate of 128 Hz. The remaining two units were configured with a sampling rate of 512 Hz to obtain EMG signals in four muscles (tibialis anterior, rectus femoris, biceps femoris, and gastrocnemius) of the participant’s dominant leg. One IMU was located on the outer lateral part of the thighs and one on the calves’ outer side with elastic bands, the EMG signal was recorded from a pair of Ag–AgCl electrodes (interelectrode distance 3 cm) after cleansing the skin by alcohol. In addition, participants were equipped with a Zephyr HxM BT (Medtronic, Ireland) on their chest with an elastic band, with a sample rate of 1 Hz. The selection of the Zephyr BT sensor is based on accuracy, reliability, cost, availability, and comfort [88,89]. The experimental setup described is illustrated in Figure 1.

Participants were then asked to walk for at least 120 s on the already configured treadmill at a zero-degree inclination, where different parameters were assessed. The first two parameters directly indicate the fatigue level: blood lactate and a perceptible fatigue scale using the Borg CR10. Therefore, they were used as reference values to diagnose and classify the participants’ fatigue level. At the same time, kinematic/spatiotemporal parameters, and EMG signals were recorded. Afterward, participants’ fatigue inducement was carried out. Volunteers had to perform as fast as possible a physical exercise circuit composed of four exercises: high knees, jumping jacks, squats, and short runs. At the end of the physical exercise circuit the participants returned to the treadmill and the whole process was repeated.



**Figure 1.** Experimental setup. Each participant was instrumented on their dominant side with an IMU placed on the dorsal side of their foot and with an IMU placed on their spine between L2 and S1. two more IMUS located on each thigh and shank with different electrodes situated in tibialis anterior, rectus femoris, biceps femoris, and gastrocnemius. Participant’s heart rate was captured using the Zephyr sensor located on their chest.

The above was repeated four times to increase the participants’ fatigue level. The difference between each round was the execution time of the physical circuit that increased progressively. In other words, the time corresponding to the performance of each exercise increased by 15 s each round; for the first time, each exercise was performed for 30 s, the second round for 45 s, the third round for 60 s, and the last round for 75 s. Note that if the participants’ HR overcame 90% of the  $HR_{max}$ , or a 10 Borg value was notified, the test was immediately concluded. The entire experiment, including donning/doffing times related to instrumentation procedures and walking tasks, was completed within 60 min for all volunteers.

It is worth highlighting that to get an approximation of each volunteer’s  $HR_{max}$ , Tanaka’s formula using the user’s age (in years) was implemented as is shown in Equation (1). Tanaka’s equation is recommended for healthy individuals such as those involved in this study because this equation significantly overpredicts maximal HR. Therefore, for people who present some diseases, it is recommendable to adapt this method when using exercise testing [90].

$$HR_{max} = 206.9 - (0.7 * age) \quad (1)$$

Regarding the measured parameters, they were measured every time that participant returned to the treadmill. The blood lactate sample was taken from the participant’s earlobe with a new lancet. The blood was collected with a new test strip, and finally, the strip was inserted into the Lactate Pro2 (Arkray, Japan) to measure the blood lactate level. On the other hand, the Borg CR10 scale was obtained by asking the volunteer how tired they felt according to the scale; Table 3 was used to explain the values meaning to each volunteer according to the four fatigue levels (low, moderate, high, very high). The participant was also instructed only to focus on the total effort sensation and not on shortness of breath or muscle pain. Concerning the data acquisition, it only started once

the self-selected speed was reached, and the treadmill speed was only reduced after all data were acquired to prevent data capture during the transient state.

**Table 3.** Borg scale description and classification

Borg CR10 Value	Description	Classification
0	No exertion at all	Low
1	Very easy	
2	Easy	
3	Somewhat moderate	Moderate
4	Moderate	
5	Somewhat hard	
6	Hard	High
7	Very Hard	
8	Very Very Hard	
9	Extremely hard	Very High
10	Maximum exertion	

### 3.2. Proposed Approach for Fatigue Classifier

Figure 2 presents an overview of the proposed framework for managing physical fatigue. This framework was divided into two phases. The first phase consisted of the selection of an appropriate classifier model for prospective analysis. To this end, the first phase is comprised of three main steps: (i) data preprocessing and feature generation, where the sensors' data are prepared for analysis and generated the dataset; (ii) model construction and validation, where statistical and data analytic models are trained for distinguishing between the four fatigue states (non-fatigued, low-fatigue, moderate-fatigue, high-fatigue); (iii) data analysis, where the classifier models are evaluated based on accuracy, precision, recall and F-score. In the second phase, where the best model is established, the subset of features/predictors that are most frequently used in predicting the fatigue state are identified.

### 3.3. Phase 1: Fatigue Detection

#### 3.3.1. Data Preprocessing

The first step in analyzing data is to ensure that the data is correct and cleaned. Therefore, the fourth main cleaning step was proposed. First, the gyroscope and accelerometer outputs were treated with a second-order low-pass filter Butterworth for noise removal [91,92]. Second, the collected data were visualized to check any additional erroneous data (e.g., faulty sensor values (too high and too low)), i.e., data that were not corrected through the automated filtering in the previous step. Third, the data from the different sensors were synchronized and any observations that were captured outside of the experimental window were eliminated. Fourth, the data were partitioned using a non-overlapping time window, where the selection of the length of the time window depended on the identification of each participant's gait event. Considering that gait presents a repetitive behavior, an automated procedure was implemented to detect each participant's gait cycle. Specifically, the process consisted of determining each heel-strike event by detecting the second minimum value in the angular velocity signals of each stride time. This process outcome is represented in Figure 3, where a sample of each heel strike of the corresponding signal from a participant's test record can be observed. In addition, a zoom of a signal part is presented in Figure 3, where gait event detection throughout the angular velocity of an inertial sensing system over two gait cycles can be appreciated [60]. The selection of the correct threshold value was carried out as in [93], whose study validated all possible thresholds of the gait cycle within a range, and whose limits were visually established from signals acquired in a preliminary analysis.

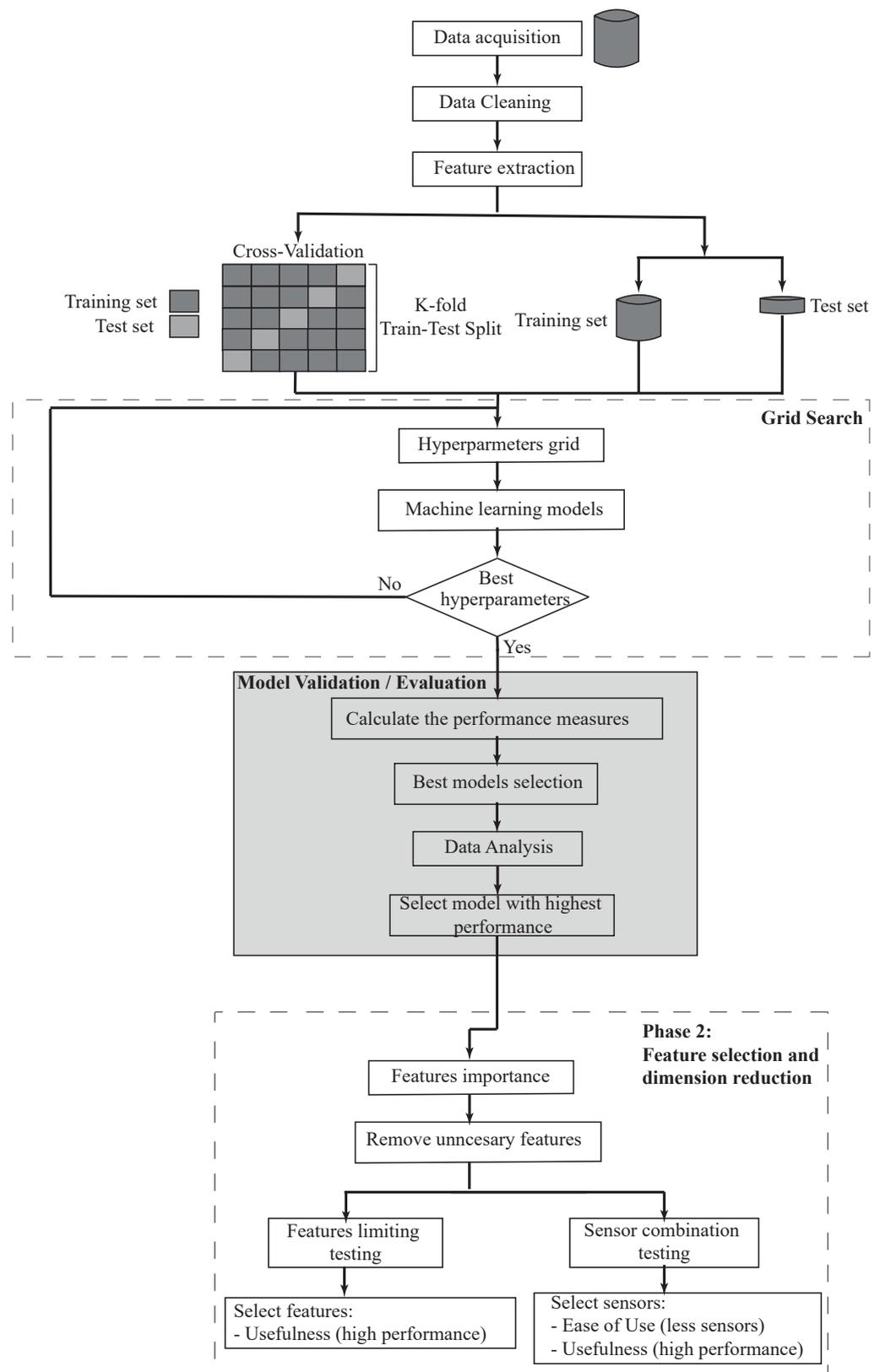
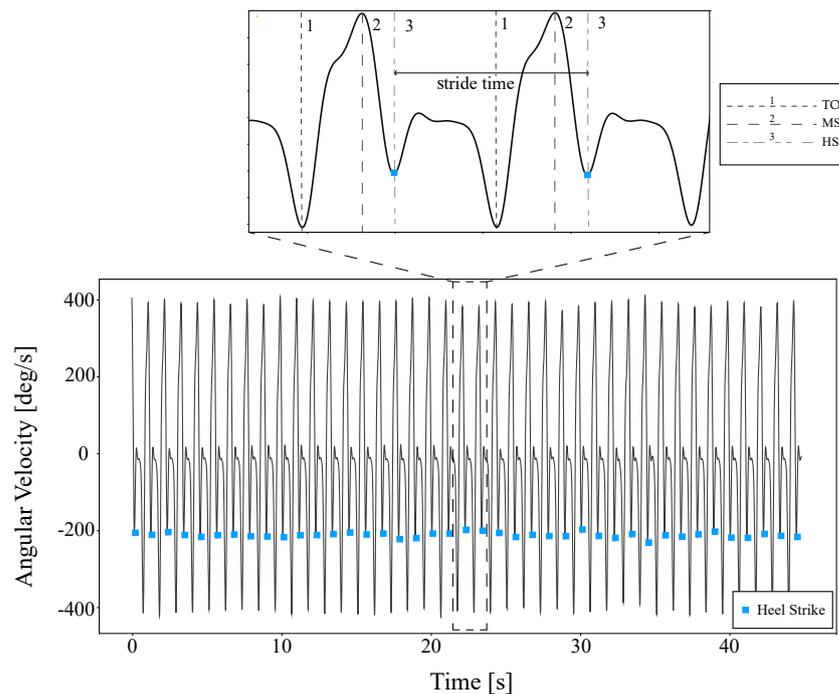


Figure 2. Flowchart that illustrates an overview of proposed method.



**Figure 3.** Heel strike detection using an inertial detection system on a participant’s test record. The zoom part represents two gait cycles and the identification of each gait phase: TO = Toe-Off (first dashed line), MS = Mid-Swing (second dashed line), and HS = Heel Strike (third dashed line).

Regarding feature generation, two IMUs were attached at the participants’ foot instep and around L5-S1 to measure the acceleration associated with a person’s dynamic motion (spatiotemporal and kinematic parameters). These features captured the intensity and spread, which are commonly used in the fatigue detection literature [74,94,95]. Likewise, two more Shimmer3s were used and configured to obtain EMG signals in four muscles (tibialis anterior, rectus femoris, biceps femoris, and gastrocnemius) to register electrical potentials [96]. These potentials are directly related to muscular strength, which allows to estimate the effort and evaluate the participants’ exercise performance [97]. The description of the proposed features is provided in Table 4. Note that these features are calculated for each time window, i.e., for each gait cycle.

**Table 4.** Features generated.

N <sup>o</sup>	Feature	Description	Ref.
0	gait_mean_acce	Average gait acceleration	
1	gait_std_acce	Average gait acceleration std	
2	gait_max_acce	Average gait maximum acceleration	
3	gait_var_acce	Average gait acceleration variance	
4	gait_median_acce	Average median gait acceleration	
5	gait_energy_acce	Average gait acceleration energy	
6	gait_entropy_acce	Average gait acceleration entropy	
7	gait_kurtosis_acce	Average gait acceleration kurtosis	
8	gait_maxfreq_acce	Average gait acceleration maxfreq	
9	gait_stdfreq_acce	Average gait gyro stdfreq	
10	gait_mean_gyro	Average gait angular velocity mean	
11	gait_std_gyro	Average gait angular velocity std	
12	gait_max_gyro	Average gait maximum angular velocity	
13	gait_var_gyro	Average gait angular velocity variance	
14	gait_median_gyro	Average median gait angular velocity	

Table 4. Cont.

N°	Feature	Description	Ref.
15	gait_energy_gyro	Average gait angular velocity energy	
16	gait_entropy_gyro	Average gait angular velocity entropy	
17	gait_kurtosis_gyro	Average gait angular velocity kurtosis	
18	gait_maxfreq_gyro	Average gait angular velocity maxfreq	
19	l2_mean_acce	Average ts acceleration	[9,95]
20	l2_std_acce	Average ts acceleration std	[98,99]
21	l2_max_acce	Average ts maximum acceleration	[94,100]
22	l2_var_acce	Average ts acceleration variance	[101,102]
23	l2_median_acce	Average ts acceleration velocity	[76,103]
24	l2_energy_acce	Average median ts acceleration energy	[70].
25	l2_entropy_acce	Average ts acceleration entropy	
26	l2_kurtosis_acce	Average ts acceleration kurtosis	
27	l2_maxfreq_acce	Average ts acceleration maxfreq	
28	l2_stdfreq_acce	Average ts acceleration maxfreq std	
29	l2_mean_gyro	Average ts angular velocity mean	
30	l2_std_gyro	Average ts angular velocity std	
31	l2_max_gyro	Average ts maximum angular velocity	
32	l2_var_gyro	Average ts angular velocity variance	
33	l2_median_gyro	Average median ts angular velocity	
34	l2_energy_gyro	Average ts angular velocity energy	
35	l2_entropy_gyro	Average ts angular velocity entropy	
36	l2_Kurtosis_gyro	Average ts angular velocity kurtosis	
37	l2_maxfreq_gyro	Average angular velocity maxfreq	
38	l2_stdfreq_gyro	Average ts angular velocity maxfreq std	
39	rms_gastro	RMS envelope of the gastrocnemius signal	
40	rms_tibilisAnterior	RMS envelope of the tibilis anterior signal	
41	rms_rectusFemoris	RMS envelope of the rectus femoris signal	
42	rms_bicepsFemoris	RMSenvelope of the biceps femoris signal	

Ref. = References; ts = torso swing; std = standard deviation; Maxfreq = maximum frequency; stdfreq = frequency standard deviation; RMS = root-mean-square.

It is essential to highlight that feature variability caused by each participant's physical condition makes it difficult to perform a direct comparison among the volunteer registers, which requires a normalization of the data according to each initial subject performance [104–106]. All features extracted were normalized by dividing them with the corresponding initial value (see Equation (2)). Note that the test number zero of each volunteer was taken as a reference since it was considered that the volunteer did not have fatigue, which was corroborated with the blood lactate, the Borg CR10 scale, and the multidimensional fatigue inventory questionnaire.

$$F_i = \left( \frac{F_i}{F_{reference}} \right) \quad (2)$$

where  $F_i$  corresponded to each feature extracted in each time window and  $F_{reference}$  is the average of the first ten time window values of the test number zero where volunteers were not fatigued.

### 3.3.2. Model Construction and Validation

In machine learning, there are several methods of data partitioning for experimentation. The most popular methods are typically referred to as training/test partitioning or cross-validation [107,108]; both were implemented in this research to evaluate the best classifier performance.

The training/test partitioning typically involves the partitioning of the data into a training set and a test set in a specific ratio, e.g., 70% of the data are used as the training set and 30% of the data are used as the test set. This data partitioning can be done randomly or fixed. The fixed way is typically avoided (except when order matters) as it may introduce

systematic differences between the training set and the test set, which leads to sampling representativeness issues. To avoid such systematic differences, the random assignment of instances into training and test sets are typically used [108,109].

Cross-validation is conducted by partitioning a data set into  $n$  folds (or subsets), followed by an iterative process of combining the folds into different training and test sets. In other words, each of the  $n$  folds is, in turn, used as the test set at one of the  $n$  iterations, while the rest of the folds are combined as the training set [110]. A typical approach to cross-validation is dividing the dataset into 10 folds, where the models are selected based on the average/median prediction performance across 10 non-overlapping test datasets. The literature suggests that 10-fold cross-validation may reduce the variation between the training and test performance [111]. However, cross-validation is generally more expensive in terms of computational cost than training/test partitioning.

Regarding the selection of the machine learning models, several classification methods are viable candidates for utilization in fatigue prediction. However, from our framework's perspective, it is impossible to predetermine which methods will work best for fatigue prediction in the walking task. Because these methods are data-driven and thus, are application-dependent, i.e., dependent on the exercise, extracted features, sensors, scenarios, among others, several methods were applied during our preliminary analysis of the data to develop the fatigue prediction model. The models evaluated included: logistic regression (LR), decision trees (DT), k-nearest neighbors (KNN), support vector machine (SVM), naive Bayes (NB), linear discriminant (LDA), artificial neuronal network (ANN), and random forest (RF). The open-source python library "scikit-learn" [112] was used to execute a quick general training for these classifiers. Afterward, according to the accuracy metric and due to their relatively poor performance, LDA and NB were eliminated. Hence, our case study focused on using the best six classification models (LR, KNN, SVM, RF, ANN, and DT), adjusted and retrained, by modifying their training hyperparameter automatically through computational iterations. For a detailed introduction on the classifier mentioned above, the reader is referred to [113,114].

### 3.3.3. Data Analysis

Developments in machine learning classifiers from imbalanced data have been mainly motivated by numerous real-life applications since it faces the problem of unequal representation of the data [115]. Most of the machine learning models used for classification have been designed around the assumption of an equal examples distribution for each class [116]. This means that an incorrect application of a model may focus on learning the characteristics of the abundant observations only, neglecting the examples from the minority class increasing false positives. However, a slight imbalance, i.e., the distribution of examples is uneven by a small amount in the training dataset (e.g., 4:6), is often not a concern [117,118]. In this context, to evaluate the performance of the proposed fatigue detection models, it was essential to consider the following performance measures.

Recall or true positive rate (TPR), which captures the ability to detect the fatigued cases, i.e., quantifies the number of positive class predictions made of all positive examples in the dataset, is computed as follows

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

where a true positive ( $TP$ ) is considered if the classifier prediction and the reference value match. Otherwise, such classification is regarded a false positive ( $FP$ ), i.e., it is the number of registers that belongs to other groups and were wrongly estimated. Likewise, accuracy presents the percentage of correct classifications made by a given model (Equation (4), where  $n$  represents the complete amount of data).

$$Accuracy = \frac{1}{n} \left( \frac{TP + TN}{TP + FN + FP + FN} \right) \quad (4)$$

where the non-fatigue state is similarly detected by a classifier and reference signal correspond to true negative (*TN*); otherwise, they have been accounted for by false negatives (*FN*). Precision, which quantifies the number of positive class predictions that actually belongs to the positive class, was measured as follows

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

The last metric considered was F-score (Equation (6)) which provides a single score that balances both the concerns of precision and recall in one number.

$$F1_{micro}score = \frac{Precision * Recall}{Precision + Recall} * 2 \quad (6)$$

### 3.4. Phase 2: Feature Selection and Dimension Reduction

Once the best prediction model is identified, it is essential to consider that a critical aspect for technology adoption is usability. When the number of potential features/predictors is large, the computational complexity for model training increases. Feature reduction is typically applied to reduce the computational burden. In general, models are more interpretable if the number of features is smaller, which could (i) equalize or even increase the performance metrics of the classifier by removing unnecessary features from the data, and (ii) increase the generalization capability [119]. In the context of our framework, usability can be measured using a total number of features selected; therefore, we hypothesize that the chosen prediction model will have a relatively low number of features. Considering that the proposed framework enables the diagnosis of fatigue and the recommendation of an appropriate intervention, the feature selection/reduction was performed through univariate statistical approaches where the feature selection was based on their relationship to the response and their prediction performance. From this step, practitioners can understand which features affect fatigue and how they are associated with changes in the potential classifier. Therefore, any unchanged features in the fatigued and non-fatigued states should be removed. This suggests that the result would be a more interpretable fatigue classifier with relatively large prediction performance, i.e., a good fatigue detection classifier with a low false alarm rate.

## 4. Results

Table 5 summarizes the samples numbers of the four fatigue states. The “Low” class presents most of the registers with 43%, followed by the “Moderate” with and “High” classes with similar proportions (26.41% and 25.08% ). The “Very High” class has the lowest value, with a difference of 10% between the “Moderate” and “High” classes, and a difference of 16.70% regarding the “Low” class. The difference between the classes represents an imbalanced dataset.

**Table 5.** Data distribution in the dataset according to each fatigue state.

Class	Number of Samples
Very High	463 (15.86%)
High	732 (25.08%)
Moderate	771 (26.41%)
Low	953 (32.65%)
Total	<b>2919 (100%)</b>

The predictive performance of the six models is summarized in Table 6. The table shows the mean for each of four metrics. The reported results are based on 2919 dataset samples from the respective partitioning method: training/test (i.e., 80% of the data were used as the training set and 20% of the data was used as a test set). For the first three numeric columns, a higher value is desired since it reflects a better prediction performance.

Note that since the data set obtained is imbalanced, which causes an increase of the false positives, the F-score was considered as the most relevant feature for the performance classifier selection. Moreover, Table 6 reports the parameters and the performance of the six classifiers implemented. The logistic regression (LR) classifier implements the large-scale bound-constrained optimization as a penalty algorithm (solver = newton-cg), and a value of 1,000,000 for its inverse of the regularization strength learning parameter ( $C = 1,000,000$ ). The k-nearest neighbor (KNN) method using Euclidean distance classified the registers by a majority vote of its nearest elements with three neighbors ( $K = 3$ ). The decision tree (DT) method using the function to ensure the quality of a split (criterion = entropy) and the tree depth to control the size of the tree to prevent overfitting (max depth = 12) can create arbitrarily small leaves (min samples split = 11) and guarantees that each leaf has a minimum size (min samples leaf = 4), avoiding low-variance over-fit leaf nodes in regression problems. The support vector machine (SVM) has a radial basis function kernel (kernel = balanced) and a constrain value of 64 ( $C = 64$ ). The artificial neuronal network has a stochastic gradient-based optimizer (solver = adam), 100, 100, and 100 as hidden layer sizes (HLS = (100, 100, 100)), activation function for the hidden layer (activation = tanh), learning rate schedule for weight updates (learning rate = constant), and regulation term parameter ( $\alpha = 0.0001$ ). Finally, the best model is a random forest classifier with 100 estimators (n estimators = 100), which means that the model integrates 100 decision tree models to merge their prediction. Note that all the hyperparameters used for the generation of the classifiers are presented in Table 6, allowing for their easy replication.

**Table 6.** Mean performance of the classification methods for fatigue detection in walking task. Bold values show the best score for each performance metric

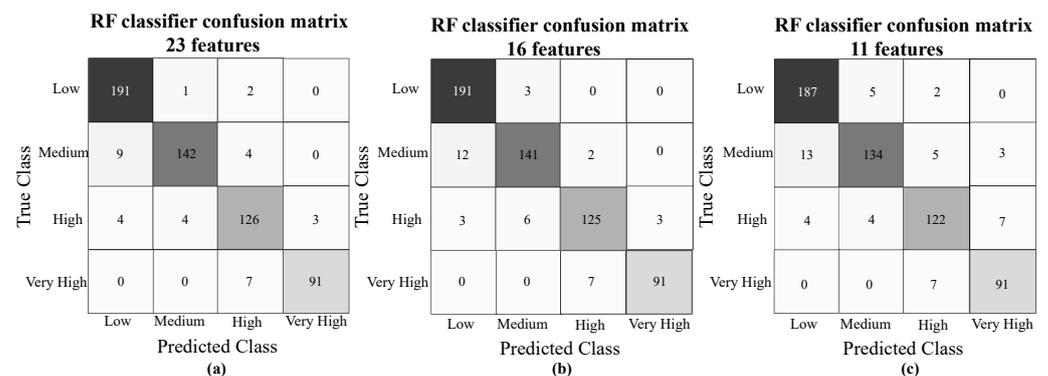
Model	Hyperparameters	Accuracy	Precision	Recall	F1-Score
RF	<b>estimators = 100</b> activation = tanh solver = adam	<b>0.965</b>	<b>0.931</b>	<b>0.929</b>	<b>0.928</b>
ANN	HLS = (100, 100, 100) alpha = 0.0001 learning_rate = 'constant' max_iter = 1000 kernel = rbf	0.949	0.896	0.898	0.894
SVM	class_weight = 'balanced' C = 64	0.907	0.809	0.809	0.806
DT	criterion = entropy max depth = 12 min samples split = 11 min samples leaf = 4	0.907	0.806	0.805	0.804
KNN	neighbors = 3	0.908	0.807	0.805	0.804
LR	solver = newton-cg C = 1,000,000	0.822	0.626	0.624	0.620

The next step is to analyze how the prediction performance varies while limiting the number of features or the number of sensors used in the study. The random forest method was selected to analyze the effect of removing features according to the highest performance metric reported in Table 6. The results of this approach are presented in Table 7, with (i) feature reduction using all sensors, and (ii) where features are limited to those from one and two sensor combinations.

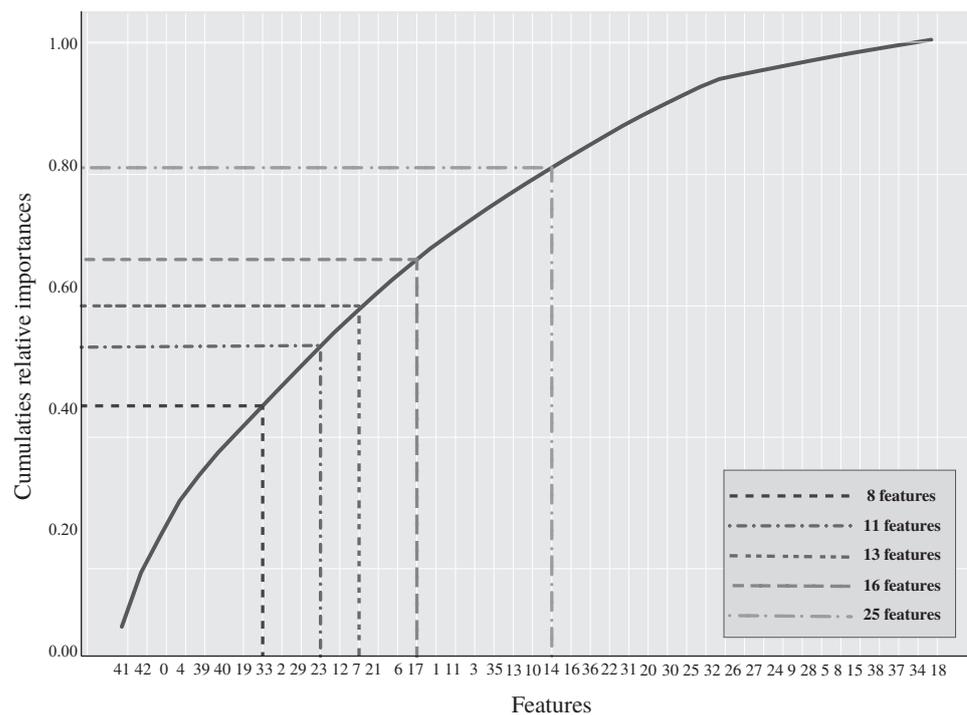
**Table 7.** Mean performance of the random forest model for fatigue detection in walking task using feature reduction and different sensors combinations. The best performance model is in **bold**.

Sensors	Estimators	Features	Accuracy	Precision	Recall	F1-Score
Thigh (EMG), Shank (EMG), L5-S1, Foot	60	25	0.965	0.934	0.928	0.930
	<b>40</b>	<b>16</b>	<b>0.965</b>	<b>0.932</b>	<b>0.927</b>	<b>0.929</b>
	80	13	0.960	0.921	0.916	0.917
	80	11	0.963	0.926	0.925	0.925
	100	8	0.946	0.895	0.883	0.888
L5-S1, Foot	80	17	0.940	0.883	0.876	0.879
L5-S1	80	16	0.839	0.678	0.658	0.664
Foot	80	19	0.921	0.856	0.827	0.838

The confusion matrix obtained from the best three RF classifiers models with feature reduction implemented after exploring in a grid search manner is shown in Figure 4, where along the x-axis are listed the true class labels and along the y-axis are the class prediction. Along the first diagonal are the correct classifications, whereas all the other entries show misclassifications.

**Figure 4.** RF classifier confusion matrix for (a) 23 features, (b) 13 features (c) 11 features.

Regarding the feature reduction used, it was essential to consider the feature importance properties for the initial RF model. The feature importance property measures a relative weight value to each feature, which represents a direct relationship with the importance of the corresponding feature for this classical machine learning model. Figure 5 presents a cumulative graph which represents the relative importance values obtained for each feature, sorted from the highest to the lowest values.



**Figure 5.** Features relative importance for random forest classifier using the original train data.

## 5. Discussion

The determination of the four fatigue states was achieved using qualitative methods such as the Borg scale, as well as quantitative methods using blood lactate measurement. This last parameter was the most considered and used to measure the performance and fatigue of individuals. This is because in response to progressively increasing exercise, lactate will increase exponentially. An individual's endurance performance is well correlated with their blood lactate [50]; therefore, lactate monitoring increases the confidence of healthcare personnel in assessing the patient's effort in physical therapies [51]—in other words, blood lactate is a direct indicator of fatigue. This, in turn, makes the fatigue classification model more accurate and reliable, since the data delivered by the sensors are directly related to what is happening physiologically with the user.

The heart rate was not considered as a feature in the classification models considering that this variable is more related to controlling the intensity of the exercise rather than determining the individual's physical fatigue (i.e., it is not directly proportional to the individual's fatigue state) [49]. For instance, given a scenario where the individual is exercising at a constant speed, the HR may remain stable and the individual may be experiencing a level of fatigue which is not reflected in the measurement, and vice versa. The second reason for not considering HR was bearing in mind that it is determined by different variables such as age, the physical condition of the individual, comorbidities, and gender; therefore, when standardizing this variable it may cause a degree of uncertainty [47,48].

According to the dataset distribution presented in Table 5, the main difference between classes was 16.79%, which is considered as a slight imbalance, and it is acceptable for data analysis and training computational models [117,118]. Considering the report by Fernandez et al. [116] that a slight imbalance can often be treated like a normal classification predictive modeling problem as long as true negatives are considered, the selection of the best model was performed mainly based on the F-score.

The fourth main observations from the predictive performance of our six models presented in Table 6 need to be highlighted. First, as expected from the preliminary analysis, the classifiers implemented with training/test presented a higher performance in all metrics than the classifiers implemented with cross-validation. Hence, a simple training/test split is sufficient for larger datasets. Second, the performance of the five

models, except for the LR classifier, is relatively high with an overall average F-score greater than 80.4%. Third, according to the literature review, it was expected that the LR classifier presented a better performance given the positive results obtained with this classifier in [38,73,74]. However, this model presented the lowest performance with a 62% F-score, which would not represent an optimal or good classifier. This can be associated with two main factors: (i) These studies considered only two fatigue conditions: fatigue or non-fatigued, whereas this work contemplated four fatigue states. It increased the probability of failing in the estimation and suggested that this classifier performed better for binary classifications problems. (ii) Regarding the lack of understanding of how an individual's performance deteriorates with fatigue accumulation, which can vary based on user conditions, physical activity, features extracted, and fatigue detection technique, it is essential to have a general framework for fatigue estimation classifiers as presented in Section 3.2 to guide the implementation, evaluation, and continuous improvement of fatigue monitoring in rehabilitation scenarios regardless of physical activity or user conditions. Fourth, the RF model presented the best performance in all features with an overall of 92% considered as an optimum classifier for the fourth state estimation of fatigue in walking tasks.

Once the best prediction model is identified, the next logical research question was to examine how the prediction performance varies while limiting the number of features and sensors used. To evaluate this question, we utilize the random forest model since Table 6 showed that it had the highest mean accuracy, precision, recall, and F-Score when compared to the other classifiers. From this, a list of predictive/important features was established from most important to least important (see Figure 5) and a feature reduction was applied to have a more interpretable classifier. Table 7 reports the prediction results when features are limited to those from 25, 16, 13, 11, and 8. From the results in Table 7, one can see that the prediction performance does not vary significantly as the number of features' is reduced. Some even match the RF performance utilizing all features (i.e., 43 features). Considering that the RF model presented a similar performance among 23, 16, and 11 features, their confusion matrix was analyzed to observed the prediction performance in two minority classes, i.e., very high fatigue and high fatigue states. These predictions are more interesting and more valuable since they are essential to avoid any risk in rehabilitation scenarios. The RF classifier presented fewer misclassifications between the low class and the two minority classes with 16 features illustrated in Figure 4b. Based on this observation, it is recommended to use the RF classifier with 16 features with an overall of 91.5% in all features. While the prediction performance is almost the same, the unnecessary features from the data were removed, optimizing computational costs and running time. Moreover, a smaller number of features facilitates the interpretation of the model, which is essential in the fatigue identification and diagnosis phases.

The prediction performance variation, while limiting the sensor used, is also reported in Table 6. Note that the sensor used to measure EMG is removed considering that EMG is invasive, which complicates its daily usage for real-time fatigue detection. Hence, the characteristics obtained by the EMG sensor were not considered. This means that the main features that detected fatigue were eliminated according to the sensor used. As expected, without EMG measurements the classifier presented a performance reduction since these features presented great importance in fatigue detection (see Figure 5). However, the classifier performance remains positive using two sensors (IMUS placed on footstep and around L2-S1) with an F-score of 88% and an overall average of 94%. Similarly, using one IMU located in the footstep has a performance of 84% and an overall average of 92%. Based on this observation, it suggests using only the IMU located on foot. While the prediction performance is almost the same, the costs incurred by the clinics are much lower, and the usability of the system by using only one sensor is significantly improved. These results are comparable with the studies that proposed fatigue estimation models during the walking task [69,71,73,75], which have shown accuracy values among 77% and 92%, with only two fatigue conditions—fatigued and non-fatigued. Besides, these studies [69,71,75] use motion

capture systems as fatigue detection techniques that often require special setups, which makes them better suited for controlled environments than real rehabilitation scenarios.

Considering these observations, one can indicate three main contributions. First, the framework proposed has shown higher detection performance (with fewer features) and detects four fatigue diagnosis states in walking tasks, which allows clinicians to better monitor the patient and pinpoint the hazard, and to prescribe and manage interventions according to each individual's needs. Second, the insights from the fatigue identification phase of our framework can be used to inform sensor placement and selection. Third, and more importantly, our framework presents a systematic approach that can answer the question: "what are the gains associated with wearing an extra sensor?" In essence, this question is left open to the researchers and practitioners to attempt to quantify whether the hassle and cost associated with wearing an extra sensor can be justified by a significant/practical improvement in fatigue detection when developing models for detecting/managing fatigue in other settings or their target application. The results and data used in this study can be accessed through the link provided in the Data Availability Statement Section. These codes can be used to develop predictive physical fatigue models.

The implementation of our proposed model requires an understanding of the specific features that were selected in the feature reduction process. Significant features contributing to the determination of physical fatigue in this model include:

- EMG RMS signals (features 41, 42, 39, 40) represent the square root of the average power of the EMG signal for a given period. Decrease over time of these signals led to the detection of muscle fatigue.
- Gait Acceleration Mean (feature 0) reflects the mean duration of each gait cycle. The fatigued musculoskeletal system is less able to attenuate heel strike-initiated shock waves, which could be observed as an increase in the amplitude of the acceleration measured at the foot. If the mean gait cycle time increased significantly with elapsed walking time indicates that the individual is fatigued.
- Gait Acceleration Median (feature 4) is the median value for each gait cycle acceleration.
- Spine Acceleration Mean (features 19 and 29) represents the torso acceleration over each gait cycle. These features show that if participants kept consistent torso movement over gait cycles, then it likely corresponds to their walking behavior and the patient is less likely to report physical fatigue.
- Spine Acceleration Median (features 33 and 23) is a measure of the central tendency of the torso acceleration distribution. Where the participant maintains a high level of spine acceleration, they are more likely to feel physically fatigued.
- Gait Maximum Acceleration (features 2 and 12) as the gait cycle time increased significantly with increasing fatigue, gait acceleration decreases. If the participant reduces their walking speed, then a decrease in peak gait acceleration is generated, indicating that the participant is fatigued.

The selected features for the best model for physical fatigue detection were shown in Figure 5. They are consistent with previous studies that have used IMUs for monitoring physical activity. Common features computed from the acceleration signal are the mean [68,94,95,98,100], variance or standard deviation [68,95,101,120] and the entropy and energy of the data [61,68].

Concerning the results using only one IMU located around the center of L5-S1, according to the literature review, it was expected to have a better classifier performance considering that the torso is the body part commonly used for physical fatigue detection and development models as reported in the following studies [38,71,73–75]. Specifically, Mamman et al. [38] reported that one IMU located on the torso was enough to detect fatigue in manual material handling environments. However, our results differ from the results obtained in previous studies. In this study we found that the mean back rotational position was selected as a mainly important feature for the classifier, as illustrated in Figure 5 with features numbers 19, 33, 29, and 23. The results showed that the accuracy and F-score were reduced to 88% and 66%, respectively, which is not promising. The above may suggest

that (i) The features extracted from this sensor differ from other related studies; hence, the feature extraction methods influenced the classification performance, and should be improved. (ii) The use of a single sensor placed at L5-S1 in walking tasks may not be sufficient and consequently, the performance of the classifier decreases significantly. (iii) The fourth fatigue diagnoses states may increase the probability of failing in fatigue estimation with the use of this sensor in relation to only two fatigue diagnostic conditions that have been presented in the previous studies. Therefore, as future work, it is suggested to revise the obtained features to improve the performance of the classification model.

Similarly, we found that the mean back rotational position was selected as the most important feature, as illustrated in Figure 5 with features numbers 19, 33, 29, and 23. Moreover, Mamman et al. [38] reported that an IMU located the torso was enough for detecting fatigue in manual material handling environments. However, the features utilized differ from the features extracted in this work. The above may suggest three things: (i) the feature extraction methods influenced the classification performance, and hence feature extraction with this sensor needs to be improved; (ii) the classifier performance decrease significantly with the use of only one sensor placed on L5-S1 in the walking tasks; and (iii) fourth fatigue diagnoses states may increase the probability of failing in fatigue estimation concerning only two fatigue diagnoses conditions. Therefore, as future work, it is suggested to review the characteristics obtained to improve the performance of the classification model.

On the other hand, there are a few limitations that must be acknowledged for this study. First, the sample size is small due to the confinement caused by the COVID-19 bio-safety protocols and the time required for each participant. However, the sample size is consistent with other studies that have focused on lab-based modeling of physical fatigue presented in Table 1. Second, the effect of different demographic variables (i.e., age, sex, physical condition) needs to be explored in future models of physical fatigue due to all volunteers being healthy people, and the features may show different behaviors and patterns with patients or other groups with various physical conditions. Third, the evaluation of our framework's performance was limited to focused lab experiments. Future studies should evaluate how this framework performs in clinical scenarios.

Finally, a future and great application would be to implement this physical fatigue detection model in SAR systems. In essence, the system is proposed to be composed of a single-sensor interface, aiming to measure all features relevant to the RF classifier. The interface architecture comprises different nodes (e.g., sensor and SAR nodes) to adequately handle the difference in sampling rates and the amount of information acquired by the system. In the case of the sensor interface, the output of each node is the data processed, and for the the corresponding behavior and feedback of the SAR node. The data from the sensors would be stored in a database to be analyzed online to manage the interaction of the SAR with the user. Regarding the SAR's behavior, the node should be designed considering three situations [81,121] (e.g., motivation, warning, and emergency, which are triggered depending on the data provided by the interface) while monitoring the users' performance, with the aim of positively influencing the constant monitoring of user's performance and providing feedback to motivate them while the physical exercise intensity and the correct exercise execution is controlled. Besides, SARs should be designed to communicate with the therapists if an emergency event occurs during therapy (e.g., physical fatigue above the maximum allowable level, dizziness or any parameter is out of normal). The aforementioned allows to achieve a successful rehabilitation program through continuous patient monitoring and providing feedback to clinicians about the patient's performance to avoid overtraining, which can affect their rehabilitation and even induce physical suffering or physiological complications. It should be emphasized that this proposal is also applicable to other robot-assisted physical exercise scenarios, since this model allows for a close measurement of users' physiological events.

## 6. Conclusions

Physical fatigue is a significant safety concern in rehabilitation environments and monitoring physical fatigue is essential to prevent an accident, injury occurrences, and for a rehabilitation program to be successful. The utilization of predictive models for physical fatigue modeling can provide a chance to understand the physiology and psychology of fatigue better. In this paper, an integrated framework was proposed with the main steps for developing a fatigue classification model using few features. It facilitates the interpretation model, which is essential in fatigue identification and diagnosis through visual analytical approaches allowing practitioners to identify risks which should be addressed through an appropriate intervention strategy.

The classifier models implemented considered four fatigue stages, instead of only determining whether the user is fatigued or not fatigued, as previous studies have presented, that allow to improve the fatigue monitoring and enable clinicians to prescribe an optimal therapeutic dose according to the individual's needs, avoiding injuries and affectations to the process. The classifier is able to obtain greater accuracy, despite the fact that identifying the four fatigue states increases the probability of error. In particular, an accuracy of 96% and an F-score of 93% were obtained with random forest model, generating an optimal classifier.

The determination of the four fatigue states was achieved using blood lactate measurement. This parameter was the most used to measure performance and fatigue of individuals, thus being a direct indicator of fatigue. This in turn makes the method more accurate and reliable, since the data delivered by the sensors are directly related to what happens physiologically with the user.

The classifier performance model was also analyzed, checking the number of features and sensors used in future real-time applications. The dimensionality of data is vital to optimizing computational costs and running time. In addition, with fewer features, the total number of sensors required to estimate them could be reduced. The results showed that the classifier is capable of using a single sensor, maintaining an accuracy and F-score of 94% and 92%, respectively. Therefore, clinicians will also be more inclined to adopt the framework if it requires fewer sensors since it will: (i) be much cheaper, for example, requiring one or two IMUs instead of four, which would reduce the cost; (ii) make the process less invasive to the patient; and (iii) reduce the time needed for the patient to wear and strap on all the sensors.

**Author Contributions:** Conceptualization, M.M. and C.A.C.; methodology, M.M. and M.J.P.-B.; software, M.J.P.-B. and O.P.; validation, M.M., O.P. and M.J.P.-B.; formal analysis, M.M., O.P., M.J.P.-B. and C.A.C.; supervision, M.M. and C.A.C.; writing—original draft preparation, M.J.P.-B.; writing—review and editing, M.M., M.J.P.-B., O.P., M.R.-R. and C.A.C. All authors contributed to the article and approved the submitted version. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Ministerio de Ciencia Tecnología e Innovación-Colombia (MinCiencias Grant ID No. 813-2017).

**Institutional Review Board Statement:** All participants gave their informed consent for inclusion before they participated in the study, i.e, informed consent was obtained for the patient(s) to publish this paper. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the ethics Committee of Colombian School of Engineering Julio Garavito (Project identification code: 813-2017; Protocol code: 04-19; date: 23 September 2019).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. This data can be found here: [https://figshare.com/projects/Fourth\\_fatigue\\_diagnosis\\_states/119580](https://figshare.com/projects/Fourth_fatigue_diagnosis_states/119580) (accessed on 12 September 2021).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Salakari, M.R.; Surakka, T.; Nurminen, R.; Pylkkänen, L. Effects of rehabilitation among patients with advanced cancer: A systematic review. *Acta Oncol.* **2015**, *54*, 618–628. [[CrossRef](#)]
2. Zanuso, S.; Balducci, S.; Jimenez, A. Physical activity, a key factor to quality of life in type 2 diabetic patients. *Diabetes/Metab. Res. Rev.* **2009**, *25*, S24–S28. [[CrossRef](#)]
3. Zanuso, S.; Jimenez, A.; Pugliese, G.; Corigliano, G.; Balducci, S. Exercise for the management of type 2 diabetes: A review of the evidence. *Acta Diabetol.* **2010**, *47*, 15–22. [[CrossRef](#)] [[PubMed](#)]
4. Warburton, D.E.; Nicol, C.W.; Bredin, S.S. Health benefits of physical activity: The evidence. *CMAJ* **2006**, *174*, 801–809. [[CrossRef](#)] [[PubMed](#)]
5. Bauman, A.E. Updating the evidence that physical activity is good for health: An epidemiological review 2000–2003. *J. Sci. Med. Sport* **2004**, *7*, 6–19. [[CrossRef](#)]
6. Oguma, Y.; Shinoda-Tagawa, T. Physical activity decreases cardiovascular disease risk in women: review and meta-analysis. *Am. J. Prev. Med.* **2004**, *26*, 407–418. [[CrossRef](#)] [[PubMed](#)]
7. Vuori, I. Physical inactivity is a cause and physical activity is a remedy for major public health problems. *Kinesiology* **2004**, *36*, 123–153.
8. Haskell, W.L.; Lee, I.M.; Pate, R.R.; Powell, K.E.; Blair, S.N.; Franklin, B.A.; Macera, C.A.; Heath, G.W.; Thompson, P.D.; Bauman, A. Physical Activity and Public Health. *Med. Sci. Sport. Exerc.* **2007**, *39*, 1423–1434. [[CrossRef](#)]
9. Pinto-Bernal, M.J.; Aguirre, A.; Cifuentes, C.A.; Munera, M. Wearable Sensors for Monitoring Exercise and Fatigue Estimation in Rehabilitation. In *Internet of Medical Things*; CRC Press: Boca Raton, FL, USA, 2021; pp. 83–110.
10. Kristensen, T.; Kornitzer, M.; Alfredsson, L.; Marmot, M.; Logstrup, S.; Williams, C. *Social Factors, Work, Stress and Cardiovascular Disease Prevention in the European Union*; European Heart Network: Brussels, Belgium, 1998.
11. Priest, N.; Armstrong, R.; Doyle, J.; Waters, E. Interventions implemented through sporting organisations for increasing participation in sport. *Cochrane Database Syst. Rev.* **2008**, *18*, CD004812. [[CrossRef](#)]
12. Livingstone, M.; Robson, P.; Wallace, J.; McKinley, M. How active are we? Levels of routine physical activity in children and adults. *Proc. Nutr. Soc.* **2003**, *62*, 681–701. [[CrossRef](#)]
13. Pollock, M.L.; Gaesser, G.A.; Butcher, J.D.; Després, J.P.; Dishman, R.K.; Franklin, B.A.; Garber, C.E. The recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness, and flexibility in healthy adults. *Schweiz. Z. Sportmed.* **1998**, *41*, 127–137. [[CrossRef](#)] [[PubMed](#)]
14. Andersen, L.B.; Schnohr, P.; Schroll, M.; Hein, H.O. All-Cause Mortality Associated with Physical Activity during Leisure Time, Work, Sports, and Cycling to Work. *Arch. Intern. Med.* **2000**, *160*, 1621–1628. [[CrossRef](#)] [[PubMed](#)]
15. Schnohr, P.; Marott, J.L.; Jensen, J.S.; Jensen, G.B. Intensity versus duration of cycling, impact on all-cause and coronary heart disease mortality: The Copenhagen City Heart Study. *Eur. J. Prev. Cardiol.* **2012**, *19*, 73–80. [[CrossRef](#)] [[PubMed](#)]
16. Warburton, D.E. Prescribing exercise as preventive therapy. *Can. Med. Assoc. J.* **2006**, *174*, 961–974. [[CrossRef](#)] [[PubMed](#)]
17. Cup, E.H.; Pieterse, A.J.; ten Broek-Pastoor, J.M.; Munneke, M.; van Engelen, B.G.; Hendricks, H.T.; van der Wilt, G.J.; Oostendorp, R.A. Exercise Therapy and Other Types of Physical Therapy for Patients With Neuromuscular Diseases: A Systematic Review. *Arch. Phys. Med. Rehabil.* **2007**, *88*, 1452–1464. [[CrossRef](#)] [[PubMed](#)]
18. Manley, A.F. *Physical Activity and Health: A Report of the Surgeon General*; Diane Publishing: Darby, PA, USA, 1996.
19. Lee, I.M.; Sesso, H.D.; Oguma, Y.; Paffenbarger, R.S. Relative intensity of physical activity and risk of coronary heart disease. *Circulation* **2003**, *107*, 1110–1116. [[CrossRef](#)]
20. American College of Sports Medicine. *ACSM's Health-Related Physical Fitness Assessment Manual*; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2013.
21. Balducci, S.; Sacchetti, M.; Haxhi, J.; Orlando, G.; D'Errico, V.; Fallucca, S.; Menini, S.; Pugliese, G. Physical exercise as therapy for type 2 diabetes mellitus. *Diabetes/Metab. Res. Rev.* **2014**, *30*, 13–23. [[CrossRef](#)]
22. Dun, Y.; Smith, J.R.; Liu, S.; Olson, T.P. High-Intensity Interval Training in Cardiac Rehabilitation. *Sports Med.* **2019**, *42*, 587–605. [[CrossRef](#)]
23. Tanasescu, M.; Leitzmann, M.F.; Rimm, E.B.; Willett, W.C.; Stampfer, M.J.; Hu, F.B. Exercise type and intensity in relation to coronary heart disease in men. *J. Am. Med. Assoc.* **2002**, *288*, 1994–2000. [[CrossRef](#)]
24. Oldervoll, L.; Kaasa, S.; Hjerstad, M.; Lund, J.; Loge, J. Physical exercise results in the improved subjective well-being of a few or is effective rehabilitation for all cancer patients? *Eur. J. Cancer* **2004**, *40*, 951–962. [[CrossRef](#)]
25. Fleig, L.; Lippke, S.; Pomp, S.; Schwarzer, R. Exercise maintenance after rehabilitation: How experience can make a difference. *Psychol. Sport Exerc.* **2011**, *12*, 293–299. [[CrossRef](#)]
26. Göhner, W.; Seelig, H.; Fuchs, R. Intervention effects on cognitive antecedents of physical exercise: A 1-year follow-up study. *Appl. Psychol. Health Well-Being* **2009**, *1*, 233–256. [[CrossRef](#)]
27. Abd-Elfattah, H.M.; Abdelazeim, F.H.; Elshennawy, S. Physical and cognitive consequences of fatigue: A review. *J. Adv. Res.* **2015**, *6*, 351–358. [[CrossRef](#)]
28. Baussard, L.; Carayol, M.; Porro, B.; Baguet, F.; Cousson-gelie, F. European Journal of Oncology Nursing Fatigue in cancer patients : Development and validation of a short form of the Multidimensional Fatigue Inventory ( MFI-10 ). *Eur. J. Oncol. Nurs.* **2018**, *36*, 62–67. [[CrossRef](#)] [[PubMed](#)]

29. Alghannam, A.F.; Tsintzas, K.; Thompson, D.; Bilzon, J.; Betts, J.A. Exploring mechanisms of fatigue during repeated exercise and the dose dependent effects of carbohydrate and protein ingestion: Study protocol for a randomised controlled trial. *Trials* **2014**, *15*, 95. [[CrossRef](#)]
30. Ozalp, O.; Inal-Ince, D.; Calik, E.; Vardar-Yagli, N.; Saglam, M.; Savci, S.; Arikan, H.; Bosnak-Guclu, M.; Coplu, L. Extrapulmonary features of bronchiectasis: Muscle function, exercise capacity, fatigue, and health status. *Multidiscip. Respir. Med.* **2012**, *7*, 3. [[CrossRef](#)]
31. Lu, L.; Megahed, F.M.; Sesek, R.F.; Cavuoto, L.A. A survey of the prevalence of fatigue, its precursors and individual coping mechanisms among US manufacturing workers. *Appl. Ergon.* **2017**, *65*, 139–151. [[CrossRef](#)]
32. Zamuner, A.R.; Moreno, M.A.; Camargo, T.M.; Graetz, J.P.; Rebelo, A.C.; Tamburús, N.Y.; da Silva, E. Assessment of subjective perceived exertion at the anaerobic threshold with the Borg CR-10 scale. *J. Sport. Sci. Med.* **2011**, *10*, 130–136.
33. Curt, G.A.; Breitbart, W.; Cella, D.; Groopman, J.E.; Horning, S.J.; Itri, L.M.; Johnson, D.H.; Miaskowski, C.; Scherr, S.L.; Portenoy, R.K.; et al. Impact of cancer-related fatigue on the lives of patients: New findings from the Fatigue Coalition. *Oncologist* **2000**, *5*, 353–360. [[CrossRef](#)] [[PubMed](#)]
34. Annett, J. Subjective rating scales: Science or art? *Ergonomics* **2002**, *45*, 966–987. [[CrossRef](#)] [[PubMed](#)]
35. Williams, N. The Borg rating of perceived exertion (RPE) scale. *Occup. Med.* **2017**, *67*, 404–405. [[CrossRef](#)]
36. Borg, G. Borg's range model and scales. *Int. J. Sport Psychol.* **2001**, *32*, 110–126.
37. Schle, A.; Vieten, M.; Sailer, S.; Mündermann, A.; Dettmers, C. Objective assessment of motor fatigue in multiple sclerosis: The Fatigue index Kliniken Schmieder (FKS). *J. Neurol.* **2014**, *261*, 1752–1762. [[CrossRef](#)] [[PubMed](#)]
38. Maman, Z.S.; Chen, Y.J.; Baghdadi, A.; Lombardo, S.; Cavuoto, L.A.; Megahed, F.M. A data analytic framework for physical fatigue management using wearable sensors. *Expert Syst. Appl.* **2020**, *155*, 113405. [[CrossRef](#)]
39. Qi, J.; Yang, P.; Waraich, A.; Deng, Z.; Zhao, Y.; Yang, Y. Examining sensor-based physical activity recognition and monitoring for healthcare using Internet of Things: A systematic review. *J. Biomed. Inform.* **2018**, *87*, 138–153. [[CrossRef](#)]
40. Zeni, A.I.; Hoffman, M.D.; Clifford, P.S. Relationships among heart rate, lactate concentration, and perceived effort for different types of rhythmic exercise in women. *Arch. Phys. Med. Rehabil.* **1996**, *77*, 237–241. [[CrossRef](#)]
41. Poole, D.C.; Burnley, M.; Vanhatalo, A.; Rossiter, H.B.; Jones, A.M. Critical power: An important fatigue threshold in exercise physiology. *Med. Sci. Sport. Exerc.* **2016**, *48*, 2320–2334. [[CrossRef](#)]
42. Pettersson, S.; Lundberg, I.; Liang, M.; Pouchot, J.; Welin Henriksson, E. Determination of the minimal clinically important difference for seven measures of fatigue in Swedish patients with systemic lupus erythematosus. *Scand. J. Rheumatol.* **2015**, *44*, 206–210. [[CrossRef](#)]
43. Yu, F.; Bilberg, A.; Stenager, E.; Rabotti, C.; Zhang, B.; Mischi, M. A wireless body measurement system to study fatigue in multiple sclerosis. *Physiol. Meas.* **2012**, *33*, 2033–2048. [[CrossRef](#)]
44. Möhler, F.; Ringhof, S.; Debertin, D.; Stein, T. Influence of fatigue on running coordination: A UCM analysis with a geometric 2D model and a subject-specific anthropometric 3D model. *Hum. Mov. Sci.* **2019**, *66*, 133–141. [[CrossRef](#)]
45. Kang, S.R.; Min, J.Y.; Yu, C.; Kwon, T.K. Effect of whole body vibration on lactate level recovery and heart rate recovery in rest after intense exercise. *Technol. Health Care* **2017**, *25*, 115–123. [[CrossRef](#)] [[PubMed](#)]
46. Glynn, A.J.; Fiddler, H. *The Physiotherapist's Pocket Guide to Exercise E-Book: Assessment, Prescription and Training*; Elsevier Health Sciences: Amsterdam, The Netherlands, 2009.
47. Aubert, A.E.; Seps, B.; Beckers, F. Heart rate variability in athletes. *Sport. Med.* **2003**, *33*, 889–919. [[CrossRef](#)]
48. Achten, J.; Jeukendrup, A.E. Heart rate monitoring. *Sport. Med.* **2003**, *33*, 517–538. [[CrossRef](#)]
49. da Cunha, F.A.; Farinatti, P.d.T.V.; Midgley, A.W. Methodological and practical application issues in exercise prescription using the heart rate reserve and oxygen uptake reserve methods. *J. Sci. Med. Sport.* **2011**, *14*, 46–57. [[CrossRef](#)]
50. Goodwin, M.L.; Harris, J.E.; Hernández, A.; Gladden, L.B. Blood lactate measurements and analysis during exercise: A guide for clinicians. *J. Diabetes Sci. Technol.* **2007**, *1*, 558–569. [[CrossRef](#)]
51. Jansen, T.C.; van Bommel, J.; Bakker, J. Blood lactate monitoring in critically ill patients: A systematic health technology assessment. *Crit. Care Med.* **2009**, *37*, 2827–2839.
52. Saey, D.; Michaud, A.; Couillard, A.; Côté, C.H.; Mador, M.J.; LeBlanc, P.; Jobin, J.; Maltais, F. Contractile fatigue, muscle morphometry, and blood lactate in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **2005**, *171*, 1109–1115. [[CrossRef](#)]
53. Helbostad, J.L.; Sturnieks, D.L.; Menant, J.; Delbaere, K.; Lord, S.R.; Pijnappels, M. Consequences of lower extremity and trunk muscle fatigue on balance and functional tasks in older people: A systematic literature review. *BMC Geriatr.* **2010**, *10*, 56. [[CrossRef](#)]
54. Wan, J.-J.; Qin, Z.; Wang, P.-Y.; Sun, Y.; Liu, X. Muscle fatigue: General understanding and treatment. *Exp. Mol. Med.* **2017**, *49*, e384. [[CrossRef](#)]
55. Karthick, P.A.; Ghosh, D.M.; Ramakrishnan, S. Surface electromyography based muscle fatigue detection using high-resolution time-frequency methods and machine learning algorithms. *Comput. Methods Programs Biomed.* **2018**, *154*, 45–56. [[CrossRef](#)] [[PubMed](#)]
56. Subasi, A.; Kiyimik, M.K. Muscle fatigue detection in EMG using time-frequency methods, ICA and neural networks. *J. Med. Syst.* **2010**, *34*, 777–785. [[CrossRef](#)] [[PubMed](#)]

57. Al-Mulla, M.R.; Sepulveda, F.; Colley, M. A Review of Non-Invasive Techniques to Detect and Predict Localised Muscle Fatigue. *Sensors* **2011**, *11*, 3545–3594. [[CrossRef](#)]
58. Camomilla, V.; Bergamini, E.; Fantozzi, S.; Vannozzi, G. Trends Supporting the In-Field Use of Wearable Inertial Sensors for Sport Performance Evaluation: A Systematic Review. *Sensors* **2018**, *18*, 873. [[CrossRef](#)] [[PubMed](#)]
59. Ejupi, A.; Gschwind, Y.J.; Valenzuela, T.; Lord, S.R.; Delbaere, K. A Kinect and Inertial Sensor-Based System for the Self-Assessment of Fall Risk: A Home-Based Study in Older People. *Hum.-Comput. Interact.* **2016**, *31*, 261–293. [[CrossRef](#)]
60. Manchola, S.; Bernal, P.; Munera, M.; Cifuentes, C.A. Gait Phase Detection for Lower-Limb Exoskeletons using Foot Motion Data from a Single Inertial Measurement Unit in Hemiparetic Individuals. *Sensors* **2019**, *19*, 2988. [[CrossRef](#)] [[PubMed](#)]
61. Aguirre, A.; Casas, J.; Céspedes, N.; Múnera, M.; Rincon-Roncancio, M.; Cuesta-Vargas, A.; Cifuentes, C.A. Feasibility study: Towards Estimation of Fatigue Level in Robot-Assisted Exercise for Cardiac Rehabilitation. In Proceedings of the 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR), Toronto, ON, Canada, 24–28 June 2019; pp. 911–916.
62. Céspedes, N.; Múnera, M.; Gómez, C.; Cifuentes, C.A. Social Human-Robot Interaction for Gait Rehabilitation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 1299–1307. [[CrossRef](#)]
63. Segala, D.B.; Chelidze, D.; Adams, A.; Schiffman, J.M.; Hasselquist, L. Tracking Physiological Fatigue in Prolonged Load Carriage Walking Using Phase Space Warping and Smooth Orthogonal Decomposition. In Proceedings of the ASME International Mechanical Engineering Congress and Exposition, Boston, MA, USA, 31 October–6 November 2008; Volume 48630, pp. 323–331.
64. Mugnosso, M.; Marini, F.; Holmes, M.; Morasso, P.; Zenzeri, J. Muscle fatigue assessment during robot-mediated movements. *J. Neuroeng. Rehabil.* **2018**, *15*, 1–14. [[CrossRef](#)]
65. Chan, V.C.; Beaudette, S.M.; Smale, K.B.; Beange, K.H.; Graham, R.B. A subject-specific approach to detect fatigue-related changes in spine motion using wearable sensors. *Sensors* **2020**, *20*, 2646. [[CrossRef](#)]
66. Ross, L.M.; Porter, R.R.; Durstine, J.L. High-intensity interval training (HIIT) for patients with chronic diseases. *J. Sport Health Sci.* **2016**, *5*, 139–144. [[CrossRef](#)]
67. García-López, J.; Morante, J.C.; Ogueta-Alday, A.; Rodríguez-Marroyo, J.A. The Type Of Mat (Contact vs. Photocell) Affects Vertical Jump Height Estimated From Flight Time. *J. Strength Cond. Res.* **2013**, *27*, 1162–1167. [[CrossRef](#)]
68. Aguirre, A.; Pinto, M.J.; Cifuentes, C.A.; Perdomo, O.; Díaz, C.A.; Múnera, M. Machine Learning Approach for Fatigue Estimation in Sit-to-Stand Exercise. *Sensors* **2021**, *21*, 5006. [[CrossRef](#)]
69. Zhang, J.; Lockhart, T.E.; Soangra, R. Classifying lower extremity muscle fatigue during walking using machine learning and inertial sensors. *Ann. Biomed. Eng.* **2014**, *42*, 600–612. [[CrossRef](#)] [[PubMed](#)]
70. Karg, M.; Venture, G.; Hoey, J.; Kulić, D. Human movement analysis as a measure for fatigue: A hidden Markov-based approach. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2014**, *22*, 470–481. [[CrossRef](#)] [[PubMed](#)]
71. Karg, M.; Kühnlenz, K.; Buss, M.; Seiberl, W.; Tusker, F.; Schmeelk, M.; Schwirtz, A. Expression and automatic recognition of exhaustion in natural walking. In Proceedings of the IADIS Interfaces and Human Computer Interaction (IHCI), Amsterdam, The Netherlands, 25–27 July 2008; pp. 165–172.
72. Kavanagh, J.J.; Morrison, S.; Barrett, R.S. Lumbar and cervical erector spinae fatigue elicit compensatory postural responses to assist in maintaining head stability during walking. *J. Appl. Physiol.* **2006**, *101*, 1118–1126. [[CrossRef](#)]
73. Yoshino, K.; Motoshige, T.; Araki, T.; Matsuoka, K. Effect of prolonged free-walking fatigue on gait and physiological rhythm. *J. Biomech.* **2004**, *37*, 1271–1280. [[CrossRef](#)] [[PubMed](#)]
74. Maman, Z.S.; Yazdi, M.A.A.; Cavuoto, L.A.; Megahed, F.M. A data-driven approach to modeling physical fatigue in the workplace using wearable sensors. *Appl. Ergon.* **2017**, *65*, 515–529. [[CrossRef](#)] [[PubMed](#)]
75. Lee, M.; Roan, M.; Smith, B.; Lockhart, T.E. Gait analysis to classify external load conditions using linear discriminant analysis. *Hum. Mov. Sci.* **2009**, *28*, 226–235. [[CrossRef](#)]
76. Helbostad, J.L.; Leirfall, S.; Moe-Nilssen, R.; Sletvold, O. Physical fatigue affects gait characteristics in older persons. *J. Gerontol. Ser. Biol. Sci. Med. Sci.* **2007**, *62*, 1010–1015. [[CrossRef](#)]
77. Winter, D.A. Human balance and posture control during standing and walking. *Gait Posture* **1995**, *3*, 193–214. [[CrossRef](#)]
78. Warburton, D.E.; Gledhill, N.; Quinney, A. Musculoskeletal fitness and health. *Can. J. Appl. Physiol.* **2001**, *26*, 217–237. [[CrossRef](#)]
79. Swift-Spong, K.; Short, E.; Wade, E.; Matarić, M.J. Effects of comparative feedback from a socially assistive robot on self-efficacy in post-stroke rehabilitation. In Proceedings of the 2015 IEEE International Conference on Rehabilitation Robotics (ICORR), Singapore, 11–14 August 2015; pp. 764–769.
80. Fasola, J.; Matarić, M.J. A socially assistive robot exercise coach for the elderly. *J. Hum.-Robot Interact.* **2013**, *2*, 3–32. [[CrossRef](#)]
81. Casas, J.; Senft, E.; Gutierrez, L.F.; Rincon-Roncancio, M.; Munera, M.; Belpaeme, T.; Cifuentes, C.A. Social assistive robots: Assessing the impact of a training assistant robot in cardiac rehabilitation. *Int. J. Soc. Robot.* **2020**, 1–15. [[CrossRef](#)]
82. Cifuentes, C.A.; Pinto, M.J.; Céspedes, N.; Múnera, M. Social robots in therapy and care. *Curr. Robot. Rep.* **2020**, *1*, 59–74. [[CrossRef](#)]
83. Céspedes Gómez, N.; Irfan, B.; Senft, E.; Cifuentes, C.A.; Gutierrez, L.F.; Rincon-Roncancio, M.; Belpaeme, T.; Munera, M. A Socially Assistive Robot for Long-Term Cardiac Rehabilitation in the Real World. *Front. Neurobot.* **2021**, *15*, 21.
84. Gockley, R.; Bruce, A.; Forlizzi, J.; Michalowski, M.; Mundell, A.; Rosenthal, S.; Sellner, B.; Simmons, R.; Snipes, K.; Schultz, A.C.; et al. Designing robots for long-term social interaction. In Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, AB, Canada, 2–6 August 2005; pp. 1338–1343.

85. Gockley, R.; Matarić, M.J. Encouraging physical therapy compliance with a hands-off mobile robot. In Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human–Robot Interaction, Salt Lake City, UT, USA, 2–3 March 2006; pp. 150–155.
86. Matarić, M.J.; Eriksson, J.; Feil-Seifer, D.J.; Winstein, C.J. Socially assistive robotics for post-stroke rehabilitation. *J. Neuroeng. Rehabil.* **2007**, *4*, 1–9. [[CrossRef](#)] [[PubMed](#)]
87. Smets, E.M.; Garssen, B.; Bonke, B.; De Haes, J.C. The multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *J. Psychosom. Res.* **1995**, *39*, 315–325. [[CrossRef](#)]
88. Kakria, P.; Tripathi, N.; Kitipawang, P. A real-time health monitoring system for remote cardiac patients using smartphone and wearable sensors. *Int. J. Telemed. Appl.* **2015**, *2015*, 373474. [[CrossRef](#)]
89. Moohialdin, A.S.; Suhariadi, B.T.; Siddiqui, M.K. Practical validation measurements of a physiological status monitoring sensor in real construction activities. In Proceedings of the Streamlining Information Transfer between Construction and Structural Engineering, Brisbane, Australia, 3–5 December 2018.
90. Swain, D.P.; Brawner, C.A.; American College of Sports Medicine. *ACSM's Resource Manual for Guidelines for Exercise Testing and Prescription*; Wolters Kluwer Health/Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2014.
91. Taborri, J.; Rossi, S.; Palermo, E.; Patanè, F.; Cappa, P. A Novel HMM Distributed Classifier for the Detection of Gait Phases by Means of a Wearable Inertial Sensor Network. *Sensors* **2014**, *14*, 16212–16234. [[CrossRef](#)] [[PubMed](#)]
92. Sabatini, A.; Martelloni, C.; Scapellato, S.; Cavallo, F. Assessment of Walking Features From Foot Inertial Sensing. *IEEE Trans. Biomed. Eng.* **2005**, *52*, 486–494. [[CrossRef](#)] [[PubMed](#)]
93. Kotiadis, D.; Hermens, H.; Veltink, P. Inertial Gait Phase Detection for control of a drop foot stimulator. *Med. Eng. Phys.* **2010**, *32*, 287–297. [[CrossRef](#)]
94. Bao, L.; Intille, S.S. Activity recognition from user-annotated acceleration data. In Proceedings of the International Conference on Pervasive Computing, Vienna, Austria, 21–23 April 2004; Springer: Berlin/Heidelberg, Germany, 2004.
95. Pirttikangas, S.; Fujinami, K.; Nakajima, T. Feature selection and activity recognition from wearable sensors. In *International Symposium on Ubiquitous Computing Systems*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 516–527.
96. Reaz, M.B.I.; Hussain, M.S.; Mohd-Yasin, F. Techniques of EMG signal analysis: Detection, processing, classification and applications. *Biol. Proced. Online* **2006**, *8*, 11–35. [[CrossRef](#)] [[PubMed](#)]
97. Wojtys, E.M.; Wylie, B.B.; Huston, L.J. The effects of muscle fatigue on neuromuscular function and anterior tibial translation in healthy knees. *Am. J. Sport. Med.* **1996**, *24*, 615–621. [[CrossRef](#)] [[PubMed](#)]
98. Kern, N.; Schiele, B.; Schmidt, A. Multi-sensor activity context detection for wearable computing. In *European Symposium on Ambient Intelligence*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 220–232.
99. Marras, W.S.; Lavender, S.A.; Leurgans, S.E.; Rajulu, S.L.; Allread, S.W.G.; Fathallah, F.A.; Ferguson, S.A. The role of dynamic three-dimensional trunk motion in occupationally-related. *Spine* **1993**, *18*, 617–628. [[CrossRef](#)] [[PubMed](#)]
100. Huynh, T.; Schiele, B. Analyzing features for activity recognition. In Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence: Innovative Context-Aware Services: Usages and Technologies, Grenoble, France, 12–14 October 2005; pp. 159–163.
101. Heinz, E.A.; Kunze, K.S.; Sulistyo, S.; Junker, H.; Lukowicz, P.; Tröster, G. Experimental evaluation of variations in primary features used for accelerometric context recognition In *European Symposium on Ambient Intelligence*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 252–263.
102. Krause, A.; Siewiorek, D.P.; Smailagic, A.; Farringdon, J. Unsupervised, Dynamic Identification of Physiological and Activity Context in Wearable Computing. *ISWC* **2003**, *3*, 88.
103. Lee, S.W.; Mase, K. Activity and location recognition using wearable sensors. *IEEE Pervasive Comput.* **2002**, *1*, 24–32.
104. Lessley, D.; Crandall, J.; Shaw, G.; Kent, R.; Funk, J. *A Normalization Technique for Developing Corridors from Individual Subject Responses*; Technical Report; SAE Technical Paper: Detroit, MI, USA, 2004.
105. Moorhouse, K. An improved normalization methodology for developing mean human response curves. In Proceedings of the International Technical Conference on the Enhanced Safety of Vehicles, Seoul, Korea, 27–30 May 2013.
106. Yoganandan, N.; Arun, M.W.; Pintar, F.A. Normalizing and scaling of data to derive human response corridors from impact tests. *J. Biomech.* **2014**, *47*, 1749–1756. [[CrossRef](#)]
107. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence, IJCAI, Montreal, QC, Canada, 20–25 August 1995; Volume 14, pp. 1137–1145.
108. Jain, A.K.; Duin, R.P.W.; Mao, J. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 4–37. [[CrossRef](#)]
109. Liu, H.; Cocea, M. Semi-random partitioning of data into training and test sets in granular computing context. *Granul. Comput.* **2017**, *2*, 357–386. [[CrossRef](#)]
110. Browne, M.W. Cross-validation methods. *J. Math. Psychol.* **2000**, *44*, 108–132. [[CrossRef](#)]
111. Dag, A.; Topuz, K.; Oztekin, A.; Bulur, S.; Megahed, F.M. A probabilistic data-driven framework for scoring the preoperative recipient-donor heart transplant survival. *Decis. Support Syst.* **2016**, *86*, 1–12. [[CrossRef](#)]
112. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
113. Han, J.; Pei, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Elsevier: Amsterdam, The Netherlands, 2011.

114. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 112.
115. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 26.
116. Fernández, A.; García, S.; Galar, M.; Prati, R.C.; Krawczyk, B.; Herrera, F. *Learning from Imbalanced Data Sets*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 10.
117. Krawczyk, B. Learning from imbalanced data: Open challenges and future directions. *Prog. Artif. Intell.* **2016**, *5*, 221–232. [[CrossRef](#)]
118. Skiena, S.S. *The Data Science Design Manual*; Springer: Berlin/Heidelberg, Germany, 2017.
119. Khalid, S.; Khalil, T.; Nasreen, S. A survey of feature selection and feature extraction techniques in machine learning. In Proceedings of the 2014 Science and Information Conference, London, UK, 27–29 August 2014; pp. 372–378.
120. Ravi, N.; Dandekar, N.; Mysore, P.; Littman, M.L. *Activity Recognition from Accelerometer Data*; AAAI: Pittsburgh, PA, USA, 2005; Volume 5, pp. 1541–1546.
121. Casas, J.; Irfan, B.; Senft, E.; Gutiérrez, L.; Rincon-Roncancio, M.; Munera, M.; Belpaeme, T.; Cifuentes, C.A. Social Assistive Robot for Cardiac Rehabilitation: A Pilot Study with Patients with Angioplasty. In Proceedings of the Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, HRI'18, Chicago, IL, USA, 5–8 March 2018; pp. 79–80.