



# A machine learning approach for detecting fatigue during repetitive physical tasks

Guobin Liu<sup>1</sup> · Chelsea Dobbins<sup>1</sup> · Matthew D'Souza<sup>1</sup> · Ngoc Phuong<sup>2</sup>

Received: 16 February 2022 / Accepted: 18 March 2023 / Published online: 28 March 2023  
© The Author(s) 2023

## Abstract

Prolonged and repetitive stress on muscles, tendons, ligaments, and nerves can have long-term adverse effects on the human body. This can be exasperated while working if the environment and nature of the tasks puts significant strain on the body, which may lead to work-related musculoskeletal disorders (WMSDs). Workers with WMSDs can experience generalized pain, loss of muscle strength, and loss of ability to continue working. Most WMSDs injuries are caused by ergonomic risks, such as repetitive physical movements, awkward postures, inadequate recovery time, and muscular stress. Fatigue can be seen as a detector of ergonomic risk, as the accumulation of fatigue can significantly increase the possibility of injury. Thirty participants completed a series of repetitive physical tasks over a six-hour period while wearing sensors to capture data related to heart rate and movement, while external embedded sensors captured ground reaction and hand exertion force. They also provided subjective ratings of fatigue at the start and end of the experiment. Classifiers for fatigue (high vs low) were constructed using three methods: linear discriminant analysis (LDA), k-nearest neighbor (kNN), and polynomial kernel-based SVM (P-SVM) and were validated using a tenfold cross-validation technique that was repeated a hundred times. Results of our supervised machine learning approach demonstrated a maximum accuracy of 94.15% using P-SVM for the binary classification of fatigue.

**Keywords** Pervasive computing · Ergonomics · Sensors · Classification · Machine learning · Fatigue

## 1 Introduction

Musculoskeletal disorders (MSDs) occur due to injuries to muscles, tendons, ligaments, nerves, cartilage, and blood vessels [1]. Work-related musculoskeletal disorders (WMSDs) occur due to the working environment and physical tasks that are required to be performed [1].

WMSDs can involve a wide range of disorders and symptoms, including local or generalized pain, loss or hypersensitivity of sensation to touch, heat or pressure, loss of muscle strength, endurance, and/or flexibility, loss of ability to control movements, physical changes to muscle tone, and abnormal alignment of joints [1]. Compared to acute injuries, the mechanism of injury from WMSDs is based on the accumulation of microdamage, which can be caused by different ergonomic risks, from both workers and the workplace, including excessive force, repetition, awkward/static postures, compression, vibration, cognitive demands, and low temperature [1, 2]. Certain industries are more susceptible to their workers being more at risk of developing WMSDs, including heavy and light manufacturing, mining, and food processing [3]. Physical ergonomic features of the work undertaken in such industries can contribute to the increased risk of developing WMSDs, including rapid work pace and repetitive motions, forceful exertions, non-neutral body postures, and vibrations [3]. WMSDs signify a large public health issue and economic burden to employers and workers, which affects health and safety, efficiency, and

---

✉ Chelsea Dobbins  
c.m.dobbins@uq.edu.au

Guobin Liu  
guobin.liu@uq.net.au

Matthew D'Souza  
m.dsouza@uq.edu.au

Ngoc Phuong  
ngoc.h.phuong@boeing.com

<sup>1</sup> School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD, Australia

<sup>2</sup> Boeing Research & Technology, Brisbane, QLD, Australia

productivity [4]. In Australia, WMSDs continue to be the most serious occupational health and safety problem that affected 6.9 million people in 2014–2015, with total costs of more than \$24 billion AUD [1]. Furthermore, the conservative estimation of costs from WMSDs in the US are between \$45 and \$54 billion USD annually [5].

To minimize the probability of suffering WMSDs, it is essential to monitor and analyze ergonomic risks. Once detected, corresponding actions can be performed, such as calibrating the height of working surfaces, extending breaks between tasks, and decreasing the distance between tools and workers [2]. However, monitoring ergonomic risks is challenging due to the variability of the tasks and the way in which they are performed by individuals. Nevertheless, fatigue is seen as a reasonable indicator towards detecting WMSDs, as maintaining a high level of fatigue, for an extended period of time, has been shown to contribute to WMSDs [1]. This paper illustrates a machine learning approach to detect fatigue based on multimodal data from wearable and embedded (non-wearable) sensors. An experiment has been undertaken with thirty participants, who each performed four repetitive physical tasks that have been designed to simulate daily factory work with the purpose of inducing fatigue. Streams of multimodal human-centric data from wearable motion inertial measurement units (IMUs) and a heart rate sensor, as well as external force, have been collected. Self-reported levels of fatigue were also collected at the start and end of the experiment, via questionnaires.

This work provides a number of contributions to the area of personal and ubiquitous computing by developing a data processing pipeline that amalgamates multiple streams of multimodal data from wearable, physiological and external embedded sensors to detect fatigue. This includes:

1. Developing a new signal processing pipeline for multimodal data fusion that is capable of synchronizing, integrating, and processing multiple data streams. This includes recording a combination of human-centric and external measures, which are sensitive to artefacts, and amalgamating these distinct data streams to a common time basis.
2. Implementing a personalized data labelling approach that takes into account individual differences about feelings towards fatigue.
3. Providing a comparison of performance between multiple machine learning algorithms to determine the most suitable approach for detecting fatigue when undertaking repetitive tasks.
4. Providing a comparison of performance to ascertain the best type of sensors to utilize with different models. This is important to understand how different sensors and hardware contribute to the classification of fatigue.

As such, the work addresses the technical challenges that are involved in processing multiple streams of distinct data related to both human behavior and external factors in order to detect fatigue when participants are undertaking repetitive tasks. This has been demonstrated through the analysis of raw data using our data processing pipeline, which involves data synchronization, segmentation, resampling, filtering, feature extraction, and body fatigue classification. A personalized data labelling approach has been implemented to classify fatigue by taking into consideration individual differences towards fatigue. In addition, we demonstrate and evaluate the most appropriate type of sensors and machine learning classification algorithms for identifying high and low periods of fatigue.

The paper is organized as follows. Related work is discussed in Sect. 2. Section 3 demonstrates the methodology, including the experiment details and data pre-processing procedures. Sections 4 and 5 demonstrate the data analysis procedures and the classification results from machine learning models with discussions. Finally, conclusions and future work are shown in Sect. 6.

## 2 Related work

This section provides an overview of related work in the area of sensor-based ergonomic monitoring systems. Current research into ergonomic analysis uses either wearable sensors to capture motion and/or physiological data or embedded sensor systems, to record external force that is being exerted on the human body.

### 2.1 Wearable sensor systems

Wearable sensors to measure physical activity involve inertial measurement units (IMUs) and physiological sensors [6]. IMUs measure acceleration and angular velocity, while body part orientation can be deduced by mounting them to the human body. Valero et al. [7] used eight IMUs attached to the upper back, lower back, upper legs, arms, and lower legs to recognize back, shoulder, and knee activities. Scores have been allocated to each joint based on thresholds. A posture score can then be calculated as the weighted sum of joint scores. Results illustrated that risky postures can be detected by the system from their high posture scores. Rather than analyzing a task by joint angle thresholds, other researchers record motion data and pass them into ergonomic assessment tools to rate an activity. Vignais et al. [8] mounted six IMUs and two goniometers on participant's upper body in order to perform a RULA assessment, which rates a task by the upper body posture and external load [9]. A RULA score was then generated successfully through sensor readings.

Physiological sensors have also been widely used for performing ergonomic analysis. A wearable sensor system has been proposed by Peppoloni et al. [10] that incorporated IMU and electromyography (EMG) sensors for the purpose of reconstructing upper limb motion and measuring hand exertion force against an object. Both a RULA score and Strain index were calculated based on this data. In addition, a wearable biosensor system has been implemented by Jebelli et al. [11] that measured photoplethysmogram (PPG) signal (to derive heart rate), electrodermal activity (EDA), and skin temperature. Features have been extracted from these physiological parameters and used to estimate workers' physical demand, which has been labelled by an energy-expenditure prediction program. Results demonstrated that the accuracies for binary and three-level physical demand classifications are 90% and 87%, respectively. Furthermore, four IMUs and one heart rate monitor have been applied by Maman et al. [12] to measure body fatigue during manual material handling tasks. Features from body motion and heart rate act as the input and a Borg score [13] recorded from participants' perception acts as the output. A LASSO penalized regression model has been constructed to predict body fatigue every 10 min. The mean absolute deviation between recorded and predicted scores was 2.16, meaning that the result is accurate in the case that the recorded score is highly variable.

## 2.2 Embedded sensor systems

External force applied to the human body is one of the most common causes for WMSDs [2]. For instance, a plantar pressure measurement approach has been proposed by Chen et al. [14]. The system converts pressure data to a grayscale image, which is then passed into a convolutional neural network with long- and short-term memory to recognize postures. The recognition accuracy for five quasi-static postures was 100%, as there is a clear difference between plantar force distribution. As for sequential motions, the classification accuracies were 99%, 74%, 79%, and 92% for pushing/pulling, bending, carrying, and lifting, respectively. Furthermore, the number of repetitions for each activity were counted with an accuracy of over 80%. Furthermore, Chow and Dickerson [15] analyzed the factors that can affect manual force strength during two-handed standing, with maximal horizontal pushing and pulling, including shoulder capacity, handle height, exertion type, and handle orientation. A load cell and VICON motion capture system were included in this study. The load cell was mounted to a metal pole to measure hand exertion force, while the VICON was responsible for tracking body postures. Results indicated that handle height and exertion type are two major factors that can affect manual force strength during push/pull tasks. In addition, a force plate and two sets of motion

capture systems were used by Plamondon et al. [16] to rank the importance of expertise, lifting height, and load weight in affecting the lifting tasks. Lumbar spine and knees were selected as the major joints to reveal the effect from these three factors. By constructing a statistical analysis model, results illustrated that lifting height and load weight are more important than expertise in influencing lumbar spine and knees. In addition, results also illustrated that the lower and heavier the box, the riskier the lifting task is.

In summary, both wearable and embedded sensors have separately been applied in the area of ergonomic analysis. Wearable sensors measure body motion and physiological data, while embedded sensors are mainly responsible for measuring the external force exerted on the human body. Such data can be processed to predict indirect parameters, identify ergonomic risks, calculate ergonomic scores, and detect body fatigue. However, little work incorporates data from both wearable and embedded sensors to detect fatigue. This is important because it can be postulated that fatigue detection can be more accurate by including not only human-centric parameters but also external factors, such as external force.

## 3 Methodology

We propose a machine learning approach for detecting fatigue during aviation manufacturing tasks using a combination of wearable and embedded sensors. Manufacturing is one of the industries that has the highest rate of WMSDs and so provides an ideal use case for our scenarios [3]. The experiment has been designed based on guidelines developed by the National Institute for Occupational Safety and Health [17] and through consultation with industry partners. In order to test our approach, four physical tasks have been designed to simulate daily aviation factory work, including two-handed box lifting, two-handed box carrying, two-handed box pushing/pulling, and two-handed trolley pushing/pulling. Data have been recorded from wearable and embedded sensors while participants performed the tasks, including body motion, external force, and heart rate. The experiment setup and conditions are standard and generalizable to ensure repeatability. Participant details, sensor platform, experiment protocol, and the data pre-processing approach will be discussed in this section.

### 3.1 Participants

The experiment included 30 participants (13 males and 17 females), with an age range from 20 to 54 years (mean = 25.67, SD = 6.66). Participants did not have a history of cardiovascular illness or any physical injuries prior to participating in the experiment. This experiment has

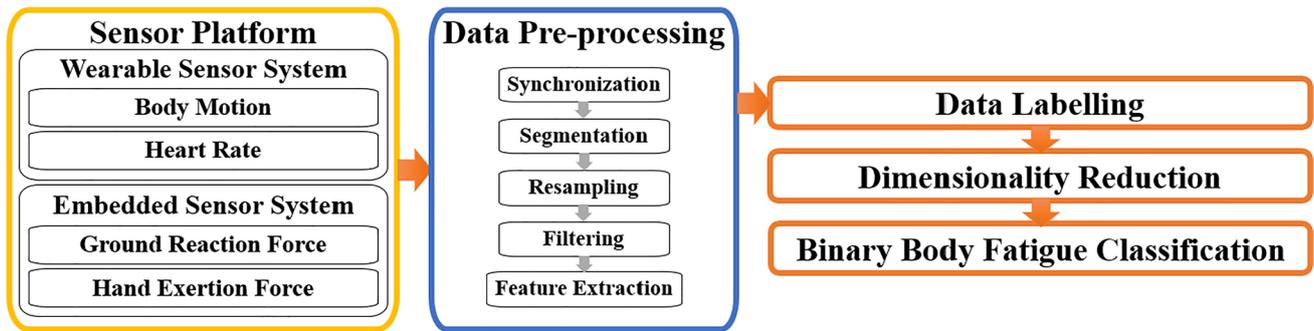
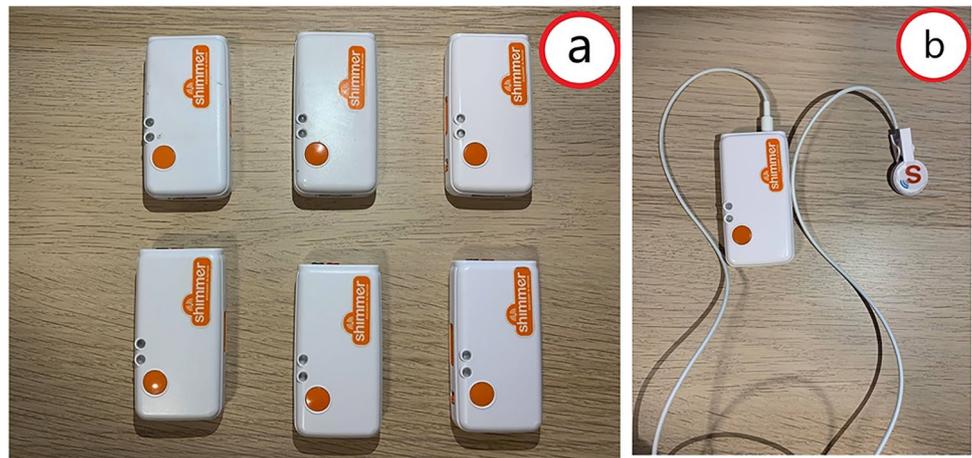


Fig. 1 Sensor platform overview

Fig. 2 a Shimmer3™ GSR + unit. b Shimmer3™ PPG-to-HR ear clip



been approved by The University of Queensland Ethics Sub-Committee.

### 3.2 Sensor platform

The platform consisted of both wearable and embedded sensors (see Fig. 1). The wearable sensor system is responsible for recording body motion and heart rate, while the embedded aspect measures ground reaction and hand exertion force. During the pre-processing stage, data have been synchronized, segmented, resampled, and filtered, and features have been extracted before undertaking the binary classification of fatigue.

The wearable aspect of the sensor platform was composed of six Shimmer3™<sup>1</sup> inertial measurement units (IMUs) (see Fig. 2a) and a Shimmer3™ photoplethysmogram (PPG) optical pulse ear clip (see Fig. 2b). Shimmer3™ sensors were selected as the experiment required raw biometric data to be collected from different areas of the body in order to develop the signal processing pipeline and model classification.

In order to capture major body motion [7, 12, 18, 19], the Shimmer3™ IMU sensors were secured to six different areas of the body, including the hip, upper spine, left arm, right arm, right forearm, and right shank (see Fig. 3). These areas were chosen in order to provide maximum coverage of the entire body. Each Shimmer3™ IMU sensor measured 3-axis acceleration and 3-axis angular velocity. In addition, the Shimmer3™ PPG-to-HR ear clip was clipped to the left ear lobe and was connected to the Shimmer3™ sensor on the left arm to measure blood volume pulse (BVP), which is used to derive heart rate (see Fig. 3a).

Based on previous work in activity recognition and ergonomic analysis [20–22], 100 Hz has been selected as the sampling frequency to ensure continuous and sufficient data collection. Data was stored on the onboard SD card of each sensor and then exported to a PC after each task.

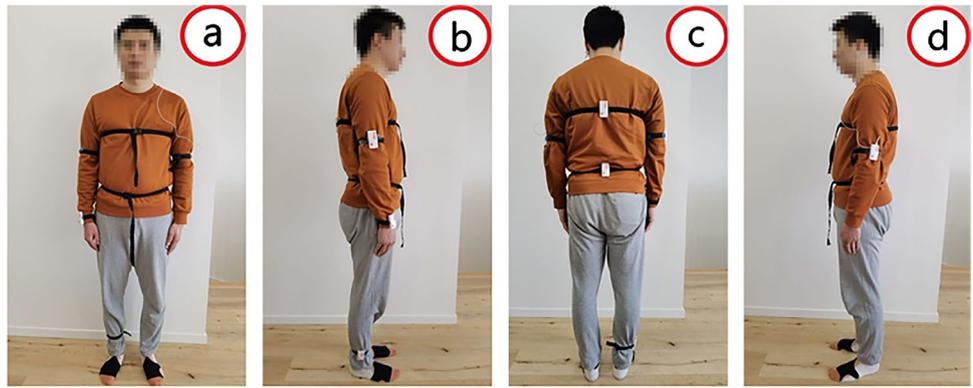
The embedded sensors used were a Sparkfun™<sup>2</sup> load cell (see Fig. 4a) and a Bertec™<sup>3</sup> force plate (see Fig. 4b). These sensors measured contact forces around the human

<sup>1</sup> [www.shimmersensing.com](http://www.shimmersensing.com)

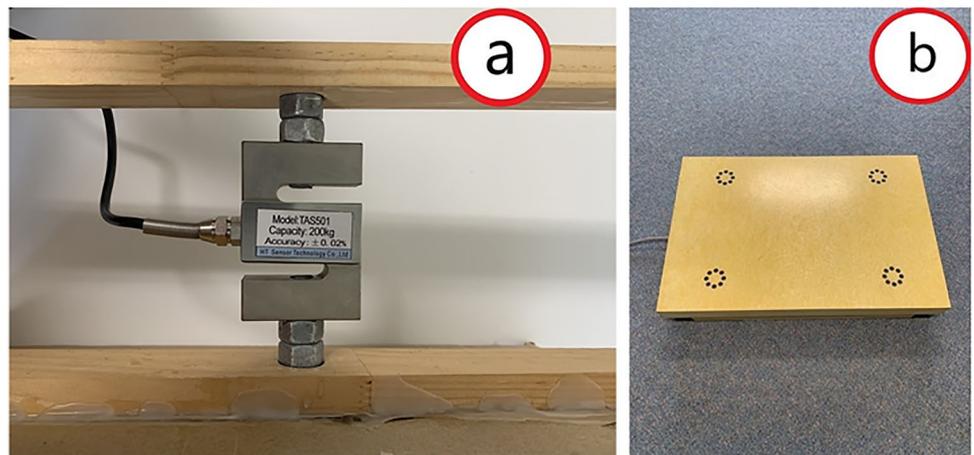
<sup>2</sup> <https://www.sparkfun.com/products/14282>

<sup>3</sup> <https://www.bertec.com/products/force-plates>

**Fig. 3** Wearable sensor placement: **a** front view; **b** right view; **c** back view; **d** left view



**Fig. 4** **a** Load cell–200 kg, S-type (TAS501). **b** Force plate FP4060-NC



body, which cannot be captured by wearable sensors. The load cell's handle was attached to a box, which measured the hand's exertion force against the box. The force plate measured 3-axis ground reaction force, 3-axis torso moment, 2-axis center of pressure, and 2-axis center of gravity. The sampling rates for the load cell and the force plate were set up to 10 Hz and 100 Hz, respectively, where 10 Hz is the maximum frequency of the load cell. Both the load cell and force plate streamed data directly to a PC through a wired connection.

The Shimmer3™ sensors and load cell were utilized in all tasks, while the force plate was not included in tasks that required participants to move off the plate (i.e., two-handed trolley pushing/pulling). This is because the force plate was not able to measure the ground reaction force while the participant was walking.

### 3.3 Experiment protocol

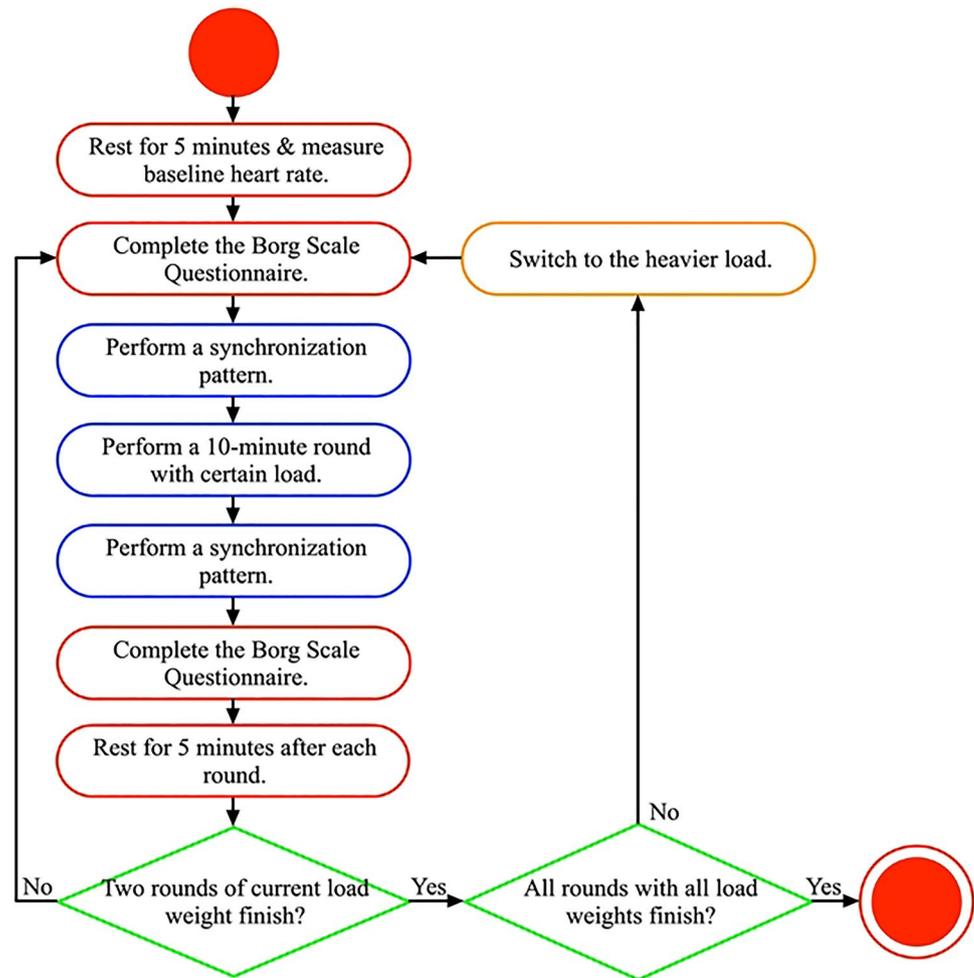
The experiment consisted of four repetitive physical tasks that have been designed to simulate real factory work and included two-handed box lifting, two-handed box carrying, two-handed box pushing/pulling, and two-handed trolley

pushing/pulling. A description of each task has been provided as follows.

1. *Two-handed box lifting*: Participants lift a box from the ground to their hip height with two hands, hold on for approximately 3 s, and then lower it down to the ground.
2. *Two-handed box carrying*: Participants carry a box with both hands and move it from one chair to another. Two chairs are fixed at either side of the participant, at a distance of 60 cm from the centerline of the body.
3. *Two-handed box pushing/pulling*: Participants use both hands to push a box further away from them to a length of approximately 60 cm and pull it back. The box is placed on a table at a height of 90 cm. There is a handle attached to the box for pushing/pulling.
4. *Two-handed trolley pushing/pulling*: Participants walk and push a trolley away for 5 m and pull it back to the original position. The handle height for the trolley is 80 cm.

Figure 5 illustrates the procedure that has been followed by the participants while they performed each task. Before each task began, baseline heart rate was recorded for 5 min,

Fig. 5 Task procedure



while they rested. Participants then performed two rounds of each task utilizing light to heavy weights and also completed the Borg Scale Questionnaire [13], at the beginning/end of each round, which recorded their subjective rating of body fatigue. The load weights differed according to the task and included:

1. Two-handed box lifting: 3 kg and 5 kg.
2. Two-handed box carrying: 3 kg and 5 kg.
3. Two-handed box pushing/pulling: 3 kg, 10 kg, and 20 kg.
4. Two-handed trolley pushing/pulling: 3 kg, 40 kg, and 80 kg.

As the recorded data is from three independent data sources, it is essential to synchronize them to a common timeline and to avoid time drift, which is a significant issue as the sampling rates differ between devices. As such, a synchronization pattern was designed whereby during the pre-processing stage, the pattern was detected, and data aligned accordingly. As such, each participant performed

the same pattern at the beginning and end of each round, which included three steps:

1. Stand still and upright at origin for 10 s.
2. Perform start signal by raising the body up slowly to stand on tiptoes and plunge downwards rapidly.
3. Stand still and upright at origin for 10 s.

Each round lasted for 10 min, and participants were required to do the task repetitively during this period. During each round, participants were allowed to control their posture and speed while performing the task. They were also permitted to stand upright at origin to rest if they could not sustain the task. A 5-min resting period after each round and a 10-min resting period after each task were also provided. Since lifting and carrying tasks can pose continuous pressure on the waist, the sequence of tasks was interwoven to avoid excessive and durable pressure on specific joints, so as not to cause injury. All participants followed this sequence during the experiment—two-handed trolley pushing/pulling, two-handed

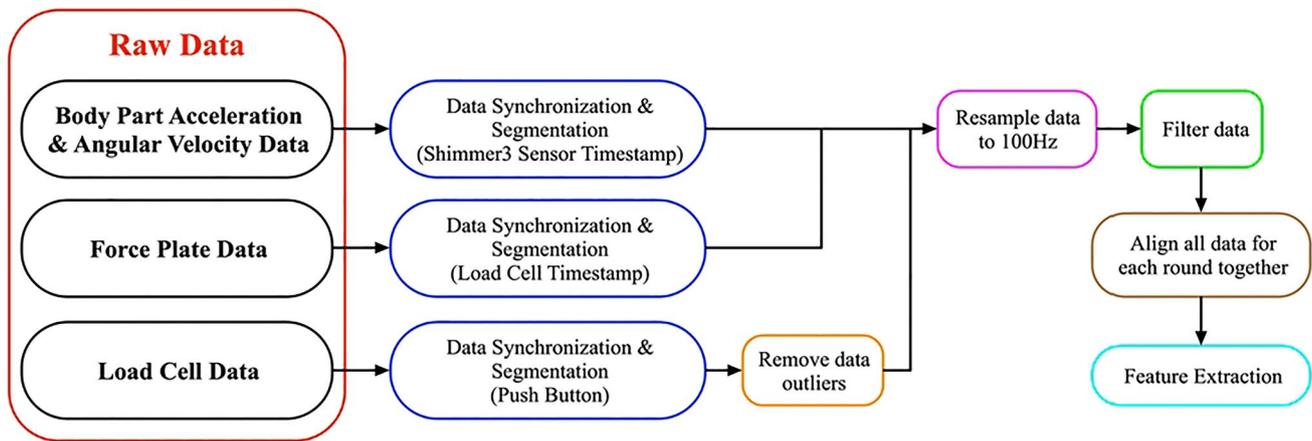
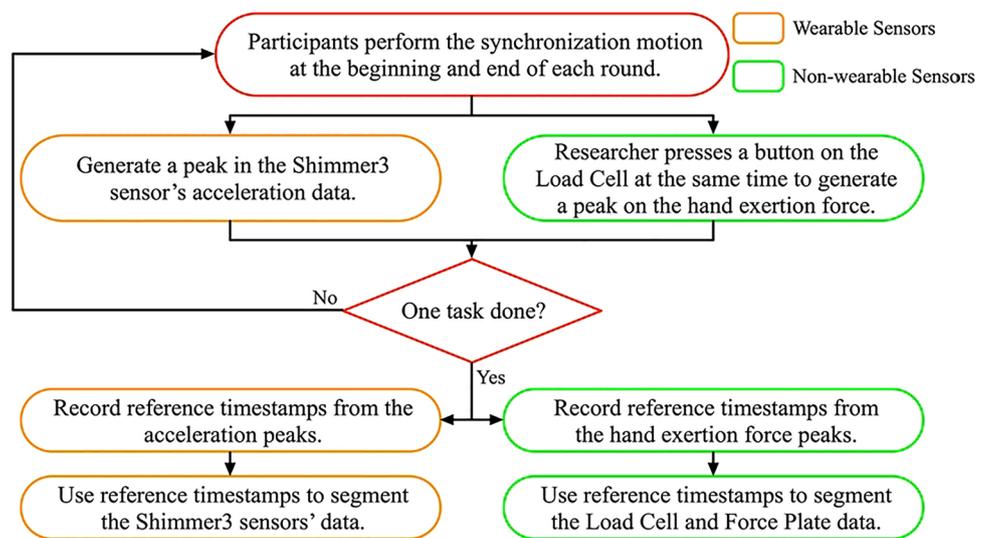


Fig. 6 Data pre-processing procedure

Fig. 7 Data synchronization and segmentation procedure



box carrying, two-handed box pushing/pulling, and two-handed box lifting.

The entire experiment lasted for 6 h. Lifting and carrying tasks took approximately 60 min each to complete, while pushing/pulling tasks took around 90 min each to complete. As an incentive, after completing the experiment, participants were paid with a \$50 (AUD) gift card to compensate their time for taking part in the experiment.

### 3.4 Data pre-processing

The pre-processing stage included a number of steps to prepare the data before it was classified. Figure 6 illustrates an overview of this process, which included data synchronization, segmentation, resampling, filtering, and feature extraction. Details for each step will be expanded upon in the following sections. This stage has been undertaken using MATLAB R2019b [23].

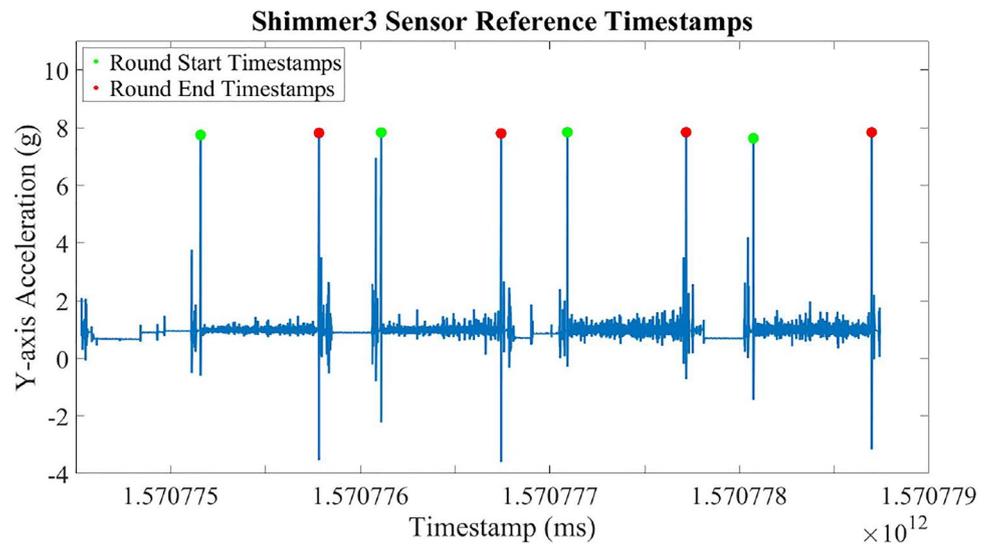
#### 3.4.1 Data synchronization and segmentation

As mentioned in Sect. 3.3, a synchronization pattern was performed by the participants at the beginning and end of each round, which generates a peak in the acceleration data. Figure 7 illustrates a flow chart for this stage.

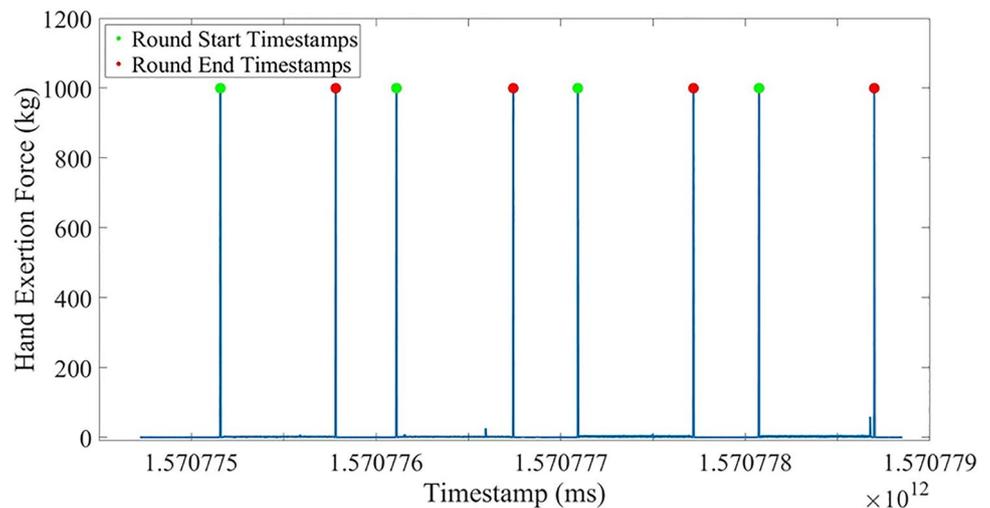
As all the Shimmer3™ sensors have a consistent timestamp, data from each of them has been segmented through the reference timestamps at the points of the acceleration peaks. Figure 8 displays an example of the y-axis acceleration of the Shimmer3™ sensor. It can be seen that there are acceleration peaks at the beginning and end of each round, which have been labelled by the green and red circles respectively. Corresponding timestamps have then been extracted as the reference timestamps to segment data from all Shimmer3™ sensors.

During the synchronization pattern, at the point when the participants performed the start signal, the researcher

**Fig. 8** Example of Shimmer3 sensor reference timestamps



**Fig. 9** Load cell reference timestamps



pressed a button on the load cell, which generated a peak in the hand exertion force data. Since the load cell and force plate's data were recorded using the same PC, the timestamps at the hand exertion force peaks have been used to segment the embedded sensors' data. Figure 9 depicts an example for extracting reference timestamps from the load cell data. There are peaks at the beginning and end of each round, which have been labelled by green and red circles, respectively. The timestamps were then used to segment the data from the force plate and load cell.

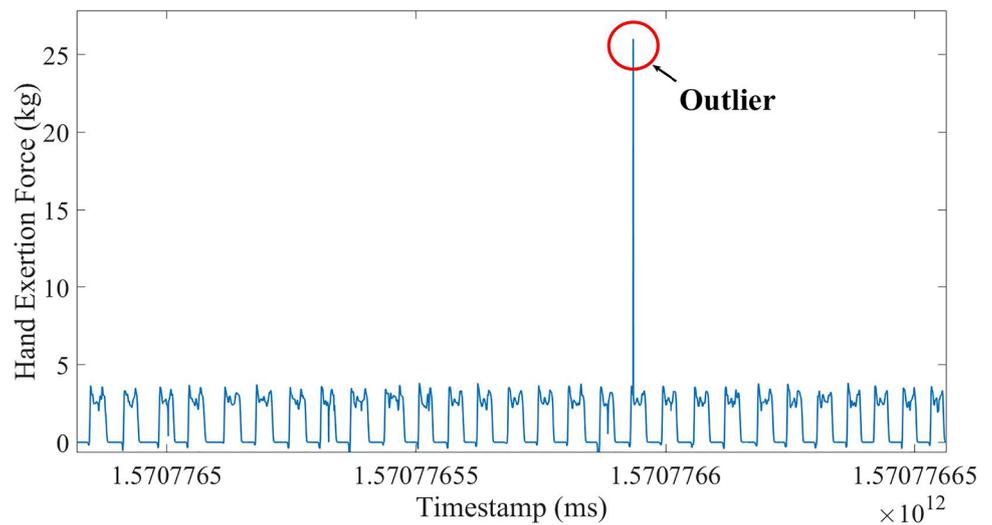
### 3.4.2 Data resampling and filtering

Once the data was synchronized to a common timeline, resampling the data was required as different sensors record data at different sampling rates. As such, the Shimmer3™ sensors' and force plate's data have been resampled to 100 Hz [20–22]. Acceleration data from the Shimmer3™ sensors

and the data from the force plate have then been filtered through a 2nd order zero-phase Butterworth low-pass filter with a cutoff frequency of 3 Hz [24–26]. Angular velocity data was filtered through a 2nd order zero-phase Butterworth high-pass filter, with a cutoff frequency of 0.1 Hz [27]. Furthermore, the PPG signal was filtered through a 2nd order zero-phase Butterworth band-pass filter, with a cutoff frequency ranging from 0.5 to 3.5 Hz [28, 29].

Additionally, outliers in the load cell data have been identified whereby the hand exertion force is greater than 10 kg (see Fig. 10). As these outliers cannot be removed by filtering alone, a sliding window algorithm, with a window size equaling 20 samples, was developed based on experimentation to remove any outliers. For each window, if the sample's absolute value was greater than two times the median absolute deviation, this was treated as an outlier and replaced by a null value. Finally, the null values were interpolated by their nearest valid sample. After the outliers were removed,

Fig. 10 Load cell outliers



the load cell’s data was resampled to 100 Hz (to be consistent with the other data) and filtered using a 2nd order zero-phase Butterworth low-pass filter, with a cutoff frequency of 3 Hz [30, 31].

Physiological data was also recorded using the PPG sensor, which measured blood volume change inside the capillary vessels as blood flows through the vessels at each heartbeat. Therefore, the filtered PPG signal was utilized to derive heart rate in beats per minute [32, 33]. This was calculated using the time interval between adjacent peaks (see Fig. 11 and Eq. (1)). In Fig. 11, each red circle indicates a peak in the PPG signal, with a corresponding timestamp in milliseconds. Therefore, the seconds per beat have been calculated by the time difference between adjacent peaks, and then converted to beats per minute using Eq. (1).

$$Heart\ rate( beats\ per\ minute) = \frac{60}{\Delta t_{peak-peak}(seconds)} \quad (1)$$

### 3.4.3 Feature extraction

Features have then been extracted from each round of synchronized data, using a 30-s window, with a 50% overlapping area, and included both time- and frequency-domain features [11, 34]. This period was selected to ensure a sufficient number of features were generated in the feature set in order to perform fatigue classification. Figure 12 illustrates the input data with corresponding features.

Motion data was recorded from each IMU, including from each axis ( $x$ ,  $y$ , and  $z$ ), as well as magnitude of body part acceleration and angular velocity. In addition, acceleration jerk and angular acceleration were also calculated, which are the first-order time derivative of acceleration and angular velocity (see Eqs. (2) and (3)) [12, 35]. In Eq. (2),  $\vec{a}(t)$

represents the acceleration and  $\vec{j}(t)$  is the first order time derivative of the acceleration, which is the acceleration jerk. In Eq. (3),  $\vec{\omega}(t)$  indicates the angular velocity and  $\vec{\alpha}(t)$  is the first-order time derivative of the angular velocity, which is named as angular acceleration.

$$Acceleration\ jerk\ \vec{j}(t) = \frac{d\vec{a}(t)}{dt} \quad (2)$$

$$Angular\ acceleration\ \vec{\alpha}(t) = \frac{d\vec{\omega}(t)}{dt} \quad (3)$$

External force included the force plate and load cell data, with each axis of ground reaction force, torso moment, center of pressure, center of gravity, and hand exertion force. Heart rate was also included in the time-domain feature extraction process. Time-domain features included 10th, 25th, 50th, 75th, 90th percentiles, mean, standard deviation, inter-quartile range, kurtosis, mean absolute deviation, and autocorrelation. In total, 1056 and 11 features were extracted

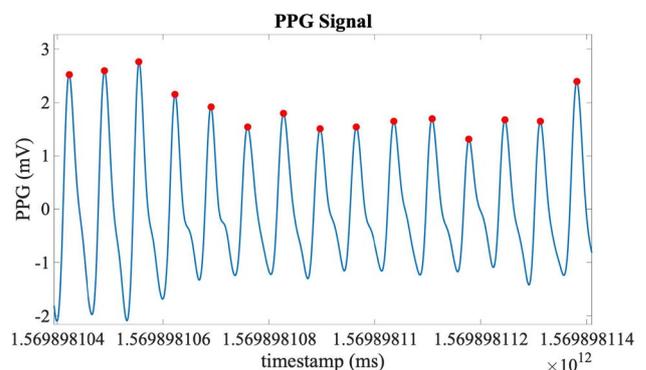


Fig. 11 Filtered PPG signal with detected peaks

Fig. 12 Feature set

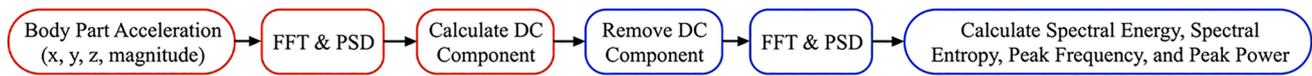
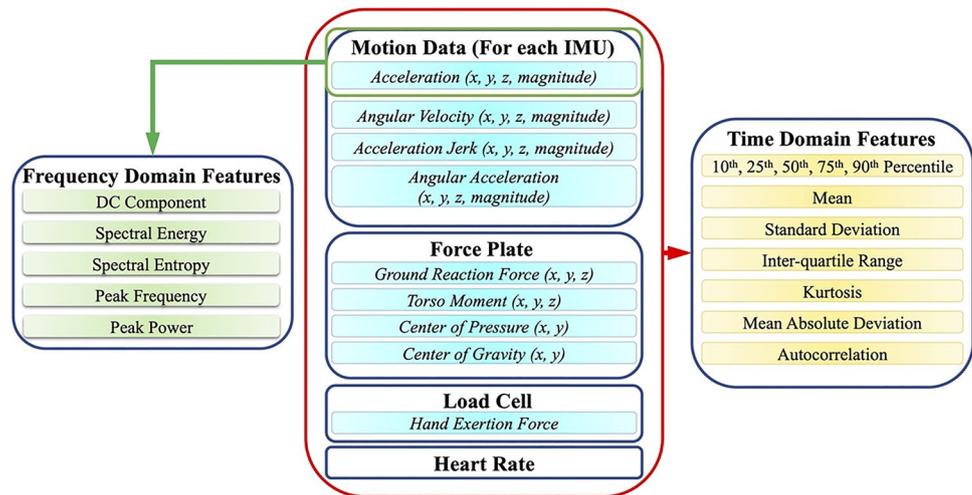


Fig. 13 Frequency-domain feature extraction procedure

from the body motion and heart rate data, respectively. For external force, 11 features were extracted from the two-handed trolley pushing/pulling tasks, as the force plate was not included, while 121 external force features were extracted for each of the other tasks.

Acceleration data from each IMU was also used in the frequency-domain feature extraction stage. Frequency domain features included the DC component, spectral energy, spectral entropy, peak frequency, and peak power. Figure 13 illustrates the procedure of frequency-domain feature extraction. Acceleration data from each IMU was processed by fast Fourier transform (FFT) to calculate the power spectral density (PSD). The DC component was then calculated at the point where frequency equaled 0 Hz. After that, the DC component was removed by subtracting the acceleration data by their mean values. FFT and PSD have been recalculated to extract spectral energy, spectral entropy, peak frequency, and peak power. Spectral energy is the sum of squares of the frequency components' magnitudes, with its calculation shown in Eq. (4), while spectral entropy measures the spectral power distribution, which is the sum of normalized frequency components multiplied by their logarithm values (see Eq. (5)) [36].

$$S_E = \sum_{i=1}^N F(n)^2 \quad (4)$$

$$S_{EN} = - \sum_{i=1}^N \hat{F}(n) * \log(\hat{F}(n)) \quad (5)$$

$F(n)$  is the magnitude of  $n^{\text{th}}$  FFT frequency component and  $\hat{F}(n)$  is the normalized value of  $F(n)$  (see Eq. (6)).

$$\hat{F}(n) = \frac{F(n)}{\sum_{i=1}^N F(n)} \quad (6)$$

This phase resulted in the extraction of 120 frequency-domain features. Overall, 1198 features were extracted from the two-handed trolley pushing/pulling task, while 1308 features were extracted for each of the other tasks. To mitigate individual differences, features from each round, per participant, were then normalized and combined into one datasheet.

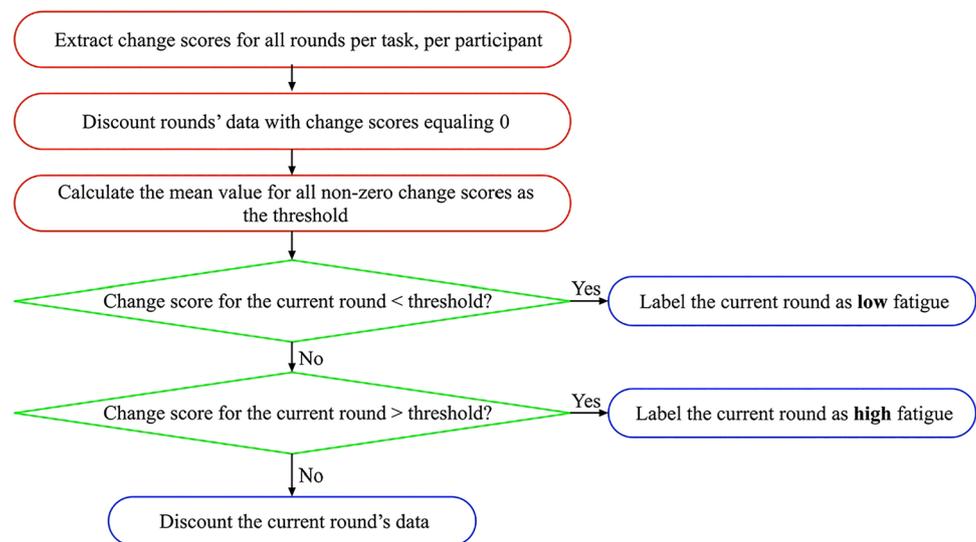
## 4 Data analysis

The data analysis stage was composed of three phases, including data labelling, dimensionality reduction, and body fatigue classification.

### 4.1 Data labelling

The Borg Scale questionnaire [13] was utilized to capture subjective ratings of body fatigue at the beginning (pre) and end (post) of each round. This questionnaire included 15 exertion levels, ranging from 6 to 20, which participants used to rate their current fatigue level. For instance, a rating of 6 corresponds to relaxing tasks, such as reading a book, while 20 indicates exhausting tasks, such as running a race.

**Fig. 14** Data labelling procedure



The scores from this questionnaire were processed to calculate a change score (post-round–pre-round). As such, change scores related to fatigue were used as subjective labels for the data to describe the level of fatigue each participant felt during each round. Figure 14 illustrates the data labelling procedure. Data were labelled as high or low fatigue on a per round, per task, per participant basis. Those rounds where a change score equaled zero were excluded, as a change in the level of fatigue did not occur. A personalized threshold was then calculated for each participant’s round of data based on the calculated mean of all the non-zero change scores for each round. Those rounds whereby (a) the change scores were greater than the threshold have been labeled as *high* fatigue, (b) those that scored lower than the threshold were labeled as *low* fatigue, and (c) those that equaled the threshold were discounted.

The data labelling procedure resulted in an imbalanced dataset (see Table 1), which can decrease model performance, as the model tends to predict the majority class better than the minority class [37]. To rectify this issue, it was necessary to balance the classes before classification. In this instance, random Under-Sampling (RUS) was implemented to randomly remove samples from the majority class (in this instance the *low* class) in order to balance it with the number of samples in the minority class (the *high* class) [38].

### 4.2 Dimensionality reduction

The dataset contained more than 1000 features that were extracted from the various streams of sensor data. However, some of them may be correlated with others and/or irrelevant. When performing fatigue classification, these features may also incorporate noise into the predictor, increase computational cost, and reduce model performance. Therefore, dimensionality reduction was undertaken using principal

component analysis (PCA), with 95% of variance being explained, to minimize the number of features, while maintaining the most discriminant information in the feature set. The number of features before/after PCA for all results can be found in Appendixes 1–3.

### 4.3 Body fatigue classification

The classification analysis involved investigating the types of sensors that are most appropriate for detecting fatigue. As such, the analysis comprised of three main phases:

1. Investigation of the variety of sensors, including separate analysis of embedded, wearable and both wearable + embedded
2. Exploration of the contribution of each individual wearable sensor to the system’s performance
3. Optimization of the wearable sensors to reduce the number of sensors required in the system, which included

As such, this analysis involved creating several individual binary models using parametric and non-parametric supervised machine learning algorithms. Parametric methods aim at summarizing the relationship between labels and features, while non-parametric methods do not make any strong assumptions between them. In other words, it can

**Table 1** Number of samples per class

Class label	Number of samples
High	8355
Low	9629
Removed	5417

**Table 2** Classification performance utilizing only the embedded sensors

Classifier	Accuracy	F1	BER
LDA	45.86%	46.02%	54.14%
kNN	<b>59.08%</b>	60.04%	<b>40.92%</b>
P-SVM	59.05%	<b>62.54%</b>	40.95%

Bold entries denotes the best performing results

learn any form of training data [40]. A mixture of parametric and non-parametric methods have been implemented in the analysis to evaluate their performance and included linear discriminant analysis (LDA) [41], k-nearest neighbor (kNN) with  $k$  equaling 5 [41], and polynomial kernel-based SVM (P-SVM) [42]. Hyperparameters for each method have been tuned using grid search based on the highest accuracy. The tuned hyperparameters for all results can be found in Appendix 4. This analysis has been undertaken using RStudio v1.3.1093 [39].

For each analysis, and to avoid bias in the prediction, the models were trained and validated utilizing repeated tenfold cross-validation, with 100 repetitions [43]. Performance has then been compared within and between datasets based on three parameters, including:

- 1) Accuracy: The ratio of correct predictions over all predictions.
- 2)  $F_1$  score: The measurement of classification accuracy by balancing precision and recall.
- 3) Balance error rate (BER): The average of the errors on each class.

#### 4.3.1 Classification of wearable vs embedded sensors

The purpose of these classification models was to assess the optimum type of sensors to detect fatigue by analyzing wearable vs embedded sensors. Table 2 illustrates that the performance of the models based only on using the embedded sensors. The performances are modestly comparable using P-SVM and kNN, which have produced similar accuracies of 59% and error rates of 41%. LDA performed the worst with a lower accuracy of 46% and a higher error rate of 54%. Additionally, P-SVM and kNN generated similar  $F_1$  scores of 60%, which is approximately 15% higher than the value from LDA. This pattern demonstrates that using only embedded sensors is not sufficient to classify fatigue.

As opposed to the embedded sensors, Table 3 illustrates that using only the wearable sensor dataset increases performance of the P-SVM and kNN algorithms with acceptable accuracies of 93% and 72% and low error rates of 28% and 7%, respectively. Once again, LDA performed the worst with an accuracy of 39% and a high error rate of 61%. The  $F_1$  score result follows the same trend as the accuracy and error rate.

**Table 3** Classification performance utilizing only the wearable sensors

Classifier	Accuracy	F1	BER
LDA	39.04%	39.15%	60.96%
kNN	72.12%	73.35%	27.88%
P-SVM	<b>93.18%</b>	<b>93.22%</b>	<b>6.82%</b>

Bold entries denotes the best performing results

**Table 4** Classification performance utilizing both wearable and embedded sensors

Classifier	Accuracy	$F_1$	BER
LDA	38.54%	38.65%	61.46%
kNN	74.10%	74.50%	25.90%
P-SVM	<b>94.15%</b>	<b>94.18%</b>	<b>5.85%</b>

Bold entries denotes the best performing results

P-SVM produced the highest score of 93%, followed by kNN of 73% and LDA of 39%. This demonstrates that the wearable sensors are better suited to classify fatigue than utilizing only the embedded sensors.

In an effort to further improve the findings, Table 4 illustrates that utilizing both the wearable and embedded sensors marginally improved the results. Once again, P-SVM and kNN performed the best with accuracies of 94% and 74%, high  $F_1$  scores of 94% and 74%, and low error rates of 6% and 26%, respectively. This is a significant improvement over the embedded results and a negligible improvement over the wearable sensors dataset. LDA performed the worst using this dataset, with poor performance of 39% accuracy, a low  $F_1$  score of 38%, and a high error rate of 62%. This pattern demonstrates that a simple linear parametric model is not sufficient to classify fatigue across all datasets. In this instance, fatigue classification seems better suited to nonparametric machine learning algorithms.

#### 4.3.2 Classification of individual wearable sensors

The results in Sect. 4.3.1 demonstrated that using both wearable and embedded sensors or only wearable sensors, exhibited improved performance over using only the embedded sensors. Based on these results, the purpose of this classification was to compare the results from each individual wearable sensor (see Table 5). As the embedded sensors did not produce adequate results on their own, and their inclusion with the wearable sensors only provided a marginal improvement, they have been omitted from further analysis. Table 5 illustrates that P-SVM again performed the best across all sensors. The placement at the right forearm produced the best result of 64%, while heart rate produced the worst at 51%. There was very little variation in the accuracies of the

**Table 5** Classification performance of individual wearable sensors

Sensor position/performance measure		LDA	kNN	P-SVM
Right forearm	Accuracy	44.59%	61.54%	<b>64.28%</b>
	F <sub>1</sub>	45.15%	62.68%	<b>64.39%</b>
	BER	55.41%	38.46%	<b>35.72%</b>
Hip	Accuracy	47.02%	61.40%	<b>63.85%</b>
	F <sub>1</sub>	47.28%	58.88%	<b>63.57%</b>
	BER	52.98%	38.60%	<b>36.15%</b>
Right arm	Accuracy	46.87%	60.76%	<b>63.65%</b>
	F <sub>1</sub>	47.48%	61.93%	<b>63.71%</b>
	BER	53.13%	39.24%	<b>36.35%</b>
Left arm	Accuracy	46.68%	61.72%	<b>63.62%</b>
	F <sub>1</sub>	47.13%	63.22%	<b>63.68%</b>
	BER	53.32%	38.28%	<b>36.38%</b>
Upper spine	Accuracy	46.88%	61.84%	<b>63.24%</b>
	F <sub>1</sub>	47.37%	60.76%	<b>63.13%</b>
	BER	53.12%	38.16%	<b>36.76%</b>
Right shank	Accuracy	43.91%	60.91%	<b>62.96%</b>
	F <sub>1</sub>	44.05%	61.51%	<b>63.09%</b>
	BER	56.09%	39.09%	<b>37.04%</b>
Heart rate	Accuracy	48.34%	50.39%	<b>51.29%</b>
	F <sub>1</sub>	47.26%	50.41%	<b>52.13%</b>
	BER	51.66%	49.61%	<b>48.72%</b>

Bold entries denotes the best performing results

**Table 6** Classification performance of wearable motion sensors

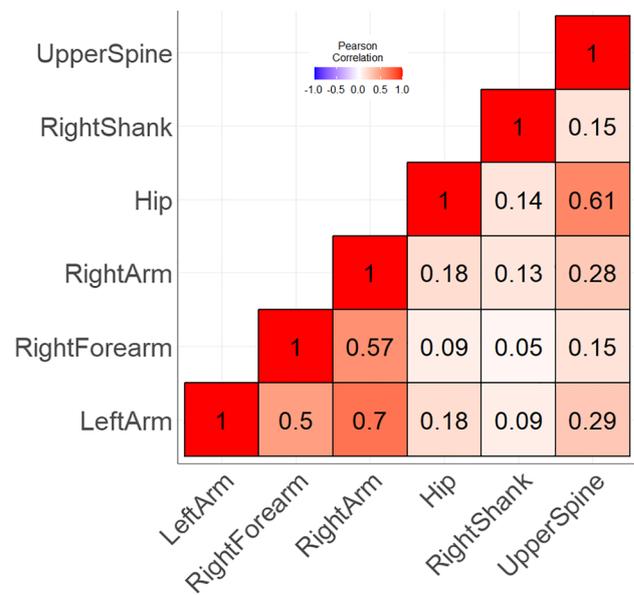
Classifier	Accuracy	F <sub>1</sub>	BER
LDA	39.11%	39.23%	60.89%
kNN	71.64%	72.98%	28.36%
P-SVM	<b>93.17%</b>	<b>93.20%</b>	<b>6.83%</b>

Bold entries denotes the best performing results

sensors around the body that were used to detect movement (omitting heart rate), which ranged from 63 to 64%.

### 4.3.3 Classification of wearable sensors

The results in Sect. 4.3.2 demonstrated that individually the wearable sensors produced sub-optimal results. Based on these results, the purpose of this classification was to optimize the system to minimize the number of wearable sensors used, while maintaining model performance. As the heart rate sensor’s results were barely above chance this sensor has been omitted from further analysis. Table 6 illustrates the performance of the models using only the wearable motion sensors. Once again, the performance of the P-SVM model surpassed the two other models, with the highest accuracy of 93%, and lowest error rate of 7%. The LDA model performed the worst,



**Fig. 15** Correlation of wearable motion sensors

**Table 7** Classification performance of wearable motion sensors without upper spine

Classifier	Accuracy	F <sub>1</sub>	BER
LDA	40.10%	40.22%	59.90%
kNN	71.12%	72.62%	28.88%
P-SVM	<b>90.95%</b>	<b>90.98%</b>	<b>9.05%</b>

Bold entries denotes the best performing results

with the lowest accuracy of 39%, and the highest error rate of 61%. kNN’s performance was intermediate, with a modest accuracy of 72% and error rate of 28%.

Table 6 demonstrated the results of removing the physiological sensor and using only the wearable motion sensors. The next stage of the analysis is to minimize the number of wearable motion sensors. This analysis first included examining the correlation between the sensors and to remove those that were highly correlated (see Fig. 15). Using Pearson’s correlation, Fig. 15 demonstrates that the left arm and right arm (0.7), hip and upper spine (0.61), and right forearm and right arm (0.57) had moderate to high correlations. According to feedback from the participants, the upper spine sensor was extremely uncomfortable to wear during the experiment. As such, this sensor has been removed first, and the model’s performance has been re-evaluated.

Table 7 illustrates the performance of the models with the upper spine sensor removed. Once again, the performance of the P-SVM model surpassed the two other models, with the highest accuracy of 91%, and lowest error rate of 9%. The LDA model performed the worst, with the lowest accuracy of 40%, and the highest error rate of 60%.

**Table 8** Classification performance of wearable motion sensors without upper spine and right arm

Classifier	Accuracy	F <sub>1</sub>	BER
LDA	41.18%	41.30%	58.82%
kNN	70.65%	71.87%	29.35%
P-SVM	<b>88.12%</b>	<b>88.17%</b>	<b>11.88%</b>

Bold entries denotes the best performing results

The other highly correlated sensor was on the right arm and so the next stage of the analysis was to remove this sensor as well and re-evaluate the system's performance. Table 8 illustrates the performance of the models with the upper spine and right arm sensors having been removed. Once again, the performance of the P-SVM model surpassed the two other models, with the highest accuracy of 88%, and lowest error rate of 12%. The LDA model performed the worst, with the lowest accuracy of 41%, and the highest error rate of 59%.

## 5 Discussion

We show the results of binary fatigue classification using a variety of data streams, including body motion, external force, and heart rate. Four repetitive physical tasks (two-handed box lifting, two-handed box carrying, two-handed box pushing/pulling, and two-handed trolley pushing/pulling) were designed to simulate daily factory work to collect human-centric data from wearable and embedded sensors. Data was then pre-processed before features were extracted. In an effort to reduce the number of features in each dataset, PCA was utilized to eliminate redundant features so that each dataset was composed of only the discriminate features. The analysis then involved implementing several models to examine the optimum sensors and algorithms to use for detecting fatigue. This is important for the ubiquity of the system so that more accurate models can be developed that do not include redundant sensors that contribute very little to the overall result. Overall, the results demonstrate that P-SVM consistently outperformed other algorithms, while LDA was the worst. This can be attributed to the design of the algorithms. LDA was found to not be able to handle a complicated dataset with hundreds of features well. In contrast, P-SVM is a complex machine learning method that is capable of solving difficult classification problems. However, kNN is a tradeoff between LDA and P-SVM, as it has a lower computational cost than P-SVM, without diminishing performance.

The first analysis included evaluating the sensors in terms of their broad categories of embedded (Table 2), wearable (Table 3) and both wearable + embedded (Table 4). The results of the embedded sensor dataset were relatively poor,

achieving a maximum accuracy of 59% using both kNN and P-SVM. However, P-SVM did produce a marginally higher F<sub>1</sub> score of 63% over kNN (60%), which demonstrates that this classifier was better at correctly detecting genuine fatigue when it occurred. The wearable sensors, which included heart rate, and the wearable + embedded datasets produced similar results with maximum accuracies of 93% and 94%, respectively. These results demonstrate a 34% increase in accuracy over just using the embedded sensors. The inclusion of the embedded sensors only increased performance by 1% over the wearable sensors, which demonstrates that wearable sensors alone are sufficient for detecting fatigue.

The first set of results demonstrated that the embedded sensors contributed very little to the overall performance of the system. With this in mind, the second analysis aimed to understand how each individual wearable sensor contributed to the system's performance (Table 5). Individually, the accuracy ranged from 51 to 64%. On its own, heart rate produced the lowest accuracy that ranged from 48% (LDA) to 51% (P-SVM), which represents a 42% decrease over using all the wearable sensors together (Table 3). In terms of the wearable sensors that were used to capture motion, individually these sensors produced accuracies that were on average 30% lower than using all the wearable sensors together (Table 3). The results in Table 5 ranged from 63% (right shank) to 64% (right forearm). The results demonstrate that individually the wearable sensor's ability to satisfactorily distinguish fatigue is inadequate and so a combination of sensors is required to detect fatigue with a reasonable degree of accuracy.

The final analysis sought to build on the individual analysis by minimizing the number of wearable sensors used to determine a minimally viable set of sensors that could be used to adequately detect fatigue (Tables 6, 7, and 8). This process included first removing the worst performing sensor (heart rate) and then the highly correlated wearable motion sensors. Firstly, using only the wearable motion sensors (Table 6), the accuracy did not change at 93% and remained steady to those in Table 3, which did include heart rate. This illustrates that heart rate wasn't contributing to the results. In terms of the wearable motion sensors, the arms, hip, and upper spine had the highest correlation with each other (Fig. 15). As the upper spine sensor was uncomfortable to wear, this was removed first. Compared to Table 6, results in Table 7 illustrated that removal of the upper spine sensor marginally reduced the accuracy to 91%, which is a negligible decrease of 2% and illustrates that the upper spine also was not a strong contributor to the results. Additionally, the right arm sensor was removed due to its correlation with multiple sensors. The results whereby the upper spine and right arm were removed (Table 8) still produced a respectable maximum accuracy of 88%, which

**Table 9** Comparison between previous studies

Article	Sensor measurements	Output	Accuracy
This study	IMUs–body motion PPG–heart rate Force plate–ground reaction force Load cell–hand exertion force	Body fatigue (Borg scale/personalized threshold)	88.12–94.15%
[8] (2013)	IMUs and goniometers–joint angles	RULA score	–
[11] (2019)	PPG–heart rate EDA–skin conductance Skin temperature	Physical demand (energy expenditure program)	90%
[12] (2017)	IMUs–body motion ECG–heart rate	Body fatigue (Borg scale/fixed threshold)	79% (specificity)
[14] (2017)	Force plate–plantar pressure	Quasi-static postures Dynamic postures	100% 92%
[15] (2016)	Load cell–hand exertion force VICON–body motion	Manual force strength	–
[44] (2020)	IMUs–body motion ECG–heart rate	Body fatigue (Borg scale/fixed threshold)	87.90%
[47] (2017)	RSP–breathing rate PPG–heart rate EDA–skin conductance	Body fatigue (Borg scale/fixed threshold)	75%
[48] (2017)	PPG–heart rate EEG–brainwave signal Skin temperature	Body fatigue (Borg scale/fixed threshold)	82.60%

compared to Table 6 is a decrease of 5%. These results are important to understand the ubiquity of the system for deployment in an aviation manufacturing environment as too many wearable sensors can be obtrusive and cause discomfort, while deployment of embedded solutions can incur a high infrastructure cost.

Table 9 illustrates a comparison between our work and previous studies in the area. Previous studies utilized joint angles that have been derived from IMUs to identify risky postures [8]. However, using joint angles alone does not provide the necessary granularity for detecting WMSDs, as they are caused by a combination of factors, including body motion and external force. Rather than using joint angles, this study aims to understand the nature of the data using body motion, physiological and external force measurements.

Heart rate has also been widely applied to ergonomic analysis. Most research measures heart rate using electrocardiogram (ECG) sensors [12, 44]. Although ECG is the gold standard for measuring the heart via electrodes attached to the chest [45], the utilization of such devices in a real-world factory setting is not suitable. For instance, if these devices are worn for an extended period of time throughout the day, their attachment to the skin can degrade over time [46]. However, our approach uses a PPG sensor for optical sensing, which is more comfortable, can be worn for a long duration and is a sensor that is found in the majority of smartwatches. Although we have taken raw PPG from the ear, the approach can be adapted to a real-world setup, where the use of smartwatches to obtain similar data is an

interesting avenue to further explore. In contrast to [12, 44, 47, 48] who labelled their data using a subjective rating at the end of the task, our approach to labelling the data included applying the results of processing subjective fatigue levels to obtain a change score between the start and end of a task that represented the level of exertion for each task. This has been beneficial because fatigue fluctuates and changes from the beginning of a task to the end and thus calculation of change scores better represented the physical demand of a task. Additionally, a fixed threshold of fatigue classification has been implemented in [11, 12, 44, 47, 48] that has been used for all participants. This can lead to bias and unrepresentative results as fatigue and feelings of fatigue differ between individuals. Therefore, we have labelled the data based on the change scores for *each* participant, which takes personal differences into consideration and thus provides a more personalized approach. Furthermore, our maximum classification accuracy of 94.15% provides an improvement over previous work [11, 44, 48]. Research shown in [11, 48] used physiological data only for physical demand classification, while we included more types of data sources, which provided more aspects of body-related information.

Our proposed approach has implications for advancing the area of fatigue detection, by transforming raw human-centric data into in-depth body-related information. While capturing data is relatively straightforward, processing enormous amounts of multimodal sensor data is challenging. This paper addresses these issues through the development of a data processing pipeline that synchronized independent data streams together on a common time basis

for the purpose of detecting fatigue. This has great potential to reduce WMSDs by providing a method of capturing body-related information and analyzing this data to improve workers' situational awareness of their body conditions. This is important to facilitate them in actively calibrating their method of performing tasks. For example, by taking longer breaks, adjusting their body postures, and controlling their motion speed. While the results are promising there are some limitations. For instance, the dimensions of the force plate meant that this sensor could not be used in all the tasks. As such, it was excluded from tasks that required the participants to move off it, such as two-handed trolley pushing/pulling. Moving forward, results will also need to be visualized within the system to enable real-time ergonomic monitoring that provides feedback to workers. This feedback can be delivered through smart devices, such as smartphones and smartwatches, as a method of alerting them once they have sustained a high body fatigue level for an extended duration. In this way, workers will have a clearer insight into their workload distribution during a shift. This will greatly improve the health and safety of workplaces by empowering workers to be in control of their health to avoid excessive workloads. Lastly, factory-wide situational awareness of operations can be hugely improved through the collection of body fatigue levels, which can systematically analyze the effect of tasks towards a group of workers. For example, if most workers display a high body fatigue for a certain task, then the task can be redesigned to ensure its ergonomic safety.

## 6 Conclusions and future work

WMSDs pose a negative and long-term effect on both productivity and workers' health. This paper presents a machine learning approach for detecting fatigue utilizing body part motion, external force, and physiological data. An experiment has been conducted to simulate daily factory work and data has been collected from thirty participants while they performed a series of manual physical tasks. A data processing pipeline has then been implemented to pre-process multimodal data and extract features for the machine learning models. Results indicated that the P-SVM model was able to provide a maximum accuracy of 94.15% for the binary detection of fatigue when utilizing the dataset that combined both wearable and embedded sensors, which is an improvement over similar work in this area.

While the results are encouraging, future work aims to build on this work to further improve the system. For instance, in this study, six wearable and two embedded sensors were included to measure different body parameters and provide systematic fatigue detection. Although the results were promising, deploying that many sensors, per worker, in the real world can be costly, obtrusive, and uncomfortable, and it may disturb or

impede them doing their job. Future work therefore aims to explore the efficacy of commercial devices while undertaking such tasks. Moreover, it is impractical to deploy force plates and load cells in a factory environment, due to the high cost and impracticalities of gathering data, as workers need to remain on the plate in order to gather data. Therefore, optimizing the sensor configuration is required before it can be considered in a real-world application. Additionally, the current system works offline. A further line of enquiry would be to improve the system to construct a real-time feedback loop that includes a visualization component, which can detect and display the current level of fatigue as workers complete their tasks. Furthermore, compared to a laboratory environment, a real factory environment can pose different types of external signal interference, including noise and high temperature. As such, it is essential to test and validate the system's performance in a real factory environment.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00779-023-01718-z>.

**Acknowledgements** The authors would like to thank all of the participants for agreeing to take part in the study. Authors listed by contribution.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions This work has been supported by a Boeing – UQ Research Alliance PhD Scholarship and The University of Queensland.

**Data availability** The data that support the findings of this study are available from The Boeing Company, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of The Boeing Company.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Oakman J, Clune S, Stuckery R (2019) "Work-related musculoskeletal disorders in Australia"
2. U.S. Department of Labor Occupational Safety and Health Administration, "Ergonomics: The Study of Work", 2000.

3. Punnett L, Wegman DH (2004) Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. *J Electromyogr Kinesiol* 14(1):13–23
4. Sultan-Taïeb H et al (2017) Economic evaluations of ergonomic interventions preventing work-related musculoskeletal disorders: a systematic review of organizational-level interventions. *BMC Public Health* 17(1):935
5. National Research Council, Musculoskeletal disorders and the workplace: low back and upper extremities. National Academies Press, 2001
6. Valero E, Sivanathan A, Bosché F, Abdel-Wahab M (2016) Musculoskeletal disorders in construction: A review and a novel system for activity tracking with body area network. *Appl Ergon* 54:120–130
7. Valero E, Sivanathan A, Bosché F, Abdel-Wahab M (2017) Analysis of construction trade worker body motions using a wearable and wireless motion sensor network. *Autom Constr* 83:48–55
8. Vignais N, Miezial M, Bleser G, Mura K, Gorecky D, Marin F (2013) Innovative system for real-time ergonomic feedback in industrial manufacturing. *Appl Ergon* 44(4):566–574
9. McAtamney L, Nigel Corlett E (1993) "RULA: a survey method for the investigation of work-related upper limb disorders. *Appl Ergonomics* 24(2):91–99
10. Peppoloni L, Filippeschi A, Ruffaldi E, Avizzano CA (2016) A novel wearable system for the online assessment of risk for bio-mechanical load in repetitive efforts. *Int J Ind Ergon* 52:1–11
11. Jebelli H, Choi B, Lee S (2019) "Application of wearable biosensors to construction sites. II: Assessing workers' physical demand". *J Constr Eng Manag* 145(12)
12. Maman ZS, Yazdi MAA, Cavuoto LA, Megahed FM (2017) A data-driven approach to modeling physical fatigue in the workplace using wearable sensors. *Appl Ergon* 65:515–529
13. Borg G (1998) Borg's perceived exertion and pain scales. *Human Kinetics*
14. Chen D, Chen J, Jiang H, Huang M (2017) "Risk factors identification for work-related musculoskeletal disorders with wearable and connected gait analytics system," in *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. 330–339
15. Chow AY, Dickerson CR (2016) Determinants and magnitudes of manual force strengths and joint moments during two-handed standing maximal horizontal pushing and pulling. *Ergonomics* 59(4):534–544
16. Plamondon A, Larivière C, Delisle A, Denis D, Gagnon D (2012) Relative importance of expertise, lifting height and weight lifted on posture and lumbar external loading during a transfer task in manual material handling. *Ergonomics* 55(1):87–102
17. National Institute for Occupational Safety and Health, "Ergonomic guidelines for manual material handling", 2007
18. Attal F, Mohammed S, Dedabrishvili M, Chamroukhi F, Oukhelou L, Amirat Y (2015) Physical human activity recognition using wearable sensors. *Sensors* 15(12):31314–31338
19. Yan X, Li H, Li AR, Zhang H (2017) Wearable IMU-based real-time motion warning system for construction workers' musculoskeletal disorders prevention. *Autom Constr* 74:2–11
20. Zhang M and Sawchuk AA (2012) "USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors," presented at the Proceedings of the 2012 ACM Conference on Ubiquitous Computing, Pittsburgh, Pennsylvania.
21. Nath ND, Akhavian R, Behzadan AH (2017) Ergonomic analysis of construction worker's body postures using wearable mobile sensors. *Appl Ergon* 62:107–117
22. Akhavian R and Behzadan A (2015) "Wearable sensor-based activity recognition for data-driven simulation of construction workers' activities," in *2015 Winter Simulation Conference (WSC)*, 6–9 3333–3344.
23. The MathWorks, Inc. "MATLAB." <https://au.mathworks.com/products/matlab.html> (accessed Jan. 19, 2021).
24. Antonsson EK, Mann RW (1985) The frequency content of gait. *J Biomech* 18(1):39–47
25. Mayagoitia RE, Nene AV, Veltink PH (2002) Accelerometer and rate gyroscope measurement of kinematics: an inexpensive alternative to optical motion analysis systems. *J Biomech* 35(4):537–542
26. Zhang Y, Markovic S, Sapir I, Wagenaar RC, Little TDC (2011) "Continuous functional activity monitoring based on wearable tri-axial accelerometer and gyroscope," in *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, 23–26 370–373
27. Brodie MAD, Psarakis M, Hoang P (2016) Gyroscopic corrections improve wearable sensor data prior to measuring dynamic sway in the gait of people with Multiple Sclerosis. *Comput Methods Biomech Biomed Engin* 19(12):1339–1346
28. Salehizadeh SMA, Dao D, Bolkhovskiy J, Cho C, Mendelson Y, Chon KH (2016) A novel time-varying spectral filtering algorithm for reconstruction of motion artifact corrupted heart rate signals during intense physical activities using a wearable photoplethysmogram sensor. *Sensors* 16(1):10
29. Cennini G, Arguel J, Akşit K, Leest AV (2010) "Heart rate monitoring via remote photoplethysmography with motion artifacts reduction. *Optics Express* 18(5):4867–4875
30. Pietrzak P, Meller M, Niedźwiecki M (2014) Dynamic mass measurement in checkweighers using a discrete time-variant low-pass filter. *Mech Syst Signal Process* 48(1):67–76
31. Alaziz M, Jia Z, Liu J, Howard R, Chen Y, Zhang Y (2016) "Motion scale: a body motion monitoring system using bed-mounted wireless load cells," in *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)* 183–192
32. Fukushima H, Kawanaka H, Bhuiyan MS, Oguri K (2012) "Estimating heart rate using wrist-type Photoplethysmography and acceleration sensor while running," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2901–2904
33. Ye Y, Cheng Y, He W, Hou M, Zhang Z (2016) Combining non-linear adaptive filtering and signal decomposition for motion artifact removal in wearable photoplethysmography. *IEEE Sens J* 16(19):7133–7141
34. Lara OD, Labrador MA (2013) A survey on human activity recognition using wearable sensors. *IEEE Commu Surveys Tutorials* 15(3):1192–1209
35. Hassan MM, Uddin MZ, Mohamed A, Almogren A (2018) A robust human activity recognition system using smartphone sensors and deep learning. *Futur Gener Comput Syst* 81:307–313
36. Cook DJ, Krishnan NC (2015) Activity learning : discovering, recognizing, and predicting human behavior from sensor data. John Wiley & Sons Inc, Hoboken, New Jersey
37. Chawla NV, Lazarevic A, Hall LO, Bowyer KW (2003) "SMOTE-Boost: improving prediction of the minority class in boosting," in *Knowledge Discovery in Databases: PKDD 2003*, Springer Berlin Heidelberg 107–119
38. Kotsiantis S, Kanellopoulos D, Pintelas P (2006) Handling imbalanced datasets: a review. *GESTS Int Trans Comp Sci Eng* 30(1):25–36
39. RStudio. "RStudio." <https://rstudio.com/> (accessed Jan. 19, 2021).
40. Russell S and Norvig P (2016) *Artificial intelligence: a modern approach, global edition*. Harlow, United Kingdom, UNITED KINGDOM: Pearson Education Limited
41. Kim S, Nussbaum MA (2014) An evaluation of classification algorithms for manual material handling tasks based on data obtained using wearable technologies. *Ergonomics* 57(7):1040–1051

42. Janidarmian M, Roshan Fekr A, Radecka K, Zilic Z (2017) "A comprehensive analysis on wearable acceleration sensors in human activity recognition. *Sensors* 17(3):529
  43. Kim J-H (2009) Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Comput Stat Data Anal* 53(11):3735–3745
  44. Maman ZS, Chen Y-J, Baghdadi A, Lombardo S, Cavuoto LA, Megahed FM (2020) A data analytic framework for physical fatigue management using wearable sensors. *Expert Syst Appl* 155:113405
  45. Greene S, Thapliyal H, Caban-Holt A (2016) A survey of affective computing for stress detection: evaluating technologies in stress detection for better health. *IEEE Consumer Electronics Magazine* 5(4):44–56
  46. Hovsepian K, Al'Absi M, Ertin E, Kamarck T, Nakajima M, Kumar S, (2015) "cStress: towards a gold standard for continuous stress assessment in the mobile environment *Proc ACM Int Conf Ubiquitous Comput* 2015 493 504
  47. Tsao L, Ma L, Papp C-T (2018) "Using non-invasive wearable sensors to estimate perceived fatigue level in manual material handling tasks," in *International Conference on Applied Human Factors and Ergonomics*, 65–74
  48. Aryal A, Ghahramani A, Becerik-Gerber B (2017) Monitoring fatigue in construction workers using physiological measurements. *Autom Constr* 82:154–165
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.