Advanced Engineering Informatics 38 (2018) 514–526

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Full length article

Automated ergonomic risk monitoring using body-mounted sensors and machine learning

Nipun D. Nath[a], Theodora Chaspari[b], Amir H. Behzadan[a,*]

[a] Department of Construction Science, Texas A&M University, 3137 TAMU, College Station, TX 77843, USA
[b] Department of Computer Science, Texas A&M University, 3112 TAMU, College Station, TX 77843, USA

ARTICLE INFO

Keywords:
Construction health
Wearable sensors
Ergonomics
Overexertion
Human activity recognition
Machine learning

ABSTRACT

Workers in various industries are often subject to challenging physical motions that may lead to work-related musculoskeletal disorders (WMSDs). To prevent WMSDs, health and safety organizations have established rules and guidelines that regulate duration and frequency of labor-intensive activities. In this paper, a methodology is introduced to unobtrusively evaluate the ergonomic risk levels caused by overexertion. This is achieved by collecting time-stamped motion data from body-mounted smartphones (i.e., accelerometer, linear accelerometer, and gyroscope signals), automatically detecting workers' activities through a classification framework, and estimating activity duration and frequency information. This study also investigates various data acquisition and processing settings (e.g., smartphone's position, calibration, window size, and feature types) through a leave-one-subject-out cross-validation framework. Results indicate that signals collected from arm-mounted smartphone device, when calibrated, can yield accuracy up to 90.2% in the considered 3-class classification task. Further post-processing the output of activity classification yields very accurate estimation of the corresponding ergonomic risk levels. This work contributes to the body of knowledge by expanding the current state in workplace health assessment by designing and testing ubiquitous wearable technology to improve the timeliness and quality of ergonomic-related data collection and analysis.

1. Introduction

With advancements in mobile technology, modern smartphones are now equipped with a host of sensors which can capture location and motion-related data of a person within the environment. These devices have the potential to facilitate everyday life in various ways by giving users contextual information about their activities, interests, and surroundings without being obtrusive and interruptive. In addition, compared to other classes of data-capturing devices, smartphones are more ubiquitous (thus more affordable) and intuitive to use, can be controlled and operated remotely (using the cloud technology), and require a relatively lower maintenance and operating costs. The value of using smartphones in domains such as healthcare, wellbeing, and behavioral analysis has been investigated over the past few years. For example, smartphones are being used for monitoring patients and elderly people [1–3]. In addition to health monitoring, smartphones can also be used in managing and promoting human well-being [4,5]. Also, smartphone technology can be integrated with behavioral health care [6]. For instance, Timmons et al. [7] used audio and global positioning system (GPS) data from smartphones to unobtrusively and remotely monitor the behavior of young couples. Furthermore, smartphone's built-in inertial measurement unit (IMU) can be utilized to prevent work-related injuries, for example, fall from a height [8], shoulder injury [9], and upper-limb injury [10]. Particularly, recent studies have explored the potentiality of smartphone sensor in preventing musculoskeletal disorders (MSDs) associated with awkward posture [11,12].

MSDs are major health issues that affect a large number of individuals across many occupations and industries (e.g., from office space work to manufacture and construction), leading to long-term disability and economic loss [13]. MSDs refer to a group of disorders or injuries resulting from the stress in a person's inner body parts (e.g., muscles, nerves, tendons, joints, cartilages, and spinal discs) while the person moves [14,15]. Examples of MSDs include Carpal Tunnel Syndrome (CTS), Tendonitis, and Bursitis [16,17]. MSDs caused particularly due to the activities in a workplace are referred to as work-related musculoskeletal disorders (WMSDs). In 2009 alone, direct workers' compensation costs due to WMSDs were amounted to be more than $50 billion in the U.S. [18]. Moreover, workers exposed to major WMSDs may face permanent disability that can prevent them from carrying out their professional tasks and, in severe cases, regular everyday tasks

* Corresponding author.
E-mail addresses: nipundebnath@tamu.edu (N.D. Nath), chaspari@tamu.edu (T. Chaspari), abehzadan@tamu.edu (A.H. Behzadan).

https://doi.org/10.1016/j.aei.2018.08.020
Received 4 February 2018; Received in revised form 2 June 2018; Accepted 31 August 2018

[15]. In 2015, workers employed by the private sector in the U.S. required a median of 12 days to recover from WMSDs before they could return to work [19].

The construction industry is considered as one of the most ergonomically hazardous occupations [20]. Compared to other industries, construction activities are more physical and labor-intensive. Moreover, with increasing complexity and scope of construction projects, workers (especially those with limited skills and training) often find themselves performing tasks that are beyond their natural physical limits [21]. This sustained physical labor over a long period of time can trigger WMSDs which in turn adversely affect the project budget, schedule, and productivity. In 2015, the WMSD-related incident rate (number of illnesses and injuries per 10,000 equivalent full-time workers) was 34.6 [19]. WMSDs are the major source of concern in other industries as well. Among all goods-producing sectors, workers in the manufacturing, agriculture, forestry, fishing, and hunting sectors, and among all the service-providing sectors, workers in the transportation, warehousing, healthcare and social assistance sectors are reported to be more exposed to WMSDs [20]. Nursing assistants, laborers, and freight, stock and material movers experienced the highest number of WMSD cases in 2013 [20].

To prevent WMSDs, various health and safety organizations have established rules and guidelines to identify the risks associated with performing certain tasks. Such efforts aim at the ergonomic design of project tasks, tools, and workplace to match physical jobs with workers' natural body capacities. As an example, the prevention through design (PtD) initiative, introduced by the National Institute for Occupational Safety and Health (NIOSH), aims at limiting and ultimately preventing occupational injuries, illnesses, and fatalities that can be achieved by eliminating the potential risks to workers at the source as early as possible in a project life cycle [22]. Since a proper PtD practice requires prior identification of the risk factors, it is necessary to collect adequate spatiotemporal work-related data. The collected data, if properly analyzed and interpreted, can be used to promote workers' safety and health by improving the quality of job training and eliminating potential ergonomic risks in the workplace.

Field practices of data collection are traditionally based on self-reporting, manual observation, or the use of sophisticated sensor networks. Such practices, however, are time-consuming, naturally obtrusive, and require technical knowledge that may not be available among construction practitioners. Therefore, the objective of this research is to design and test a methodology where an unobtrusive and automated data processing framework is used to calculate ergonomic risks associated with occupational tasks, in particular, those comprising the use of excessive force (overexertion). In the designed methodology, mobile technology (smartphones) is used to collect multi-modal time-motion data from the workers while they perform different activities. Next, machine learning will be used to recognize workers' activities, and then, activity duration and frequency information will be extracted. The output of this step will be subsequently used to identify the ergonomic risk levels for each worker. Calculated risk levels can be used to identify major sources of ergonomic risks which can help workers and decision-makers (e.g., project managers, safety officers, superintendents) to take proper actions to preemptively limit and ultimately eliminate such risks by redesigning high-risk activities and/or workspaces.

2. Literature review

With 33% of all cases, the U.S. Bureau of Labor Statistics [19] ranks overexertion first in the leading events or exposures that cause WMSDs. By definition, overexertion is the event category that includes injuries related to exerting an excessive force beyond the body's capacity. Activities that require force can be categorized into two groups: lifting/lowering/carrying (category-1), and pushing/pulling (category-2) [23]. A risk factor is defined as a condition present in the workplace that is directly responsible for health hazards [17]. For example, applying excessive force to lift a heavy object can be considered as a risk factor for overexertion. However, the mere presence of a risk factor is not sufficient to evaluate the risk associated with a task, rather the risk also depends on the extent of the risk factor [17]. Determining if an exposure or a risk factor will result in WMSDs depends on intensity, duration, and frequency, or a combination of these factors [24]. Intensity, duration, and frequency refer to how much, how long, and how often, respectively, one is exposed to a risk factor. Generally, risk level rises with the increase of these factors. For instance, if a worker forcefully (i.e., intensity factor) and repetitively (i.e., frequency factor) pushes a heavy object for a long period of time (i.e., duration factor), the worker is exposed to WMSDs (e.g., back pain). These are regulated by the Occupational Safety and Health Administration (OSHA), which has provided a set of empirical rules assessing the risk of activities according to their type, duration, and frequency.

Towards this goal, three different approaches have been practiced in general: (1) self-assessment, (2) observational, and (3) direct measurement [25]. In self-assessment, workers are asked to provide risk-related data. Though this approach has low initial cost and is straightforward, researchers have stated that workers' self-assessments on exposure levels are often imprecise, unreliable, and biased [26]. The observation-based approach involves real-time assessment or analysis of the recorded video. But it is mostly impractical in nature due to the substantial cost, time, and technical knowledge required for post-analysis of large amounts of non-heterogeneous data [24].

Unlike the previous two approaches, direct measurement uses tools to collect workers' posture- and motion-related data. Examples of this approach include but are not limited to using off-the-shelf micro-electro-mechanical sensors (MEMS), e.g., IMUs, and vision-based sensors. Vision-based sensors such as Red-Green-Blue (RGB) camera and Kinect suffer in extreme lighting conditions and optical occlusions [27]. For this reason, wearable sensors such as IMUs have gained more popularity for being inexpensive, easy to install and maintain, and requiring minimum training for data collection and human activity recognition (HAR) [28]. Moreover, previous studies have shown that when compared to the depth-based sensors (e.g., Kinect), IMUs are superior for detecting movements of body parts because they are more sensitive than Kinect (i.e., capable of capturing subtle movements), more robust (i.e., capable of providing stable data), and have higher sampling rate (e.g., > 50 Hz, while the maximum frequency for Kinect is 30 Hz) [27]. While previous studies in this area have revealed some of the shortcomings of the direct measurement approach including high initial investment cost, maintenance cost, and technical knowledge to interpret data, compared to other approaches, this method by far yields the most valid assessment of risk factors [24,29].

In order to overcome the implementation challenges of direct-measurement approach, the authors used smartphones as a data collection device. Recent work has explored the merit of built-in smartphone IMU sensors to collect input data for machine learning algorithms to identify field activities and to estimate activity durations [30–33]. We have to note that compared to traditional physical activity recognition (e.g., walking, running, and sitting) [34–38], the activities performed in construction sites are much more complex in nature (e.g., loading, unloading, lifting, lowering, carrying, pushing, and pulling). Previous efforts in identifying construction activities include the use of single-sensor (i.e., accelerometer) data to identify masonry work [39,40]. Particularly, Ryu et al. [40] have used data from wrist-worn accelerometer sensors to classify more subtle mason's actions (e.g., spreading mortar, laying bricks, adjusting bricks, and removing excess mortar). While past work has mainly focused on activity recognition, the literature is rather limited and fragmented about the prospect of identifying ergonomic risks, particularly those associated with overexertion, from the outcome of multi-sensor HAR. Therefore, the applicability and robustness of existing methods are to a large extent unexplored in overexertion-related ergonomic risk assessment. Given

the depth and breadth of knowledge and building upon the findings of previous work, the authors extended the previously developed methodology to automatically evaluate the ergonomic risk levels of various construction activities. Moreover, since placement variations of electronic sensors can result in significantly different acceleration signals and, thus, impacting the final performance of HAR [41], the optimum position for placing the sensors need to be investigated. With this objective, the authors also identify the optimal data acquisition and data processing settings (e.g., smartphone's position, window size, and feature set) that yield reliable estimation of ergonomic assessment. The novelty of this research lies in identifying ergonomic risk levels due to overexertion using raw time-motion multi-sensor data captured by body-mounted smartphones. Findings of this research will have significant practical implications in the construction domain since developed methods can be used to design unobtrusive and inexpensive real-world systems for automatic estimation and generation of real-time warnings to construction workers regarding potential overexertion and ergonomic risks merely through the use of a smartphone device.

3. Overall framework and problem formulation

The main building blocks of the designed framework are shown in Fig. 1. These include optimization of data acquisition and data processing settings, human activity recognition (HAR), estimation of activity durations and frequencies, and assessment of overexertion and ergonomic risks. The first step of this process involves identifying the optimal data-acquisition setting (i.e., the optimal position to place a smartphone on workers' body) and data-processing settings (e.g., the optimal window size for data segmentation, and the best feature set) that would yield highly accurate classification result. This starts with collecting time-stamped data using smartphone sensors (worn by workers), followed by pre-processing the collected data and converting them into a set of distinctive features using different combinations of settings. Next, for each combination, a classifier model is built and performance of the model in terms of weighted accuracy, precision, recall, and F1 score, and Spearman's rank correlation is recorded. As an example, the equation for calculating weighted recall is given in Eq. (1), where m is the number of classes, $Recall_i$ is the recall for class i ($i = 1, 2, \ldots, m$), and $Samples_i$ is the number of samples in class i ($i = 1, 2, \ldots, m$).

$$Weighted \quad Recall = \frac{\sum_{i=1}^{m} (Recall_i) * (Samples_i)}{\sum_{i=1}^{m} Samples_i} \tag{1}$$

The setting that generates the best performing classifier model is considered as the optimal setting. Next, HAR takes as an input the decisions from the best performing setting, performs smoothing to remove outliers, and estimates activity duration and frequency information. Details of the activity recognition process can be found in the previous study by Akhavian et al. [30]. Finally, estimated durations and frequencies are used to determine the ergonomic risk levels associated with each activity performed by the workers.

4. Experiment design and data collection

To test the designed methodology, a field experiment is carried out. The experiment is designed based on the following criteria:

(1) Activities performed in the experiments resemble real-world activities.
(2) Activities have potential to cause ergonomic harms (i.e., activities are repetitive in nature and require excessive force).
(3) Activities involve two categories of events associated with overexertion. Examples include lift/lower/carry (category-1), and push/pull (category-2). From an ergonomics perspective, these two categories are regulated by two different sets of rules that limit the duration and frequency of the activities (as summarized in Table 3) with potential of leading to WMSDs [42].
(4) Subjects involved in the experiments are allowed to perform their assigned tasks at their own pace.

The experiment resembles a warehouse operation which includes uncertainty in the workflow. The goal of this experiment is to transport an item (i.e., a box) from a loading area to an inspection area, inspect the item and if the item is accepted, to move it through the system to a designated unloading area. As shown in Fig. 2, the cyclic operation starts with a worker loading a box onto a cart and then pushing it to the inspection area. Next, an inspector lifts the box and inspects it. During the inspection, the worker waits in the inspection area. After the inspection, the inspector either accepts the box or rejects it. Upon acceptance, the worker lowers the box onto the cart, pushes it to the unloading area, unloads the box and then pulls the cart back to the loading area. If the box is rejected, the worker pulls the cart back to the loading area. In both cases, the worker moves back to the loading area and the cycle starts over. This operation is performed for 15 cycles for two workers, referred as worker *W1* and worker *W2*, respectively. To collect data, two smartphones (e.g., Google Nexus 5X and Google Nexus 6) are mounted on each of the performer's body (one on the upper arm and another on the waist). Data are collected from the accelerometer, linear accelerometer, and gyroscope sensors. The activities are manually annotated based on inspecting the recorded video and marking the starting and ending of each activity (category 1: lift/lower/carry, category 2: push/pull, category 0: any other no-risk activity). The total duration of the data is approximately 40 min.

5. Methodology

In this section, the process starting from data collection and process optimization to ergonomic risk assessment is demonstrated in detail.

5.1. Optimization of data acquisition and data processing settings

Previous studies suggested that arm and waist are relatively better positions to attach smartphone for HAR [30,43]. To assess previous findings within the context of this work, two smartphones are mounted on each worker's body, one on the upper arm and another on the waist. The objective is to investigate which position would yield better classification results. The classifier models built using data acquired from the arm- and waist-mounted smartphone are referred to as *Arm* and *Waist* models, respectively.

The orientation of the smartphones (with relative to the body) might be different for different subjects. In fact, when smartphones are attached to different subjects, or to the same subject but at different times, it is impossible to guarantee that the orientations will be the same every time. Therefore, there is a possibility that sensor signals might be different (which would make the classification less accurate) even if the subjects physically move their body parts in a similar manner (Fig. 3**(a)**). A potential solution to this issue is to ensure that sensor signals could not be significantly different due to the variance in

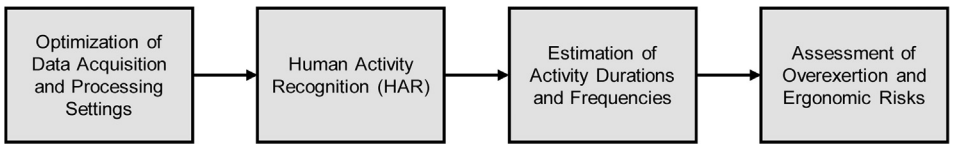

Fig. 1. Schematic framework for the assessment of ergonomic risk associated with overexertion.

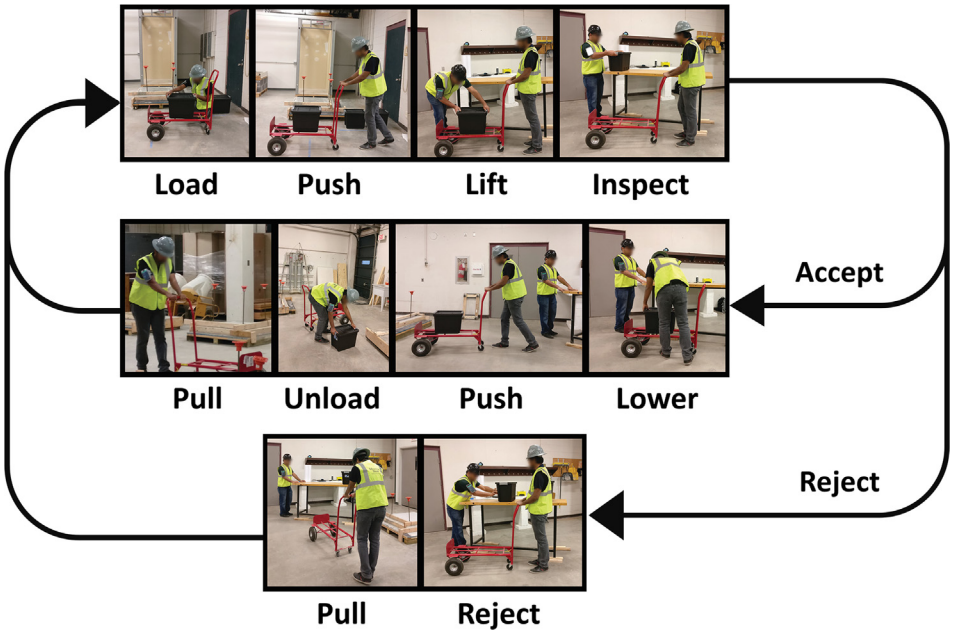

Fig. 2. The cycle of activities in the experiment.

the orientation of the smartphone. Therefore, it is assumed that all sensor readings for activity 'wait' (when workers stand still) would be zero regardless of the orientation of the smartphone. In order to implement this assumption, before starting each session of activity recognition, workers are asked to wait (i.e., standing in a still position) for a certain period of time to establish a benchmark signal reading. The average of sensor signals, recorded in that period, is deducted from the raw sensor signals (Fig. 3**(b)**) collected throughout the actual experiment. The process is referred to as *calibration* in this work. Similarly, the classifier models built with and without calibration are referred to as *Calibrated* and *Not Calibrated* models, respectively.

In the conducted experiment, data are collected using smartphone's built-in accelerometer, linear accelerometer, and gyroscope sensors. To extract statistical features such as the interquartile range (IQR), it is important to have continuous and uniform time series data. Smartphone sensors, however, occasionally freeze and stop recording data, and upon recovering, record data at a higher sampling rate to compensate [30]. Therefore, raw sensor data usually form non-uniform time series containing high-frequency data points in some time intervals (which are redundant) and low-frequency data points in some others (i.e., missing data). Similar to the previous study [44], in order to have a continuous and orderly data stream, raw sensor data is resampled into a 180 Hz of uniform time series data using linear interpolation. The sampling rate of the raw sensor data (before resampling) is given in Table 1.

Next, to obtain more motion-related features, jerk (the difference between two consecutive data points, e.g., $Jerk_a = a_t - a_{t-1}$) and magnitude of the tri-axial data (e.g., $a = \sqrt{a_x{}^2 + a_y{}^2 + a_z{}^2}$) are calculated [45]. Afterwards, the sensor data is segmented into a series of fixed-length time-windows with 50% overlap. Past research in HAR has achieved promising results by choosing window sizes between 1 and 3 s [30,31,34]. Also, previous studies have found that the accuracy of classification generally decreases with an increase in window size [32]. Therefore, in this study, in order to find the optimum window size, three different sizes, namely 1 s (180 data points), 2 s (360 data points), and 3 s (540 data points) are examined. The classifier models built with data segmented into 1-, 2-, and 3-second windows are referred to as *1-Second*, *2-Second*, and *3-Second* models, respectively. The resulting number of samples for each time-segmentation, as well as their distribution into the three activity categories, can be found in Table 2.

Next, key statistical features from each sensor stream including the mean, minimum, maximum, standard deviation, IQR, skewness, kurtosis, mean absolute deviation, and the 4th-order autoregressive coefficients are computed over the corresponding window. This results in a 12-dimensional feature vector per sensor, or 288 features in total (as shown in Fig. 4). Previous studies have indicated the good discriminatory ability of such features for human activity recognition tasks [34,35]. In order to identify the most effective features (a.k.a. distinctive features), ReliefF [46] ranks the original 288 features in order of their effectiveness (a.k.a. weight). Assume, a feature is denoted as f_r where r is the rank of the feature determined by the ReliefF algorithm. Thus, the feature space can be written as $\{f_1, f_2, f_3, ..f_{288}\}$ where f_1 is the best feature and f_{288} is the worst feature. A subset of this feature space, F_n, refers to the set of first n features, i.e., $\{f_1, f_2, ..f_n\}$. For each subset, from F_{11} (containing first 11 features) to F_{288} (containing all 288 features), a classifier model is built using the training dataset and cross-validation results are recorded. The initial value of n was set at 11 (i.e., first 11 features) because it was empirically found that using fewer features would result in a relatively less accurate model. Among all the classifier models built, the one that yields the best F1 score is referred to as the *Best Features* model. For example, as illustrated in Fig. 4, the *Best Features* model uses the feature subset $F_i = \{f_1, f_2, f_3, ..f_i\}$. On the other hand, the classifier model built with all features (i.e., F_{288}) is referred to as *All Features* model.

In all experiments, we use support vector machine (SVM) implemented in MATLAB® to classify each window to one of the three activity categories (category 1: lift/lower/carry, category 2: push/pull, category 0: any other no-risk activity). During this multiclass classification task, for each pair of classes we construct one binary SVM classifier with the polynomial cubic kernel, and make the final decision using majority voting (commonly referred to as "one-vs-one" setting). Also, a leave-one-subject-out cross-validation setup is used, according to which the feature selection process and the training of the classifier are performed in the training set of each fold, and results are computed on the corresponding test set. The final reported results are averaged across all folds.

The goal of the optimization process is to investigate all possible combinations of the aforementioned settings (illustrated in Fig. 5), build and examine classifier model for each case, and identify the best combination of settings for which classifier model produces the highest

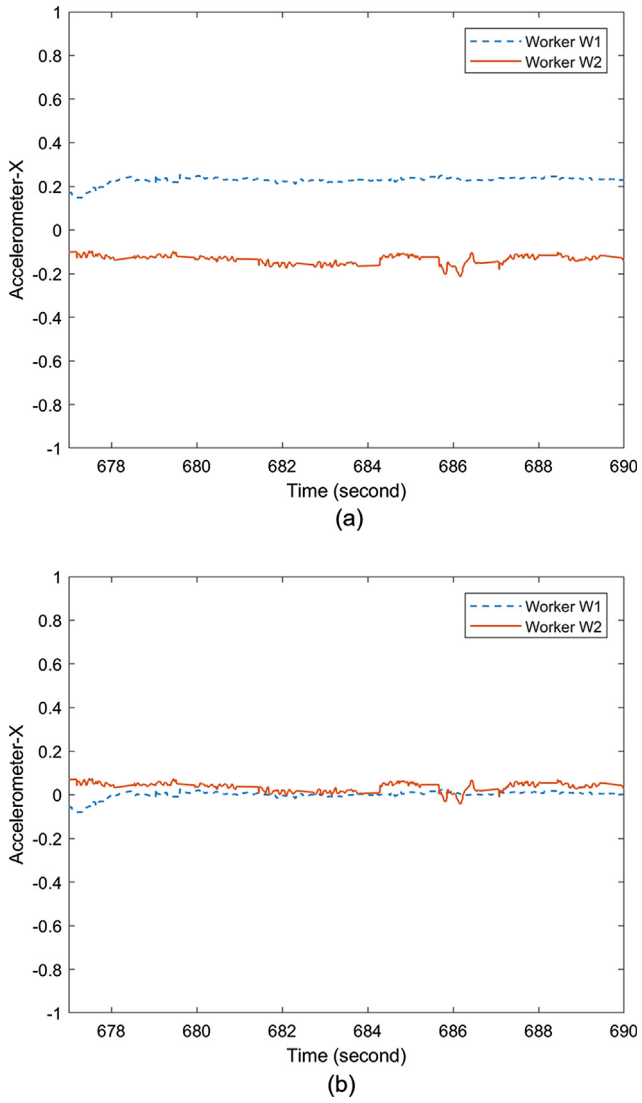

Fig. 3. Accelerometer-X readings with (a) and without (b) calibration when the workers were waiting.

F1 score. Next, applying this optimal setting, classifier model is built for activity recognition.

5.2. Activity recognition

The activities involved in the process fall into two categories of events that associate with risks due to overexertion, namely lift/lower/carry (category 1), and push/pull (category 2). Any other activity that does not associate with such risks is assigned to the "none" category (category 0). Similar to Section 5.1, in this step, we also use SVM (with the polynomial cubic kernel and "one-vs-one" setting) to classify each window to one of these three activity categories. Experiments are performed through a leave-one-subject-out cross-validation procedure. According to this procedure and given that we have two subjects, data from worker *W1* are used for training and data from worker *W2* are used for testing during the first fold, while the reverse occurs during the second fold. Median filtering is performed on the predicted labels in order to remove outliers. Because of the unbalanced number of samples from each class, the activity recognition performance is evaluated using weighted precision, recall, and F1 score measures (e.g., Eq. (1)). Spearman's rank correlation was computed between the actual and predicted time series of 0, 1, 2 values (each corresponding to one category).

5.3. Duration and frequency estimation

One instance of an activity category is defined as a group of consecutive windows which are classified as a similar category of activities (0, 1, or 2). The duration of each instance is calculated by counting the number of windows in that group and multiplying the result by half of the window size (since windows are 50% overlapped). The total duration of a category is determined by summing up durations of all instances of that category. Additionally, the frequency of a category, indicating how many times a category of activity is performed [17], is determined by counting all the instances of that category.

5.4. Ergonomic risk assessment

OSHA reports that ergonomic risk levels (low, moderate, and high) can be estimated through the frequency and duration of lifting/carrying/lowering, pushing/pulling according to Table 3 [42]. In this Table, risk level refers to the likelihood of WMSD-related injury to occur. These threshold values can be used to check ergonomic risk levels to achieve compliance with the ergonomics requirements for musculoskeletal injury in the OSHA regulations (i.e., once workers are exposed to a risk factor, the employer must assess the risk level) [42]. Moreover, this Table is a useful tool for selecting appropriate risk control measures. For example, if a field observation determines that an activity exposes a worker to high risk, the requirements for lower risk categories can be checked and applied to resolve that particular situation. Furthermore, for each category, the calculated risk level can be compared before and after ergonomics improvements to better quantify the risk reduction.

For calculating the ergonomic risk levels, durations estimated in the previous step are expressed as percentages of work shift where a shift is the total duration of the experiment. Also, estimated frequencies are expressed as frequency per minute of the shift. Next, based on Table 3, corresponding risk levels are calculated.

6. Results and discussions

This section reports on the results of activity recognition, duration and frequency estimation, and ergonomic risk assessment. Furthermore, the effect of the various data acquisition and processing settings (e.g., smartphone's position, calibration, window size, and feature set) on the considered performance metrics is examined. Of note, all experiments are performed using a leave-one-subject-out cross-validation.

6.1. Optimal data acquisition and data processing settings

Performance metrics of classifiers with different settings for parameters are summarized in Table 4 and illustrated in Fig. 6. It can be

Table 1
Sampling rate of the raw sensor data collected from the smartphones.

Worker	Smartphone's Position	Sampling Rate (Hz)
W1	Arm	190
	Waist	35
W2	Arm	70
	Waist	180

Table 2
Number of samples per activity category for different window sizes.

Window size	Class 0	Class 1	Class 2	Total
1 s	1064	1056	2829	4949
2 s	532	527	1414	2473
3 s	355	354	938	1647

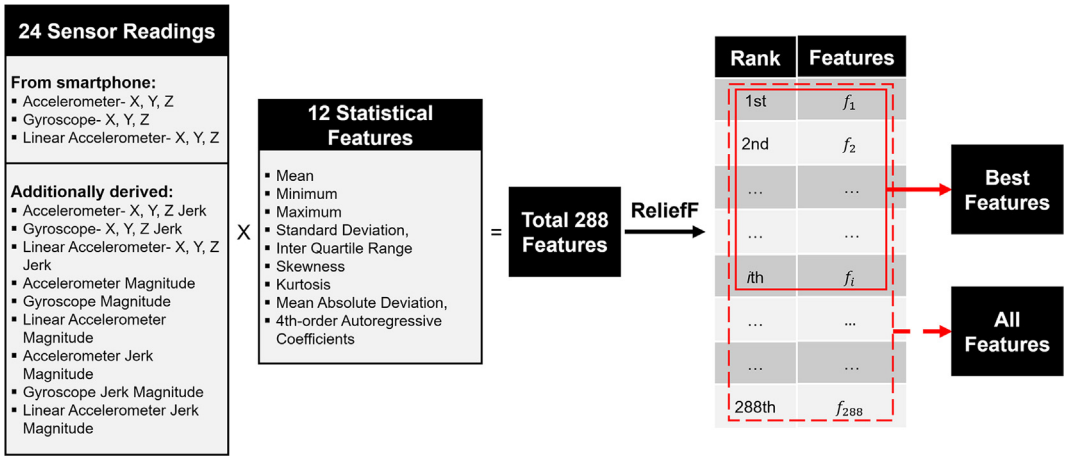

Fig. 4. Feature extraction and feature selection process.

seen that the classifier model built from the data captured with arm-mounted smartphone, calibrated sensor readings, 2-second windows, and best features performed best in terms of all performance metrics. Even without the use of feature selection, adding all the features collected from the arm as an input to the classifier still yields good performance.

6.1.1. Selection of smartphone's position

For the smartphone's position, Fig. 6 shows that the *Arm* models performed significantly better than the *Waist* models. The comparison of accuracies along with improvement in accuracy and relative reduction in error are summarized in Table 5. The overall accuracy of *Arm* models is 19.4% higher than that of *Waist* models, resulting in 49.2% reduction in error. The equation for calculating the relative reduction in error for model *A*, compared to model *B*, is given in Eq. (2), where $Error_A$ and $Error_B$ refer to the classification error of the model *A* and *B*, respectively, and $Accuracy_A$ and $Accuracy_B$ refer to the accuracies of the model *A* and *B*, respectively.

$$Relative\ \ Reduction\ \ in\ \ Error = \frac{Error_B - Error_A}{Error_B} = \frac{(1-Accuracy_B)-(1-Accuracy_A)}{1-Accuracy_B} = \frac{Accuracy_A - Accuracy_B}{1-Accuracy_B} \quad (2)$$

Confusion matrices for the *Arm* and *Waist* models (both built from calibrated sensor readings, 2-second windows, and best features) are shown in Table 6. The table shows that for arm-mounted smartphone, all categories are classified with high accuracy, i.e., 92.7%, 82.7% and 92.1% accuracy for category-0, -1 and -2, respectively. In contrary, for waist-mounted smartphone, classification accuracies are lower for all categories. Particularly, the category-1 activity (lift/lower/carry) is confused with category-2 activity (push/pull) 68.9% times (363 instances) and category-2 activity is confused with category-1 activity 13.2% times (187 instances). It suggests that the waist movements (and patterns in the sensor signals) are very similar when workers perform these two categories of activity which, ultimately, results in higher confusion rates between these two categories. However, category-1 has a lower number of training samples compared to that of category-2 (Table 2) and, therefore, the classifier has a higher tendency to misclassify category-1 as category-2. On the other hand, confusion rates between these two categories of activity are relatively low for the arm-mounted smartphone (Table 6). Therefore, it can be inferred that arm movements are fairly distinguishable while the subject is performing category-1 and -2 activities, which justifies the better classification accuracy yielded from that smartphone.

Table 3
Risk levels of category 1 (lift/carry/lower) and 2 (push/pull) activities.

Category	Parameter	Low Risk	Moderate Risk	High Risk
1	Frequency per minute	< 1	1–5	> 5
	Duration/shift	< 25%	25–50%	> 50%
2	Frequency per minute	< 1/480	1/480–10	> 10
	Duration/shift	< 25%	25–50%	> 50%

6.1.2. Calibration of sensor readings

Fig. 6 shows that, in general, *Calibrated* models performed better than *Not Calibrated* models.

Table 7 shows that the overall accuracy of *Calibrated* models is 8.9% higher than that of *Not Calibrated* models, resulting in 26.1% reduction in error. In particular, calibration improved the accuracy of *Arm, 2-Second, and Best Features* model by 11.3% and, thus, reduces the error by 53.7%.

Examples of confusion matrices for a *Calibrated* and a *Not Calibrated* model with all other parameters fixed (i.e., *Arm*, *2-Second*, and *Best*

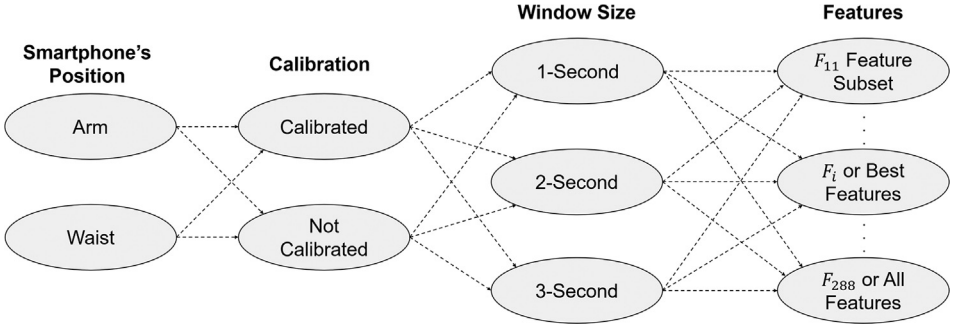

Fig. 5. Optimization process involves investigating different combination of smartphone position, calibration, window size, and feature subsets.

Table 4
Performance metrics of classifiers with different settings for parameters.

Metric	Window Size (sec)	Calibration	Arm		Waist	
			All Features	Best Features	All Features	Best Features
Accuracy/Recall (%)	1	Not calibrated	69.9	70.2	44.8	68.8
		Calibrated	78.3	88.0	63.8	60.7
	2	Not calibrated	77.7	79.0	44.6	63.0
		Calibrated	85.0	90.3	60.2	68.0
	3	Not calibrated	82.1	85.7	52.0	71.7
		Calibrated	86.0	89.1	69.5	72.9
Precision (%)	1	Not calibrated	79.8	79.7	61.7	66.4
		Calibrated	83.0	88.0	64.7	67.4
	2	Not calibrated	85.5	87.1	69.8	57.6
		Calibrated	87.8	90.5	64.1	67.4
	3	Not calibrated	86.0	87.8	74.5	72.5
		Calibrated	86.9	89.3	70.9	72.3
F1 (%)	1	Not calibrated	70.2	70.6	45.4	65.7
		Calibrated	78.9	87.9	61.2	61.5
	2	Not calibrated	78.0	79.7	45.9	57.3
		Calibrated	85.1	90.2	58.9	66.1
	3	Not calibrated	82.9	86.3	55.3	69.9
		Calibrated	86.2	89.1	68.3	71.3
Spearman Correlation	1	Not calibrated	0.69	0.69	0.65	0.35
		Calibrated	0.84	0.75	0.51	0.54
	2	Not calibrated	0.80	0.77	0.39	0.39
		Calibrated	0.85	0.82	0.61	0.52
	3	Not calibrated	0.81	0.77	0.61	0.52
		Calibrated	0.84	0.80	0.68	0.66

Features) are shown in Table 8. The table shows that calibration improved the accuracies of classifying all categories and reduces the confusion rates in most cases. Particularly, the accuracy of classifying category-2 activity (push/pull) is improved by 17.6% (from 74.5% to 92.1%). Also, the rate of confusing category-1 activity (lift/lower/carry) with category-2 activity (push/pull) is reduced by 18.0% (from 24.6% to 6.6%). It indicates that calibration minimizes the intra-class variability and maximizes the inter-class variability and, thus, improves the overall performance of classifier models.

6.1.3. Selection of window size

A general trend of improvement in the performance with the increase in the window size can be seen in Fig. 6. Similarly, Fig. 7 also illustrates that for 6 out of the 9 cases, accuracies were improved for the

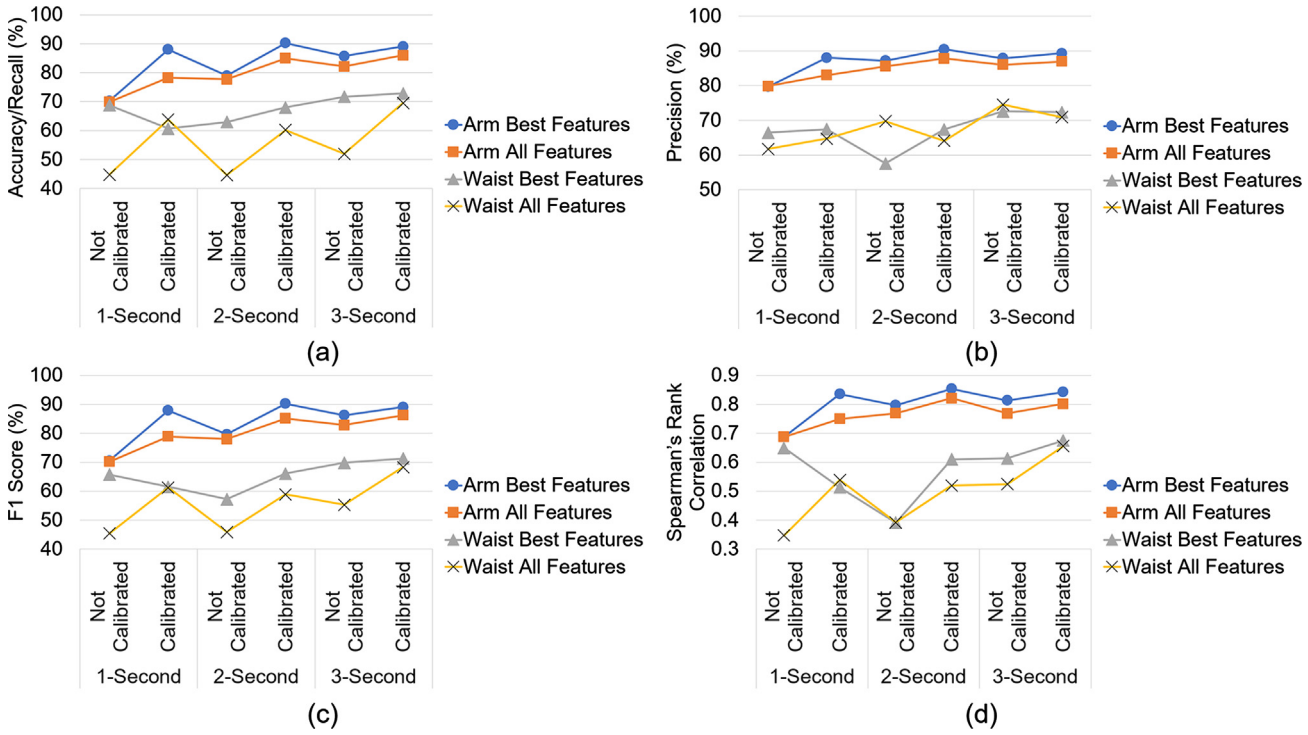

Fig. 6. Performance of classifiers with different settings for parameters in terms of (a) accuracy/recall, (b) precision, (c) F1 Score, and (d) Spearman's rank correlation.

Table 5
Improvement in accuracy and relative reduction in error for the Arm models compared to the Waist models.

	All Features				Best Features				Overall
	1 Second		2 Seconds		3 Seconds		1 Second		2 Seconds		3 Seconds		
	Not Calibrated	Calibrated	Not Calibrated	Calibrated	Not Calibrated	Calibrated	Not Calibrated	Calibrated	Not Calibrated	Calibrated	Not Calibrated	Calibrated	
Waist model (%)	44.8	63.8	44.6	60.2	52.0	69.5	68.8	60.7	63.0	68.0	71.7	72.9	60.6
Arm model (%)	69.9	78.3	77.7	85.0	82.1	86.0	70.2	88.0	79.0	90.3	85.7	89.1	80.0
Improvement (%)	25.2	14.5	33.1	24.8	30.2	16.5	1.5	27.4	16.0	22.2	14.0	16.2	19.4
Relative reduction in error (%)	45.5	40.0	59.8	62.2	62.8	54.2	4.7	69.5	43.2	69.5	49.6	59.7	49.2

Table 6
Confusion matrices for *Arm* and *Waist* models with *Calibrated*, *2-Second*, and *Best Features*.

Model	Actual Class	Predicted Class		
		0	1	2
Arm	0	**92.7%**	0.6%	6.8%
	1	1.5%	**82.7%**	15.7%
	2	1.3%	6.6%	**92.1%**
Waist	0	**63.7%**	19.0%	17.3%
	1	4.0%	**27.1%**	68.9%
	2	1.9%	13.2%	**84.9%**

2-Second models compared to the *1-Second* models and for 8 out of the 9 cases, accuracies were improved for the *3-Second* models compared to the *2-Second* models. One possible reason is that for smaller window size, inter-class variability is less and, therefore, classifier models fail to distinguish between the classes. In other words, while the workers perform category-0 (wait), -1 (lift/lower/carry), and -2 (push/pull) activities, sensor signal (or movement) patterns in smaller segments of time are similar to each other (Fig. 8**(a)**). In contrast, for larger segments, distinguishable patterns in the sensor signals (or movement) emerge (Fig. 8**(b)**). Although this observation might be true for general cases, for the particular experiment conducted in this research, the *2-Second* model (with *Arm*, *Calibrated*, and *Best Features*) performed better than other models and, therefore, 2-second has been selected as the optimal window size.

6.1.4. Selection of feature subset

Not all features are useful for classification since not all features are distinctive. In fact, for better and effective classification, only relevant and important features should be used [47]. Therefore, in the optimization process, various subsets of features are investigated to find the best feature subset. Fig. 9 demonstrates the F1 scores of *Arm*, *Calibrated*, and *2-Second* classifier models with various feature subsets. It can be seen that using all 288 features did not result in the best performance (F1 score of 85.1%). Rather, the best model (F1 score of 90.2%) uses only 28 features (i.e., the first 28 features ranked by ReliefF algorithm). Using the best feature subset instead of all features improves the F1 score by 5.1%. This supports the argument that indistinctive features increase the confusion which, ultimately results in the less accurate classification of activities.

In Table 9, the list of best features for worker *W1* and *W2* is given. It has been found that the statistical features skewness, kurtosis, and autoregressive coefficients are the least effective features. Also, the features extracted from accelerometer and gyroscope sensors are more effective than the features extracted from linear accelerometer. Moreover, 21 out of the 28 features are found to be effective for both workers which are highlighted in italic text in Table 9.

While previous studies have found promising results using a single accelerometer sensor [34–38], in this research, two additional sensors, namely gyroscope and linear accelerometer are used to increase the accuracy of results. The value of these sensors to the overall data analysis can be understood from Fig. 10. This Figure illustrates that the maximum F1 scores using accelerometer, linear accelerometer, and gyroscope sensors individually are 87.8%, 68.9%, and 87.3%, respectively. Thus, it can be inferred that features extracted from accelerometer and gyroscope data contribute the most to the classification accuracy. The presence of features from these two sensors in the best feature subset (see Table 9) also supports this conclusion. Considering the results shown in Fig. 10, when all three sensors are used, the F1 score is improved even more (i.e., 90.2%) and surpasses the F1 scores achieved using any single sensor. Therefore, it can be stated that using both gyroscope and linear accelerometer sensors in addition to the

Table 7
Improvement in accuracy and relative reduction in error for the Calibrated models compared to the Not Calibrated models.

	All Features						Best Features						Overall
	1 Second		2 Seconds		3 Seconds		1 Second		2 Seconds		3 Seconds		
	Arm	Waist	Arm	Waist	Arm	Waist	Arm	Waist	Arm	Waist	Arm	Waist	
Not calibrated (%)	69.9	44.8	77.7	44.6	82.1	52.0	70.2	68.8	79.0	63.0	85.7	71.7	65.9
Calibrated (%)	78.3	63.8	85.0	60.2	86.0	69.5	88.0	60.7	90.3	68.0	89.1	72.9	74.8
Improvement (%)	8.4	19.1	7.2	15.6	3.9	17.5	17.8	−8.1	11.3	5.1	3.3	1.2	8.9
Relative reduction in error (%)	27.8	34.5	32.5	28.1	21.8	36.5	59.7	−26.0	53.7	13.6	23.4	4.1	26.1

Table 8
Confusion matrices for *Calibrated* and *Not Calibrated* models with *Arm*, *2-Second*, and *Best Features*.

Model	Actual Class	Predicted Class		
		0	1	2
Calibrated	0 (no risk)	**92.7%**	0.6%	6.8%
	1 (lift/lower/carry)	1.5%	**82.7%**	15.7%
	2 (push/pull)	1.3%	6.6%	**92.1%**
Not calibrated	0 (no risk)	**91.0%**	6.8%	2.3%
	1 (lift/lower/carry)	1.5%	**78.7%**	19.7%
	2 (push/pull)	0.8%	24.6%	**74.5%**

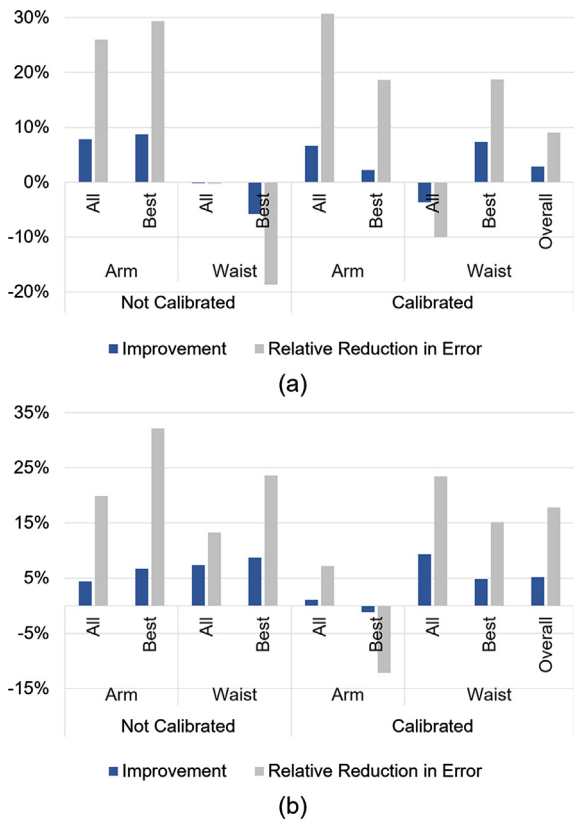

Fig. 7. Improvement in accuracy and relative reduction in error for the (a) *2-Second* models compared to the *1-Second* models and (b) *3-Second* models compared to the *2-Second* models.

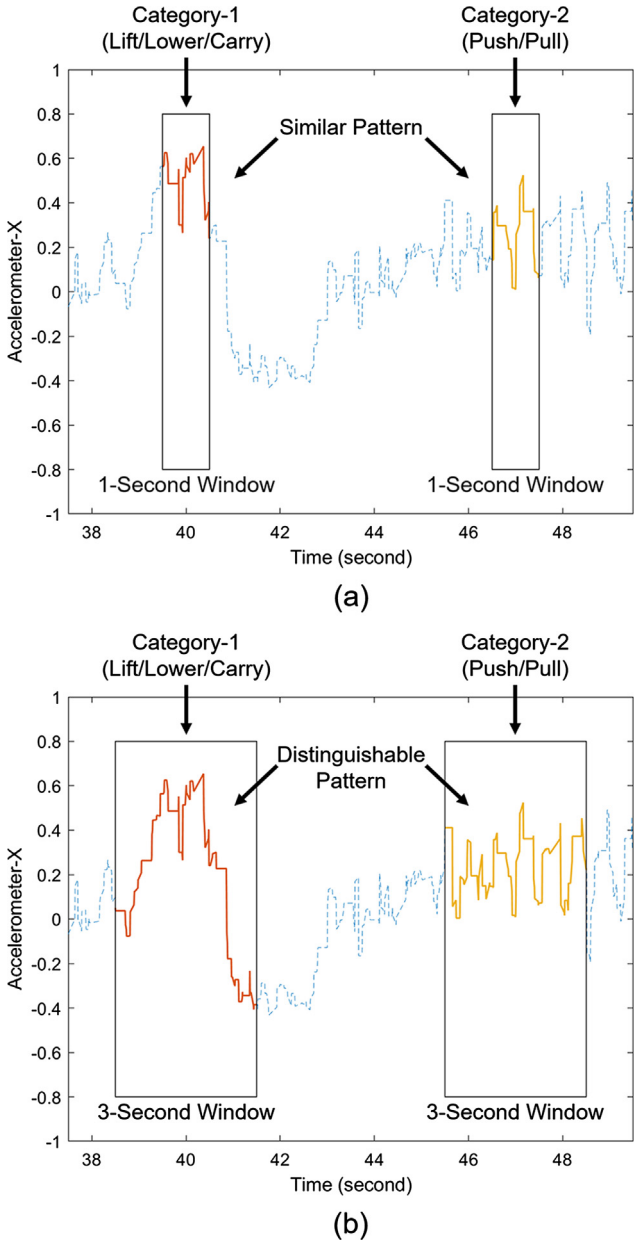

Fig. 8. Patterns in the sensor signals for worker *W1* for (a) *1-second* and (b) *3-second* windows.

accelerometer sensor positively influences the outcome of activity classification.

6.2. Activity recognition

Having found optimal settings for data acquisition and data processing, the classifier model is built and trained using sensor data captured by the arm-mounted smartphone. Data is collected and processed into a uniform time series data. Raw sensor signals are calibrated and segmented into 2-second windows with 50% overlap. Features listed in Table 9 are then extracted and fed into the SVM algorithm to build a

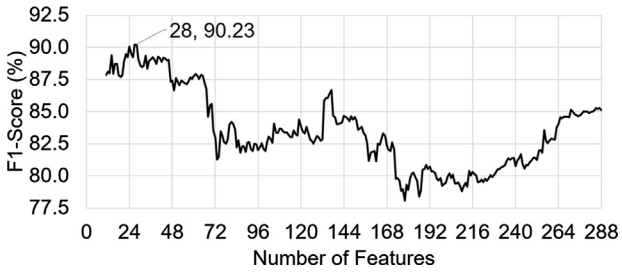

Fig. 9. F1 scores of *Arm*, *Calibrated*, and *2-Second* classifier models with various feature subsets.

classifier model. The model is then applied to the features extracted from the data of another worker to predict class labels (i.e., activities).

6.2.1. Outlier removal

An outlier is defined as an instance having a statistically small number of windows surrounded by a statistically large number of instances of another class. The threshold values for the duration of outliers can be selected by observing activity instances in the training dataset. Here, the threshold window size for the outlier is selected as 2 consecutive windows (or 3 s, given the 50% overlap between adjacent 2-second windows). A 3rd and 5th order median filter is applied consecutively to the predicted labels to remove the outliers. Fig. 11 illustrates a segment of actual activity labels for worker *W2* along with predicted labels before and after outlier removal. It shows that most of the outliers are removed after applying the medial filtering.

Performance metrics of classifier model before and after removing outliers are shown in Table 10. As shown in Table 10, after removing the outliers, all performance metrics (i.e., accuracy, precision, recall, F1, and Spearman's correlation) are improved by at least 1.82% (i.e., relative reduction in error by at least 18.67%).

6.2.2. Confusion matrices

Confusion matrices for *W1* and *W2* are shown in Table 11. The table shows that all activities are recognized with > 80% accuracy, particularly, category-0 and -2 for *W1*, and category-1 and -2 for *W2* are recognized with > 90% accuracy. Also, the tendency of classifiers to misclassify category-1 as category-2 is noticeable for both workers.

6.3. Duration and frequency estimation

Actual and estimated duration and frequency of each category of activity for worker *W1* and *W2* are listed in Table 12 which shows that all estimated durations differ by < 11% from the actual durations. Also, all estimated frequencies are within ± 3 of actual values.

6.4. Ergonomic risk level assessment

Next, corresponding ergonomic risk levels for each worker are calculated based on Table 3. Calculations of risk levels are summarized in Table 13 which shows that all extracted risk levels are identical to the actual risk levels.

Fig. 12 illustrates the actual and estimated duration per shift in percentage and the corresponding ergonomic risk levels. It can be seen that the difference between actual and estimated duration per shift is negligible compared to the difference between duration per shift for two adjacent risk levels. For example, all duration per shift differs by < 2% (Table 13). However, from low to moderate level of risk and from moderate to high level of risk, values for duration per shift differ by 25% (Table 3). Therefore, all the actual and the corresponding estimated risk falls into the same level of risk. A similar argument is also valid for the risks related to frequency per minute.

7. Summary and conclusions

In this paper, a methodology for the smart monitoring of construction activities for the purpose of ergonomic risk assessment was presented. The designed approach used wearable IMUs (built-in smartphone sensors) for time motion data collection. An experiment was carried out to test the robustness and reliability of the methodology. The primary contribution of this research is the development of a framework for automatically identifying instances, durations, and

Table 9
Best features for worker *W1* and *W2* (Features marked in italic are selected for both workers).

Rank	W1	W2
1	*Accelerometer_X_Minimum*	*Gyroscope_Magnitude_Maximum*
2	*Accelerometer_X_Mean*	*Accelerometer_X_Mean*
3	*Gyroscope_Z_Minimum*	*Gyroscope_Magnitude_StandardDeviation*
4	*Gyroscope_Magnitude_StandardDeviation*	*Gyroscope_Z_Minimum*
5	*Gyroscope_Magnitude_Maximum*	*Accelerometer_X_Maximum*
6	*Gyroscope_Y_Minimum*	Gyroscope_X_Maximum
7	Accelerometer_Y_Mean	*Accelerometer_Y_Minimum*
8	*Gyroscope_Magnitude_Mean*	*Gyroscope_Z_Maximum*
9	*Accelerometer_Z_StandardDeviation*	*Accelerometer_Magnitude_Mean*
10	*Gyroscope_X_Minimum*	*Gyroscope_Y_Maximum*
11	*Accelerometer_X_StandardDeviation*	Accelerometer_Z_Mean
12	*Gyroscope_Magnitude_IQR*	Accelerometer_Z_Minimum
13	*Gyroscope_Magnitude_MeanAbsoluteDeviation*	*Accelerometer_X_Minimum*
14	Gyroscope_Magnitude_Minimum	*Gyroscope_Y_Minimum*
15	*Accelerometer_Magnitude_StandardDeviation*	*Accelerometer_X_StandardDeviation*
16	*Accelerometer_X_Maximum*	*Gyroscope_Magnitude_IQR*
17	Gyroscope_Y_StandardDeviation	*Gyroscope_Z_StandardDeviation*
18	*Accelerometer_Z_Maximum*	*Gyroscope_X_Minimum*
19	*Gyroscope_Y_Maximum*	*Gyroscope_Magnitude_Mean*
20	*Accelerometer_Magnitude_Mean*	Accelerometer_Z_IQR
21	*Gyroscope_Z_Maximum*	Accelerometer_X_MeanAbsoluteDeviation
22	Gyroscope_Y_IQR	*Accelerometer_Magnitude_StandardDeviation*
23	Gyroscope_Y_Mean	Gyroscope_X_Mean
24	*Gyroscope_Z_StandardDeviation*	*Gyroscope_Magnitude_MeanAbsoluteDeviation*
25	Accelerometer_Magnitude_Maximum	*Accelerometer_X_Jerk_Mean*
26	LinearAccelerometer_Z_StandardDeviation	*Accelerometer_Z_Maximum*
27	*Accelerometer_X_Jerk_Mean*	Accelerometer_Magnitude_IQR
28	*Accelerometer_Y_Minimum*	*Accelerometer_Z_StandardDeviation*

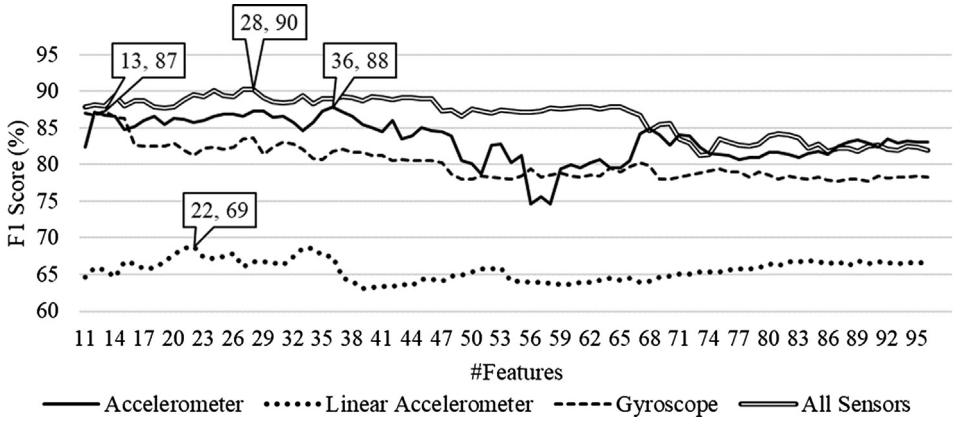

Fig. 10. F1 scores of classification (for arm-mounted phone, calibrated data, 2-second window size combination) with various feature subsets from accelerometer, linear accelerometer, gyroscope, and all three sensors.

frequencies of certain classes of human activities such as push/pull and lift/lower/carry in a work shift, and estimating corresponding levels of overexertion and ergonomic risks. It must be noted that while these activities are common across many occupations, they are inherently more complex compared to activities traditionally examined in human activity recognition studies (e.g., walking, sitting). A rigorous investigation of the optimal data collection settings and experimental parameters (i.e., smartphone's position, calibration, window size, and feature set) for reliably classifying human activities of interest was also performed. It was found that the arm-mounted smartphone provided more distinguishable patterns of signal than the waist-mounted device, which was verified by observing lower classifier's confusion between similar activities. Moreover, calibrating the sensor signals (i.e., subtracting sensor's benchmark readings when the subject is idle from all other sensor readings in non-idle times) improves classification results. Also, segmenting data into 2-second windows generates better classification results. Finally, it was found that using the best feature subset instead of all features increases the classification accuracy.

In the experiment conducted in this research, activities were video recorded and used as ground truth to evaluate the performance of the designed technique. It was found that estimated (predicted) ergonomics risk levels were identical to those observed (ground truth). Therefore, the designed approach has a great potential to replace manual observations that are time-consuming, interruptive, subjective, and require physical presence on the location. Moreover, it eliminates the necessity of implementing sophisticated sensor network which requires a significant amount of time and technical knowledge to set up, operate, and maintain. Furthermore, this research illustrated a detailed investigation on optimizing various settings to prepare datasets for cross-subject classification. The findings could be valuable for real-world implementations where it is not possible to collect adequate training samples from all subjects, rather training on one subject and testing on other subject(s) is more feasible. Moreover, the proposed approach contributes to the PtD practice by identifying the major sources of ergonomic risks that need to be eliminated. For example, in the experiment described in this paper, workers were found to be at high risk due to the long duration of category-2 (push/pull) activities. Therefore, the authority (superintendent, project manager, or safety inspector) can be alerted to prevent this risk by redesigning the workplace, task, tool, or environment so that the duration of these activities decreases to < 50% of the work shift (see OSHA threshold values in Table 3). Clearly, the manner by which such changes are implemented is expected to be different from one project to the next, as the nature of each project and the types of activities performed are distinct.

While this work demonstrated an integrated approach to using wearable technology (i.e., smartphone's built-in sensors) and machine

Table 10
Performance metrics of classifier model before and after removing outlier.

Outlier	Accuracy	Precision	Recall	F1	Spearman's Correlation
Not Removed	90.25%	90.29%	90.25%	90.27%	85.39%
Removed	92.07%	92.14%	92.07%	92.10%	88.34%
Improvement	1.82%	1.85%	1.82%	1.83%	2.95%

Table 11
Confusion matrices for *W1* and *W2*.

Worker	Actual Class	Predicted Class		
		0	1	2
W1	0 (no risk)	**99.6%**	0.4%	0.0%
	1 (lift/lower/carry)	1.1%	**82.2%**	16.7%
	2 (push/pull)	2.4%	4.3%	**93.3%**
W2	0 (no risk)	**88.0%**	3.3%	8.6%
	1 (lift/lower/carry)	0.0%	**91.6%**	8.4%
	2 (push/pull)	0.0%	5.9%	**94.1%**

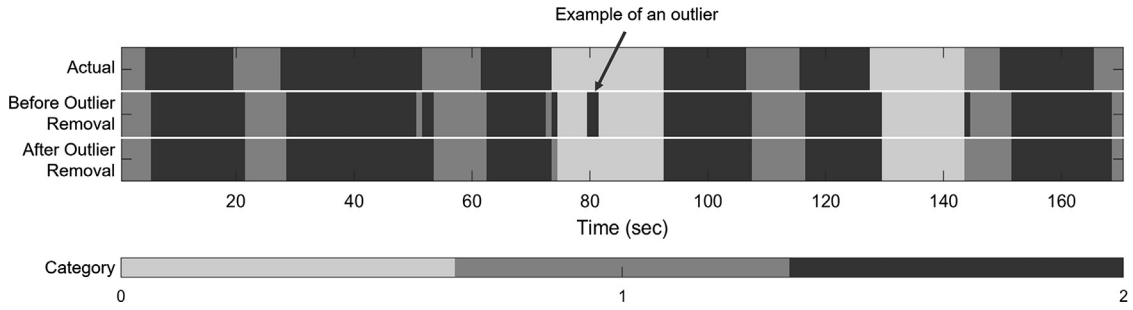

Fig. 11. Actual labels, predicted labels before outlier removal, and predicted labels after outlier removal for worker *W2*.

Table 12
Actual and estimated duration and frequency.

Worker	Cat.	Duration			Actual	Frequency	
		Actual (sec)	Estimated (sec)	Error		Estimated	Error
W1	0	246	265	−7.7%	15	15	0
	1	313	300	4.2%	37	40	−3
	2	755	754	0.1%	41	42	−1
W2	0	316	282	10.8%	15	17	−2
	1	286	310	−8.4%	35	35	0
	2	740	748	−1.1%	40	42	−2

Table 13
Calculation of ergonomic risk levels.

ID	Cat.	Duration/Shift			Risk	Frequency (per minute)			Risk
		Actual	Estimated	Diff.	Level	Actual	Estimated	Diff.	Level
W1	1	24%	23%	1%	L	1.69	1.82	0.13	M
	2	57%	57%	0%	H	1.87	1.91	0.04	M
W2	1	21%	23%	2%	L	1.56	1.57	0.01	M
	2	55%	56%	1%	H	1.79	1.88	0.09	M

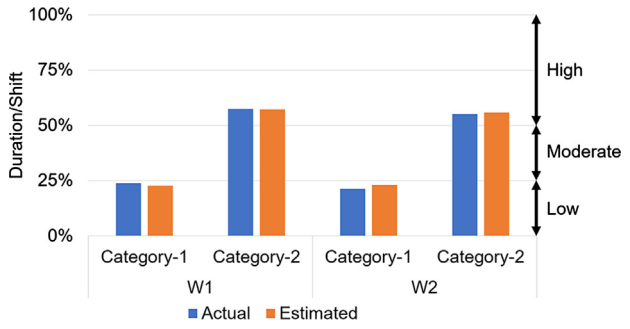

Fig. 12. Actual and estimated duration per shift and corresponding ergonomic risk levels.

learning for ergonomic-related data collection and analysis in an experimental setting, findings can be generalized and inform similar efforts in various occupations including construction, manufacturing, healthcare, transportation, and agriculture. Future work in this research will focus on investigating ways to capture the intensity factor (in addition to duration and frequency information) from field activities. Intensity can be a major cause of WMSDs resulting from activities such as lifting, carrying, throwing, pushing, and pulling. However, existing literature on this topic all point to a key limitation in sensing technologies that hinders the ability to obtain weight information (directly or indirectly) from IMU sensors mounted on people (and not objects). By mounting sensors on objects, the cost of the developed system will significantly rise, not to mention the inevitable challenges resulting from the fact that individual sensors need to be installed, calibrated, synced, and maintained one by one for each object. A more feasible approach would be to centralize the sensing process by relying on worker-mounted sensors that can significantly bring down the computation time and complexity, especially when dealing with heterogeneous data from multiple sensors. Having said that, the authors plan to investigate whether a mathematical correlation exists between signal patterns obtained from smartphone sensors (e.g., accelerometer, gyroscope, and linear accelerometer data), and the weight of objects handled. Together, duration, frequency, and intensity information will enable a more comprehensive and meaningful analysis of ergonomic hazards associated with overexertion. Also, as part of the future work, authors will explore the fusion of decisions that results from multiple windows as well as to dynamically change the window size based on the data of interest [48,49]. Moreover, while parameter optimization was performed on a relatively small number of participants in the experiments conducted in this research, a more robust optimization can be performed by collecting training samples from people of different ages, heights, genders, experience levels, and other characteristics. Finally, similar techniques can be used to automatically assess WMSDs, quantify their severity, and ultimately track the progress of corresponding treatments. To this end, the authors plan to explore the feasibility of such unobtrusive ambulatory systems and automated frameworks in clinically-relevant WMSDs applications. It is also worth mentioning that with the proliferation of wearable technology, the authors expect a paradigm shift in physically-demanding industries such as construction and manufacturing, where ergonomic risks and bodily injuries can be more precisely predicted by augmenting data from bodily motions (similar to this study) with (for instance) physiological data such as blood volume pulse (BVP), electrodermal activity (sweat), electrocardiogram (ECG), respiration, electroencephalogram (EEG) signals [50,51].

Acknowledgments

The presented work has been supported by the U.S. National Science Foundation (NSF) through grant CMMI 1800957. The authors gratefully acknowledge the support from the NSF. Any opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily represent those of the NSF.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.aei.2018.08.020.

References

[1] H.G. Kang, D.F. Mahoney, H. Hoenig, V.A. Hirth, P. Bonato, I. Hajjar, L.A. Lipsitz, In situ monitoring of health in older adults: technologies and issues, J. Am. Geriatr. Soc. 58 (8) (2010) 1579–1586, https://doi.org/10.1111/j.1532-5415.2010.02959.x.

[2] S.L. Lau, I. König, K. David, B. Parandian, C. Carius-Düssel, M. Schultz, Supporting patient monitoring using activity recognition with a smartphone, 2010 7th International Symposium on Wireless Communication Systems (ISWCS) (2010) 810–814.

[3] M.J. Mathie, A.C. Coster, N.H. Lovell, B.G. Celler, Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement, Physiol. Measur. 25 (2) (2004) R1–R20.

[4] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, P. Havinga, Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: a survey, 2010 23rd International Conference on Architecture of Computing Systems (ARCS) (2010) 1–10.

[5] N.D. Lane, M. Mohammod, M. Lin, X. Yang, H. Lu, S. Ali, A. Doryab, E. Berke, T. Choudhury, A. Campbell, Bewell: A smartphone application to monitor, model and promote wellbeing, in: 5th International ICST Conference on Pervasive Computing Technologies for Healthcare, 2011, pp 23–26.

[6] D.D. Luxton, R.A. McCann, N.E. Bush, M.C. Mishkind, G.M. Reger, mHealth for mental health: integrating smartphone technology in behavioral healthcare, Profess. Psychol.: Res. Practice 42 (6) (2011) 505–512.

[7] A.C. Timmons, T. Chaspari, S.C. Han, L. Perrone, S.S. Narayanan, G. Margolin, Using multimodal wearable technology to detect conflict among couples, Computer 50 (3) (2017) 50–59.

[8] K. Yang, C.R. Ahn, M.C. Vuran, S.S. Aria, Semi-supervised near-miss fall detection for ironworkers with a wearable inertial measurement unit, Autom. Construct. 68 (2016) 194–202.

[9] A. Alwasel, K. Elrayes, E.M. Abdel-Rahman, C. Haas, Sensing construction work-related musculoskeletal disorders (WMSDs), in: 28th International Symposium on Automation and Robotics in Construction, Seoul, Korea, 2011, pp 164–169.

[10] L. Peppoloni, A. Filippeschi, E. Ruffaldi, C. Avizzano, A novel wearable system for the online assessment of risk for biomechanical load in repetitive efforts, Int. J. Ind. Ergon. 52 (2016) 1–11.

[11] N.D. Nath, R. Akhavian, A.H. Behzadan, Ergonomic analysis of construction worker's body postures using wearable mobile sensors, Appl. Ergon. 62 (2017) 107–117.

[12] X. Yan, H. Li, A.R. Li, H. Zhang, Wearable IMU-based real-time motion warning system for construction workers' musculoskeletal disorders prevention, Autom. Constr. 74 (2017) 2–11.

[13] P. Buckle, Ergonomics and musculoskeletal disorders: overview, Occup. Med. 55 (3) (2005) 164–167.

[14] A. Luttmann, M. Jäger, B. Griefahn, G. Caffier, F. Liebers, Preventing Musculoskeletal Disorders in the Workplace vol. 5, World Health Organization, Geneva, Switzerland, 2003.

[15] OSHA, Ergonomics: The Study of Work, 2000.

[16] OHCOW, Work-related musculoskeletal disorders (WMSDs), 2005.

[17] S. Simoneau, M. St-Vincent, D. Chicoine, Work-Related Musculoskeletal Disorders (WMSDs): A Better Understanding for more Effective Prevention, IRSST, Québec, 1996.

[18] L.M. Group, 2011 Liberty Mutual Workplace Safety Index, 2011.

[19] BLS, Nonfatal Occupational Injuries and Illnesses Requiring Days Away from Work, 2016.

[20] BLS, Nonfatal Occupational Injuries and Illnesses Requiring Days Away from Work, 2014.

[21] M. Goldenhar, L.J. Williams, N.G. Swanson, Modelling relationships between job stressors and injury and near-miss outcomes for construction labourers, Work & Stress 17 (3) (2003) 218–240.

[22] NIOSH, The State of the National Initiative on Prevention Through Design, 2014.

[23] N. Jaffar, A.H. Abdul-Tharim, I.F. Mohd-Kamar, N.S. Lop, A literature review of ergonomics risk factors in construction industry, Proc. Eng. 20 (2011) 89–97.

[24] G.C. David, Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders, Occup. Med. 55 (3) (2005) 190–199.

[25] NIOSH, Musculoskeletal Disorders and Workplace Factors, 1997.

[26] I. Balogh, P. Ørbæk, K. Ohlsson, C. Nordander, J. Unge, J. Winkel, G.Å. Hansson, Self-assessed and directly measured occupational physical activities—influence of musculoskeletal complaints, age and gender, Appl. Ergon. 35 (1) (2004) 49–56.

[27] J. Chen, C.R. Ahn, S. Han, Detecting the hazards of lifting and carrying in construction through a coupled 3D sensing and IMUs sensing system, Computing in Civil and Building Engineering, Orlando, Florida (2014) 1110–1117.

[28] L. Chen, I. Khalil, Activity recognition: approaches, practices and trends, Activity Recognition in Pervasive Intelligent Environments (2011) 1–31.

[29] J. Winkel, S.E. Mathiassen, Assessment of physical work load in epidemiologic studies: concepts, issues and operational considerations, Ergonomics 37 (6) (1994) 979–988.

[30] R. Akhavian, A.H. Behzadan, Smartphone-based construction workers' activity recognition and classification, Autom. Constr. 71 (2016) 198–209.

[31] D. Anguita, A. Ghio, L. Oneto, X. Parra, J.L. Reyes-Ortiz, Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine, International Workshop on Ambient Assisted Living (2012) 216–223.

[32] X. Su, H. Tong, P. Ji, Activity recognition with smartphone sensors, Tsinghua Sci. Technol. 19 (3) (2014) 235–249.

[33] N.D. Nath, P. Shrestha, A.H. Behzadan, Human activity recognition and mobile sensing for construction simulation, in: W.K.V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, E. Page, (Eds.), 2017 Winter Simulation Conference, Las Vegas, NV, 2017, pp 2448–2459.

[34] A. Bayat, M. Pomplun, D.A. Tran, A study on human activity recognition using accelerometer data from smartphones, Proc. Comput. Sci. 34 (2014) 450–457.

[35] N. Ravi, N. Dandekar, P. Mysore, M.L. Littman, Activity recognition from accelerometer data, AAAI (2005) 1541–1546.

[36] A.M. Khan, Y.K. Lee, S.Y. Lee, T.S. Kim, Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis, 2010 5th International Conference on Future Information Technology (FutureTech) (2010) 1–6.

[37] X. Long, B. Yin, R.M. Aarts, Single-accelerometer-based daily physical activity classification, Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN (2009) 6107–6110 https://doi.org/10.1109/IEMBS.2009.5334925.

[38] A. Mannini, A.M. Sabatini, Machine learning methods for classifying human physical activity from on-body accelerometers, Sensors 10 (2) (2010) 1154–1175.

[39] L. Joshua, K. Varghese, Accelerometer-based activity recognition in construction, J. Comput. Civil Eng. 25 (5) (2010) 370–379.

[40] J. Ryu, J. Seo, M. Liu, S. Lee, Haas CT Action recognition using a wristband-type activity tracker: case study of masonry work, Construct. Res. Congr. (2016) 790–799.

[41] K. Kunze, P. Lukowicz, Sensor placement variations in wearable activity recognition, IEEE Pervasive Comput. 13 (4) (2014) 32–41.

[42] U.O. Massachusetts-Lowell, Ergonomics for Trainers, OSHA, 2012.

[43] C.-C. Yang, Y.-L. Hsu, A review of accelerometry-based wearable motion detectors for physical activity monitoring, Sensors 10 (8) (2010) 7772–7788.

[44] R. Akhavian, A.H. Behzadan, Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers, Adv. Eng. Inf. 29 (4) (2015) 867–877.

[45] D. Anguita, A. Ghio, L. Oneto, X. Parra, J.L. Reyes-Ortiz, Energy efficient smartphone-based activity recognition using fixed-point arithmetic, J. UCS 19 (9) (2013) 1295–1314.

[46] I. Kononenko, Estimating attributes: Analysis and extensions of RELIEF, Machine Learning, Springer, Heidelberg, Germany, 1994, pp. 171–182.

[47] N. Kwak, C.-H. Choi, Input feature selection for classification problems, IEEE Trans. Neural Networks 13 (1) (2002) 143–159.

[48] Q. Ni, T. Patterson, I. Cleland, C. Nugent, Dynamic detection of window starting positions and its implementation within an activity recognition framework, J. Biomed. Inform. 62 (2016) 171–180.

[49] O. Banos, J.-M. Galvez, M. Damas, A. Guillen, L.-J. Herrera, H. Pomares, I. Rojas, C. Villalonga, C.S. Hong, S. Lee, Multiwindow fusion for wearable activity recognition, International Work-Conference on Artificial Neural Networks, Springer, 2015, pp. 290–297.

[50] T. Cheng, G.C. Migliaccio, J. Teizer, U.C. Gatti, Data fusion of real-time location sensing and physiological status monitoring for ergonomics analysis of construction workers, J. Comput. Civil Eng. 27 (3) (2012) 320–335.

[51] S. Hwang, S. Lee, Wristband-type wearable health devices to measure construction workers' physical demands, Autom. Constr. 83 (2017) 330–340.