Expert Systems With Applications 155 (2020) 113405

Contents lists available at ScienceDirect

Expert Systems With Applications

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

A data analytic framework for physical fatigue management using wearable sensors

Zahra Sedighi Maman[a], Ying-Ju Chen[b], Amir Baghdadi[c,d], Seamus Lombardo[e,f], Lora A. Cavuoto[g], Fadel M. Megahed[h,*]

[a] Department of Decision Sciences and Marketing, Adelphi University, Garden City, NY 11530, USA
[b] Department of Mathematics, University of Dayton, Dayton, OH 45469, USA
[c] Department of Mechanical and Aerospace Engineering, University at Buffalo, Buffalo, NY 14260, USA
[d] Department of Clinical Neurosciences, University of Calgary, Alberta T2N 1N4, Canada
[e] Department of Mechanical and Aerospace Engineering, University at Buffalo, Buffalo, NY 14260, USA
[f] Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[g] Department of Industrial & Systems Engineering, University at Buffalo, Buffalo, NY 14260, USA
[h] Farmer School of Business, Miami University, OH 45056, USA

A R T I C L E I N F O

Article history:
Received 23 February 2019
Revised 1 March 2020
Accepted 23 March 2020
Available online 18 April 2020

Keywords:
Functional data analysis
Human performance modeling
Internet of Things (IoT)
Manufacturing
Occupational safety

A B S T R A C T

The use of expert systems in optimizing and transforming human performance has been limited in practice due to the lack of understanding of how an individual's performance deteriorates with fatigue accumulation, which can vary based on both the worker and the workplace conditions. As a first step toward realizing the human-centered approach to artificial intelligence and expert systems, this paper lays the foundation for a data analytic approach to managing fatigue in physically-demanding workplaces. The proposed framework capitalizes on continuously collected human performance data from wearable sensor technologies, and is centered around four distinct phases of fatigue: (a) detection, where machine learning methodologies are deployed to detect the occurrence of fatigue; (b) identification, where key features relating to the fatigue occurrence is to be identified; (c) diagnosis, where the fatigue mode is identified based on the knowledge generated in the previous two phases; and (d) recovery, where a suitable intervention is applied to return the worker to mitigate the detrimental effects of fatigue on the worker. Moreover, the framework establishes criteria for feature and machine learning algorithm selection for fatigue management. Two specific application cases of the framework, for two types of manufacturing-related tasks, are presented. Based on the proposed framework and a large number of test sets used in the two case studies, we have shown that: (i) only one wearable sensor is needed for fatigue detection with an average accuracy of ≥ 0.850 and a random forest model comprised of < 7 features; and (ii) the selected features are task-dependent, and thus capturing different modes of fatigue. Therefore, this research presents an important foundation for future expert systems that attempt to quantify/predict changes in workers' performance as an input to prescriptive rest-break scheduling, job-rotation, and task assignment models. To encourage future work in this important area, we provide links to our data and code as *Supplementary materials*.

1. Introduction

The advancements in automation, computation, information, sensing, and expert systems are changing the landscape of jobs and workplaces at unprecedented speeds (National Science Foundation, 2019). One of the main, and currently observed, consequences is an increase in automation, which has resulted in an increased adoption of: (a) robotic systems in manufacturing and warehousing operations (The White House, 2016; Wang, Jiang, Lee, Chew, & Tan, 2017), (b) virtual assistants (Eisman, Navarro, & Castro, 2016; Montenegro, da Costa, & da Rosa Righi, 2019), and (c) expert systems for job scheduling and task optimization (Dhurasevic & Jakobovic, 2018). Despite the undeniable fact of some job loss associated with automation, a hallmark feature of this new era (often referred to as *Industry 4.0*) is its dependence on highly-skilled labor

* Corresponding author.
E-mail addresses: zmaman@adelphi.edu (Z. Sedighi Maman), ychen4@udayton.edu (Y.-J. Chen), amirbagh@buffalo.edu (A. Baghdadi), seamuslo@mit.edu (S. Lombardo), loracavu@buffalo.edu (L.A. Cavuoto), fmegahed@miamioh.edu (F.M. Megahed).

https://doi.org/10.1016/j.eswa.2020.113405

(Ferjani, Ammar, Pierreval, & Elkosantini, 2017) who can capitalize on the technological revolution. The industry is moving towards a future that will be defined by how it optimizes its three main resources (Daugherty & Wilson, 2018; Kong, Luo, Huang, & Yang, 2018; Pacaux-Lemoine, Berdal, Enjalbert, & Trentesaux, 2018): *human workers, machines*, and *supporting technologies* (e.g., artificial intelligence, high performance computing, and expert systems).

The current inability to jointly optimize the aforementioned three resources stems from the lack of reliable and individualized models that can quantify the effects of job tasks on a worker's performance (Cavuoto & Megahed, 2017; Maman, Yazdi, Cavuoto, & Megahed, 2017). Current applications where expert systems excel in learning from human behavior to automate/optimize the decision-making process (see Saraiva et al., 2016; Weng, Ahmed, and Megahed, 2017; Weng, Lu, Wang, Megahed, and Martinez, 2018, for some recent applications); however, they do not model the impact of *automation* and *supporting technologies* on human performance. This observation is supported by Grosse, Glock, and Neumann (2017), who observed that "*human characteristics that are often a major determinant of system performance have, however, widely been ignored in this stream of research*". The problem is challenging since: (a) humans' performance changes as a function of a person's individual characteristics (e.g., age, sex, injury history, etc.), time (which can be manifested through detrimental performance due to fatigue and/or improved performance due to learning effects) and degree of task difficulty (Maman, Baghdadi, Megahed, & Cavuoto, 2016; Maman et al., 2017); (b) the literature capturing human performance in occupational settings have typically relied on surveys (Lu, Megahed, Sesek, & Cavuoto, 2017) and thus, our understanding of how an individual's performance change over the course of their day/work-shift is limited (Baghdadi et al., 2019); and (c) there is a disconnect between predictive and prescriptive models that attempt to model workplace fatigue (Lu, Megahed, & Cavuoto, 2019b; Maman, Lu, Megahed, & Cavuoto, 2019). For these reasons, many researchers and practitioners consider "human-in-the-loop" modeling to be the next frontier in artificial intelligence/expert systems research (Bavaresco, D'Oca, Ghisi, & Lamberts, 2019; Oneto, Navarin, Donini, & Anguita, 2018; Rea, 2018; Zanzotto, 2019).

As a first step toward "human-in-the-loop" modeling, this paper proposes a framework that can be used to detect and explain deterioration in an individual's work performance as a result of physical fatigue. We focus on fatigue since it is a precursor to many detrimental short-term and long-term health outcomes (Cavuoto & Megahed, 2017). Furthermore, we have chosen to focus on *advanced manufacturing* tasks since: (a) changes in work performance is task/field dependent; and (b) advanced manufacturing jobs are highly fatiguing despite the increased prevalence of automation (Kajimoto, 2008; Loriol, 2017; Lu et al., 2017; Yung, 2016). The high prevalence of fatigue at manufacturing workplaces can be explained by the transformations in labor roles, where the following changes have been observed: (a) a reduction in mundane tasks (Yakowicz, 2016), (b) an increased dependency on highly-trained workers (Ferjani et al., 2017), (c) an increase in worker's autonomy and responsibility (Waldeck, 2014), and (d) the introduction of new job duties (Waldeck, 2014).

A framework is proposed instead of a model to allow for the detection/diagnosis of multiple fatigue modes. The main premise is that advanced manufacturing firms require specialized labor (Ferjani et al., 2017; Lu et al., 2017). Thus, the jobs can then be grouped by the type of activities. This is reasonable since the main tasks performed by a CNC, computer numerical control, machinist are different from those done by a welder. The proposed framework is made of four phases: (a) *detection*, where the goal is to detect if/when a worker has become fatigued, (b) *identification*, where the most important variables for diagnosing fatigue are identified, (c) *diagnosis*, where the information captured from phases (a) and (b) is used to pinpoint the fatigue mode, and (d) *recovery*, where a suitable intervention is applied to return to a non-fatigued state. The phases are adapted from the structured methodology used by quality engineers for fault detection and diagnosis (Chiang, Russell, & Braatz, 2000). Note that none of the existing quantitative approaches for fatigue modeling present information on the *identification, diagnosis and recovery* stages needed for managing fatigue.

Our framework capitalizes on the advances and widespread use of *wearable sensors* for the purposes of data collection. There are three important justifications for the use of *wearable sensors* in our framework. First, based on a survey of U.S. manufacturing safety professionals, 54.1% of the respondents were "in favor of using wearable technologies at work to track [occupational safety and health] risk factors" (Schall, Sesek, & Cavuoto, 2018). From the responses, Schall et al. (2018) estimated that U.S. manufacturing firms would spend, on average, an estimated $68.67 per worker for a wearable device. Second, the use of *wearables* presents a unified benchmark of performance that does not depend on the cycle time of the process. The third, and perhaps the most important reason, *wearables* present an individualized view of the performance of the worker. Unlike other outcomes, e.g., work quality which may be affected by upstream performances.

The remainder of the paper is organized as follows. In Section 2, an overview of the relevant literature on fatigue management in manufacturing environments is presented. Our proposed framework for detecting, identifying and diagnosing fatigue root-causes is discussed in Section 3. In Section 4 two case studies are investigated to evaluate the utility of the framework in managing fatigue during two manufacturing tasks. Our concluding remarks and future research suggestions are presented in Section 5. We offer our code and data as *supplementary materials* to encourage adoption in practice and further investigations by researchers.

2. Background and literature review

2.1. Fatigue implications

Managing fatigued workers is an important issue with ethical, operational and financial considerations. Ethically, fatigue is a precursor to many detrimental short-term and long-term health outcomes. The short-term effects include discomfort, lowered strength and a diminished motor control function (Yung et al., 2017). In an operational environment, those short-term effects lead to "reduced performance, productivity, quality of work and increased incidence of labour accidents and human errors" (Yung, Bigelow, Hastings, & Wells, 2014, p. 1562). The long-term health consequences of fatigue include: (a) a high prevalence of musculoskeletal disorders (Naranjo-Flores & Ramírez-Cárdenas, 2014), (b) suffering from chronic-fatigue syndrome (Fukuda et al., 1994), and (c) a weakened immune function (Kajimoto, 2008). From an operational perspective, Ricci, Chee, Lorandeau, and Berger (2007) reported that the health-related lost productivity time for fatigued workers exceeds double their non-fatigued counterparts. The financial ramifications of fatigue outcomes are estimated to cost U.S. employers approximately $136 billion annually (Ricci et al., 2007).

2.2. Data collection mechanisms

An important first step in managing fatigue is the rapid and accurate detection of its occurrence. Fatigue detection techniques can be divided into two categories: qualitative and quantitative. Qualitative methods are centered around the use of fatigue surveys (Lu et al., 2017). From a practical perspective, the utility of such methods is limited to investigations aiming to assess workloads

Table 1
A summary of the two major research streams of fatigue modeling.

Category	Paper	Tasks	Sensors	Method
Exhaustion	Kavanagh, Morrison, and Barrett (2006)	Walking	EMG	Statistical test
	Karg et al. (2008)	Walking	3D optical tracking	LDA, SVM, KNN, NB
	Zhang et al. (2014)	Walking	IMUs	SVM
	Ebenbichler et al. (2002)	Lifting	EMG	Time frequency analysis
	Bonato et al. (2003)	Lifting	EMG	Statistical test
	Chow et al. (2004)	Lifting	EMG	Statistical test
	Karg et al. (2014)	Squat	Infrared cameras	Linear regression, HMM
Occ. fatigue	Yoshino et al. (2004)	Walking	EMG	Linear regression
	Helbostad et al. (2007)	Walking	Accelerometer	Statistical test
	Lee et al. (2009)	Walking	Reflective markers	Statistical test, LDA
	Baghdadi et al. (2018b)	Material handling	IMUs	SVM
	Maman et al. (2017)	Material handling, supply insertion & pickup, part assembly	IMUs, HR	Penalized logistic regression

where SVM=support vector machines, HMM=hidden markov models, LDA=linear discriminant analysis, kNN= k-nearest neighbors, & NB=naive bayes.

and/or redesign jobs. However, they are not suitable for real-time, shop-floor-wide fatigue detection, since they are not scalable and are potentially disruptive. For example, consider a situation where there are 70 workers on the shop-floor and their fatigue ratings are measured every 5 min. The administration of surveys in this situation would require a large number of surveyors, and would disrupt production (reducing the productivity of workers (Cai, Gong, Lu, & Zhong, 2018)).

The quantitative approaches, of the second category, rely on using one or more sensor technologies to model changes in human performance. The utilized sensor technologies include: (a) heart rate sensors to measure heart-rates, which are indicative of whole-body fatigue (Maman et al., 2017); (b) inertial measurement units (IMUs), which are cheap and reliable sensors that are used to capture a person's acceleration and motion data (Baghdadi, Cavuoto, & Crassidis, 2018a; Baghdadi, Megahed, Esfahani, & Cavuoto, 2018b; Maman et al., 2017); (c) electroencephalography (EEG), used to measure brain activity, which is important in detecting mental fatigue (Charbonnier, Roy, Bonnet, & Campagne, 2016; Moon, Kwon, Park, & Yoon, 2019; Zhao, Zheng, Zhao, Tu, & Liu, 2011); (d) electromyography (EMG), used to assess muscle activity and localized fatigue (Kumar & Mital, 2017; Venugopal, Navaneethakrishna, & Ramakrishnan, 2014); and (e) optical sensors, which can be used to detect sleepiness or can be utilized for motion capture (Iskander, Hossny, & Nahavandi, 2018; Koesdwiady, Soua, Karray, & Kamel, 2017). Note that some of these technologies are not suitable for daily field implementation. Specifically, EEG and EMG are invasive (Cavuoto & Megahed, 2017), which inhibits their daily usage for real-time fatigue detection. Moreover, motion capture systems often require special setups, which make them better suited for controlled environments. For these reasons, the EEG, EMG and motion capture sensors will not be further discussed. Hereafter, the phrase *wearable sensors* is used to denote a system made of one or more IMUs and a heart rate monitor.

Despite the popularity of wearable sensors in personal physical activity monitoring (e.g., Fitbit, Garmin and Jawbone trackers), workplace fatigue monitoring applications has been limited to three domains (Cavuoto & Megahed, 2017; Maman et al., 2017). These are athletics, transportation and mining. The main barrier, in other disciplines, is a lack of standardization of work activities across employees, which results in multiple modes of fatigue (e.g., different muscles or whole-body fatigue). This is different from the three domains where the technology is tailored to target a known and dominant fatigue mode. It is, therefore, difficult to develop a global model to accurately detect different fatigue modes outside of the three disciplines.

2.3. An overview and taxonomy of the physical fatigue detection literature

The literature on physical fatigue detection in manufacturing environments can be classified into: (a) exhaustion detection, and (b) occupational fatigue detection. In the first group, studies attempt to identify extreme fatigue, i.e. exhaustion, which results in an inability to generate muscle forces and consequently, a worker's inability to perform the job (Ceschi, Demerouti, Sartori, & Weller, 2017). Since exhaustion in the manufacturing workplace is often on the muscle level (localized fatigue), the associated literatures (Baghdadi et al., 2018b; Bonato et al., 2003; Chow, Man, Holmes, & Evans, 2004; Davidson, Madigan, & Nussbaum, 2004; Ebenbichler et al., 2002; Fontes et al., 2010; Karg, Venture, Hoey, & Kulic, 2014; Lee, Roan, Smith, & Lockhart, 2009; Yoshino, Motoshige, Araki, & Matsuoka, 2004; Zhang, Lockhart, & Soangra, 2014) is characterized by: (i) primarily utilizing invasive EMG and EEG sensors, (ii) focusing on one task element only (e.g., lifting or walking), and (iii) no attempt to generalize the developed models to focus on a more complex task. In the second group, the studies focused on detecting occupational fatigue, which is less extreme than exhaustion, where the workers are still able to perform their job at a diminished level. Those studies, e.g. Baghdadi et al. (2018b); Helbostad, Leirfall, Moe-Nilssen, and Sletvold (2007); Lee et al. (2009); Yoshino et al. (2004), have often utilized pervasive sensors including IMUs and heart rate monitors. In addition, recently, Maman et al. (2017) has developed a generalized model for detecting fatigue across multiple manufacturing tasks. However, their model involved over 20 predictors and lacked the interpretability that makes it effective for the consequent phases of fatigue identification, diagnosis and recovery. Table 1 summarizes the literature in the two groups. In this paper, we focus on occupational fatigue since it is: (i) a precedent to exhaustion, and (ii) more aligned to the working environment in advanced manufacturing environments. Moreover, our proposed framework is evaluated using multiple complex manufacturing tasks in an attempt to showcase its potential generalizability. The reader should note that multiple manufacturing tasks have only been examined in Maman et al. (2017).

From a detailed literature review, we could not identify any papers discussing the identification and diagnosis of fatigue. This may be attributed to the implicit assumption in the literature that management or the individual worker can handle those stages once fatigue has been detected. However, as indicated in Levenson (2017), "workplace fatigue is a systems problem", and there needs to be a systematic approach to identify its root-causes. This is a critical gap

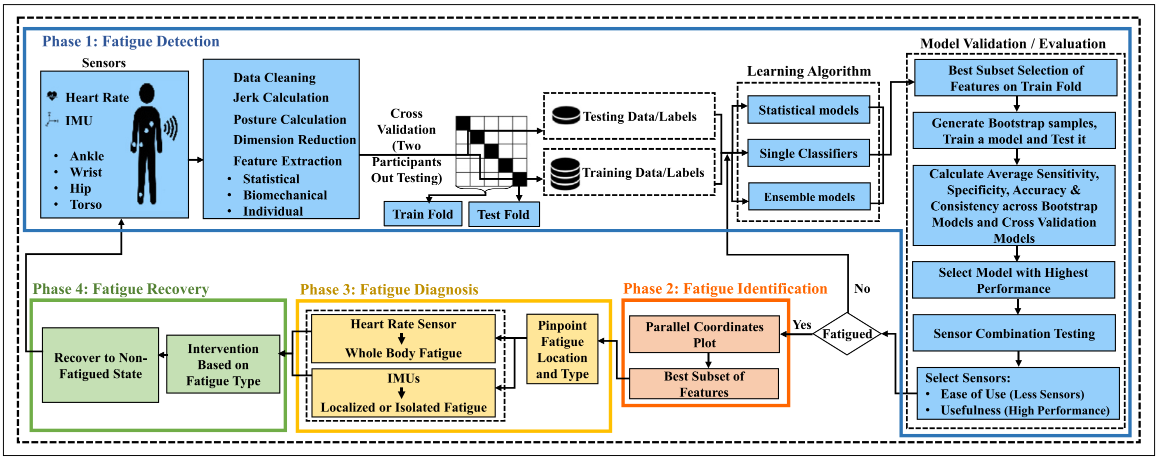

Fig. 1. An overview of proposed method.

since the end goal is intervening to prevent the unwanted negative consequences on the worker and the production process.

3. Proposed framework for fatigue management

Fig. 1 presents an overview of the four phases of the proposed framework for managing physical fatigue. The first phase is comprised of five main steps: (a) sensor selection, where practitioners should identify appropriate sensors for fatigue detection; (b) data preprocessing and feature generation, where the sensors' data are prepared for analysis; (c) model construction and validation, where statistical and data analytic models are trained for distinguishing between fatigued and non-fatigued states; (d) measuring usefulness, where models are evaluated based on accuracy, sensitivity, specificity, etc.; and (e) ease of use analysis, where the best model in step (d) is evaluated by constraining the number of sensors used. Note that steps (d) and (e) are based on the Technology Acceptance Model (TAM) (Marangunić & Granić, 2015). The outcome from Phase 1 is the selection of an appropriate model for prospective analysis. In Phase 2, the subset of features/predictors that are most frequently used in predicting the fatigue state is identified. This subset presents insights into what features are most predictive, which is an important input to the following phase. Phase 3 utilizes visual analytic methods (specifically an interactive parallel coordinates plot) to help management understand how the variation in the values of the predictors impact the fatigue state (i.e. from 0 to 1). Based on the insights gained from the fatigue diagnosis phase, a suitable evidence-based intervention can be selected in Phase 4.

3.1. Phase 1: fatigue detection

3.1.1. Sensor selection

Cavuoto and Megahed (2017) discussed several fatigue indicators, which included heart rate, heart rate variability, tremor and performance. They suggested that these indicators can be monitored using pervasive wearable sensors. In a follow-up work, Maman et al. (2017) showed that four IMU sensors (located at the ankle, hip, torso and wrist) coupled with a heart rate sensor can be used to detect fatigue in different manufacturing tasks. Similar to Maman et al. (2017), we suggest using these wearable sensors for fatigue detection. More importantly, our framework presents a systematic approach to answer the question: "what are the gains associated with wearing an extra sensor?" In essence, this question attempts to quantify whether the hassle and cost associated with wearing an extra sensor can be justified with a significant/practical improvement in fatigue detection. This question, which has not been addressed in the literature, is tackled in the usability analysis in Phase 1.

3.1.2. Data preprocessing

Cleaning The first step in analyzing data is to ensure that the data is correct and cleaned. For *wearable sensors* data, four main cleaning steps are proposed. First, a low-pass filter should be applied on the acceleration data for noise removal. Second, collected data should be visualized to check for any additional erroneous data, i.e. data that were not corrected through the automated filtering in step 1. Possible examples of erroneous data include faulty sensor values (too high and/or too low), and participants who had not experienced fatigue based on their subjective fatigue ratings. Third, the data from the different sensors should be synchronized and any observations that were captured outside of the experimental window should be eliminated. The fourth step involves the normalization of the heart rate data through the computation of: percent heart rate reserve (%HRR). Note that %HRR accounts for both an individual's resting heart rate (RHR) and his/her age-predicted maximum heart rate $HR_{\max} = 220 - \text{age}$. The %HRR can be computed as:

$$\%HRR = \frac{\text{Heart Rate} - RHR}{HR_{max} - RHR} \times 100. \quad (1)$$

The interpretation of the % HRR is a percentage of an individual's heart rate capacity being used. Since it accounts for both their resting and maximum heart rates it allows for standardizing the heart rate data. For example, if the %HRR =50, this means that the person is using 50% of their heart rate capacity, i.e. is half way between his/her resting and maximum heart rates.

Jerk and posture calculation The four IMUs (attached at the ankle, wrist, hip and torso) measure the acceleration associated with a person's dynamic motion. From the acceleration profile, other components of motion can be computed. Jerk, which is the derivative of acceleration with respect to time, should be computed since it has been shown to be effective in detecting fatigue in several occupational settings (see e.g., Catapult Sports, 2018 for several applications in professional sports). In addition, changes in work posture are also indicative of fatigue (Cavuoto & Megahed, 2017). In this paper, the approach of Baghdadi et al. (2018b) is used for posture calculation, where: (a) a Kalman filter is first used to calculate position in the three (*xyz*) directions, and then (b) posture is estimated from the positional data. The reader is referred to Baghdadi et al. (2018b) for more details on posture calculation.

Dimension reduction and feature extraction Based on the aforementioned data preprocessing steps, one would have 12 accelera-

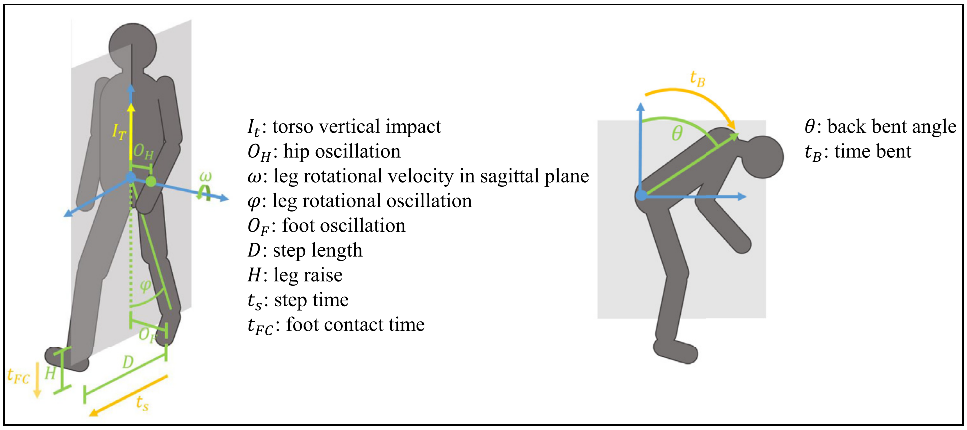

Fig. 2. Biomechanical features illustration.

tion profiles (4 IMUs × 3 directions [x y z]) and 4 jerk profiles (rate of change of the magnitude of the acceleration profile for each IMU) each sampled at 25 Hz. In addition, there is a %HRR profile sampled at 1000 Hz. These profiles cannot be directly used in predictive models and thus, features summarizing these profiles need to be generated. In this article, we propose utilizing features that would summarize the profiles based on a non-overlapping time window of the 17 profiles. The selection of the length of the time window should depend on: (a) length of the cycle for task, (b) consequences of fatigue on the worker and production, and (c) managing the trade-off between false alarms and early detection.

To capture the changes within the profile and provide insights to the later isolation and diagnosis phases, three sets of features are generated from the 17 profiles. The first set corresponds to statistical features from the acceleration, jerk, posture and %HRR. For each of these profiles, the mean and coefficient of variation (CV) are computed for each time-window to capture the intensity and variation changes. Features capturing the intensity and spread are commonly used in the fatigue detection literature (see e.g., Bao & Intille, 2004; Maman et al., 2017; Pirttikangas, Fujinami, & Nakajima, 2006). The second set corresponds to biomechanical features, which allow for identifying and diagnosing the type of fatigue. This set includes features such as: *number of steps in the time interval, mean step time and length*, and *mean foot/hip oscillations*. The biomechanical features used in our framework are depicted in Fig. 2. Note that these features are calculated for each time window. Those features are computed based on the code provided in Baghdadi et al. (2018b). The third, and last feature set contains both age and gender, which may be used to explain performance differences across different individuals (see Kent-Braun, Ng, Doyle, & Towse, 2002; Wojcik, Nussbaum, Lin, Shibata, & Madigan, 2011 for more details). A description of the proposed features for each of the three sets is provided in Table 2.

3.1.3. Model construction and validation

Cross validation A leave p-participants out cross validation approach can be used to split the preprocessed dataset into training and testing sets. Cross validation is commonly used to avoid overfitting (Weng et al., 2017). A typical approach to cross validation is dividing the dataset into 10 folds, where the models are selected based on the average/median prediction performance across 10 non-overlapping test datasets. The literature suggests that 10-fold cross validation may reduce the variation between the train and test performance (Dag, Topuz, Oztekin, Bulur, & Megahed, 2016). Note that in fatigue detection studies such as ours, each participant's data maybe autocorrelated. Thus, the plain k-fold cross validation approach is not suitable since the train and test datasets are not independent. To alleviate this problem, we recommend leaving p participants out for the cross validation, where the value of p corresponds to approximately 10% of the participants in the data analytic study.

Feature selection and dimension reduction When the number of potential features/predictors is large, the computational complexity for model training increases. Feature reduction is typically applied to reduce the computational burden. More importantly, it leads to: (a) an improved prediction performance, and (b) an increased generalization capability. Algorithms for feature selection/reduction can be categorized into three main groups (Blum & Langley, 1997): (1) *filter methods*, where univariate statistical approaches are typically used to select features based on their relationship to the response, (2) *wrapper methods*, where the important features are kept based on their prediction performance, and (3) *embedded methods*, which involve the use of methods such as LASSO for selecting the most predictive features.

Since the end goal of our proposed framework is to enable the diagnosis of fatigue and the recommendation of an appropriate intervention, we recommend a two-step approach for feature selection. In the first step, simple filter approaches (e.g., information gain or correlation analysis) should be combined with visualizations (e.g., time series charts, parallel coordinates plot, and scatter diagrams). The goal of the first step is to provide practitioners with an understanding of how fatigue affects and/or is associated with changes in the potential predictors. From this step, any features that are unchanged in the fatigued and non-fatigued states should be removed. The reader should note that the insights gained from the visualization will also be utilized in diagnosing the root-causes of fatigue. In the second step, several structured wrapper and/or embedded methods (e.g., best subset selection and LASSO) should be examined. Preference should be given to techniques that result in a small number of features (i.e. more interpretable) and a relatively large prediction performance (i.e. good fatigue detection with a low false alarm rate).

Bootstrapping To further prevent over-fitting and the bias associated with selecting a training dataset, we recommend the use of bootstrapping (Efron & Tibshirani, 1993), which is a computational procedure that uses intensive re-sampling with replacement. An important assumption behind bootstrapping is that the sample distribution is a good approximation to the population's distribution. Recent studies have shown an improved performance of analytical models when bootstrapping is deployed (e.g., see Argon & Ziya, 2009; Ødegaard & Roos, 2014).

Table 2
Generated feature sets.

Category	#	Feature	Definition	Justification
Statistical	1	%HRR.Mean	Average percent of heart rate reserve	Baghdadi, Maman, Lu, Cavuoto, and Megahed (2017), Bao (2003), Bao and Intille (2004) Bonato et al. (2003), Côté, Mathieu, Levin, and Feldman (2002) Foster (1998), Maman et al. (2017), Quagliarella, Sasanelli, and Belgiovine (2008), Young, Trudeau, Odell, Marinelli, and Dennerlein (2013), and Zhang et al. (2014)
	2	Wrist.jerk.Mean	Average wrist jerk or smoothness magnitude	
	3	Wrist.ACC.Mean	Average wrist acceleration magnitude	
	4	Wrist.xposture.Mean	Average wrist angular position in sagittal plane	
	5	Wrist.yposture.Mean	Average wrist angular position in transverse plane	
	6	Wrist.zposture.Mean	Average wrist angular position in coronal plane	
	7	Hip.jerk.Mean	Average hip jerk magnitude	
	8	Hip.ACC.Mean	Average hip acceleration magnitude	
	9	Hip.xposture.Mean	Average hip angular position in coronal plane	
	10	Hip.yposture.Mean	Average hip angular position in transverse plane	
	11	Hip.zposture.Mean	Average hip angular position in sagittal plane	
	12	Torso.jerk.Mean	Average torso jerk magnitude	
	13	Torso.ACC.Mean	Average torso acceleration magnitude	
	14	Torso.xposture.Mean	Average torso angular position in sagittal plane (bending)	
	15	Torso.yposture.Mean	Average torso angular position in transverse plane	
	16	Torso.zposture.Mean	Average torso angular position in coronal plane	
	17	Ankle.jerk.Mean	Average ankle jerk magnitude	
	18	Ankle.ACC.Mean	Average ankle acceleration magnitude	
	19	Ankle.xposture.Mean	Average ankle angular position in coronal plane	
	20	Ankle.yposture.Mean	Average ankle angular position in transverse plane	
	21	Ankle.zposture.Mean	Average ankle angular position in sagittal plane	
	22	%HRR.CV	Coefficient of variation in %HRR	
	23	Wrist.jerk.CV	Coefficient of variation in the wrist jerk	
	24	Wrist.ACC.CV	Coefficient of variation in the wrist acceleration magnitude	
	25	Wrist.xposture.CV	Coefficient of variation in the wrist angular position in sagittal plane	
	26	Wrist.yposture.CV	Coefficient of variation in the wrist angular position in transverse plane	
	27	Wrist.zposture.CV	Coefficient of variation in the wrist angular position in coronal plane	
	28	Hip.jerk.CV	Coefficient of variation in the hip jerk magnitude	
	29	Hip.ACC.CV	Coefficient of variation in the hip acceleration magnitude	
	30	Hip.xposture.CV	Coefficient of variation in the hip angular position in coronal plane	
	31	Hip.yposture.CV	Coefficient of variation in the hip angular position in transverse plane	
	32	Hip.zposture.CV	Coefficient of variation in the hip angular position in sagittal plane	
	33	Torso.jerk.CV	Coefficient of variation in the torso jerk magnitude	
	34	Torso.ACC.CV	Coefficient of variation in the torso acceleration magnitude	
	35	Torso.xposture.CV	Coefficient of variation in the torso angular position in sagittal plane	
	36	Torso.yposture.CV	Coefficient of variation in the torso angular position in transverse plane	
	37	Torso.zposture.CV	Coefficient of variation in the torso angular position in coronal plane	
	38	Ankle.jerk.CV	Coefficient of variation in the ankle jerk magnitude	
	39	Ankle.ACC.CV	Coefficient of variation in the ankle acceleration magnitude	
	40	Ankle.xposture.CV	Coefficient of variation in the ankle angular position in coronal plane	
	41	Ankle.yposture.CV	Coefficient of variation in the ankle angular position in transverse plane	
	42	Ankle.zposture.CV	Coefficient of variation in the ankle angular position in sagittal plane	
Biomechnical	43	Number of steps	Number of gait cycles during the fixed time interval	Bächlin, Förster, and Tröster (2009), Baghdadi et al. (2018b), Dolan and Adams (1998), Hallemans et al. (2009), Larivière, Gagnon, and Loisel (2000), Strohrmann, Harms, Kappeler-Setz, and Troster (2012), Willson and Kernozek (1999), and Yun, Bachmann, Moore, and Calusdian (2007)
	44	Mean step time	Average duration of each gait cycle	
	45	Mean step length	Average length of each gait cycle	
	46	Time bent	The duration spent in bent posture	
	47	Mean back bent angle	Average angle of torso in bent posture w.r.t vertical axis	
	48	Mean hip oscillation	Average side-to-side range of motion in hip	
	49	Mean foot oscillation	Average side-to-side range of motion in foot	
	50	Mean leg rotational velocity in sagittal plane	Average angular velocity of leg in sagittal plane	
	51	Mean leg rotational oscillation in sagittal plane	Average angular range of motion for leg in sagittal plane	
	52	Mean torso vertical impact	Average value of peak vertical acceleration in torso	
	53	Mean back rotational position in sagittal plane	Average range of bending posture while doing the task	
Individual	54	Age	-	Kent-Braun et al. (2002) and Wojcik et al. (2011)
	55	Gender		

Table 3
Comparing the three different analytical categories. Table is adapted from Wang (2016).

	Statistical models	Single classifiers	Ensemble models
High accuracy in general			✓
High speed of learning against # of variables and samples	✓		
High tolerance to redundant variables		✓	✓
High tolerance to collinearity		✓	✓
High dealing with overfitting		✓	
Less complexity and easy parameter handling	✓		

Analytical modeling The analytical classification models can be categorized into: (a) statistical models, (b) single classifiers, and (c) ensemble models. The pros and cons of using these methods (Kotsiantis, Zaharakis, & Pintelas, 2007; Wang, 2016) are shown in Table 3. Note that we do not include more advanced deep learning models since they often require special computing resources (i.e. graphical processing units, GPUs) and would be quite difficult to implement for a large number of workers.

Several classification methods, i.e. statistical models, single classifiers, and ensemble models, are viable candidates for utilization in fatigue prediction. From our framework's perspective, it is impossible to predetermine which methods will work best for a given application. This is due to the fact that these methods are data-driven and thus, are application-dependent. In the following paragraphs, we highlight some commonly used methods within each category.

Statistical models attempt to build a relationship between the input variables and response through the use of parametric methods. Examples include: *logistic regression* and *penalized logistic regression*. Those are classification techniques where the probability of a dichotomous outcome is a function of the predictors/features (Algamal & Lee, 2015; Hosmer Jr, Lemeshow, & Sturdivant, 2013). A key difference between the two aforementioned approaches lies in how they handle sparse datasets. Specifically, logistic regression's performance can vary significantly with sparse data (King & Zeng, 2001). On the other hand, the penalized logistic regression approach usually provides a more consistent performance (Maman et al., 2017).

In the single classifier category, some commonly used classifiers include: decision trees (DT), naive Bayes (NB), artificial neural networks (ANN), k-nearest neighbors (kNN), and support vector machines (SVM). Those non-parametric approaches are commonly used in human performance modeling applications. The reader is referred to Afsar, Cortez, and Santos (2015); Ghaderyan, Abbasi, and Saber (2018); Rescio, Leone, and Siciliano (2018); Ryu and Kim (2017) for examples of those applications. We recommend exploring one or more of those models for fatigue classification.

For the third category, ensemble models are comprised of several single classifiers, where the final classification of the response is based on some voting or weighting procedure (Dietterich, 2000). The premise for these methods is that combining a large number of single classifiers allows for a more diverse representation of the data and consequently, a more accurate prediction. Commonly used ensembles include: (a) random forests (RFs), which are ensemble classification algorithms that utilize trees as base classifiers to generate many classifiers and aggregate their results via voting (Breiman, 2001); (b) bagging (Breiman, 1996), where bootstrapping is used to generate a new training dataset, and combine several base learners to fit a weak learner to the data; and (c) boosting (Schapire, 2003), which creates different base learners by sequentially reweighing the instances in the training set. Boosting gives different weights to the base learners based on their accuracy. The final model obtained by the boosting algorithm is a linear combination of several base learners weighted by their own performance. For a detailed introduction on the aforementioned analytical models, the reader is referred to Han, Pei, and Kamber (2011); James, Witten, Hastie, and Tibshirani (2013).

3.1.4. Measuring usefulness

To evaluate the performance of the analytical models, we recommend using five performance measures: (a) accuracy, which presents the percentage of correct classifications made by a given model, (b) *sensitivity*, which captures the ability to detect the fatigued cases, (c) specificity, which measures the correct classification of non-fatigued cases, (d) G-mean, which is defined as the square root of sensitivity times specificity, and (e) a newly proposed consistency metric, which is a simple metric that captures the absolute difference between the metrics in (b) and (c). This metric can be used by practitioners to gauge whether a model is equally capable of predicting both the fatigued and non-fatigued states. The mathematical formula below show how each of these metrics is computed first for each fold, and then averaged across all folds:

$$Accuracy_j = \frac{1}{n}\sum_{i=1}^{n}\frac{TP_{ij}+TN_{ij}}{TP_{ij}+TN_{ij}+FP_{ij}+FN_{ij}}. \tag{2}$$

$$Mean\ Accuracy = \frac{1}{m}\sum_{j=1}^{m}Accuracy_j. \tag{3}$$

$$Sensitivity_j = \frac{1}{n}\sum_{i=1}^{n}\frac{TP_{ij}}{TP_{ij}+FN_{ij}}. \tag{4}$$

$$Mean\ Sensitivity = \frac{1}{m}\sum_{j=1}^{m}Sensitivity_j. \tag{5}$$

$$Specificity_j = \frac{1}{n}\sum_{i=1}^{n}\frac{TN_{ij}}{TN_{ij}+FP_{ij}}. \tag{6}$$

$$Mean\ Specificity = \frac{1}{m}\sum_{j=1}^{m}Specificity_j. \tag{7}$$

$$G\text{-}mean_j = \sqrt{Sensitivity_j \times Specificity_j}. \tag{8}$$

$$Mean\ G\text{-}mean = \frac{1}{m}\sum_{j=1}^{m}G\text{-}mean_j. \tag{9}$$

$$Consistency_j = |Sensitivity_j - Specificity_j|. \tag{10}$$

$$Mean\ Consistency = \frac{1}{m}\sum_{j=1}^{m}Consistency_j. \tag{11}$$

where TP, TN, FP, FN denote the number of true positives, true negatives, false positives, and false negatives, respectively. i denotes the number of the bootstrapping samples, j is the number of the training or testing data sets, n is the number of bootstrapped samples, and m is the number of folds in the leave p-participants-out cross validation.

3.1.5. Ease of use analysis

In addition to evaluating its usefulness, an important aspect for technology adoption is usability. In the context of our framework, usability can be measured using two metrics: (a) total number of features selected, and (b) total number of sensors needed to generate these features. In general, models are more interpretable if the number of features are smaller (assuming no significant differences in prediction capabilities). Workers and more practitioners will also be more inclined to adopt the framework if it requires less sensors since it will: (i) be much cheaper; for example, requiring one IMU instead of four, would reduce the cost by a factor of four; (ii) make the process less invasive to the worker; and (iii) reduce the time needed for the worker to wear and strap all the sensors. Therefore, our framework will not only consider prediction performance, but it will also evaluate how the prediction performance varies while restricting the number of sensors that can be used. At this stage, one would have a model that can accurately predict the fatigue state (based on the leave p-participants out cross validation approach), while having a relatively small number of features. This model can now be deployed for near real-time prediction.

3.2. Fatigue identification

Once the model is deployed and fatigue is identified, it is important to understand how the predictors' change when an individual becomes fatigued. Typically, machine learning models are thought of as "black boxes", where it is difficult to understand how the predictors affect the response. However, an important aspect of recovering from fatigue is being able to diagnose its root-causes. Since we favor having a lower number of features in our model selection (see Section 3.1.5), we hypothesize that the chosen prediction model will have a relatively low number of features. Thus, one can use a parallel coordinates plot to depict how the chosen features vary with the dichotomous response. The use of such a plot will enhance the interpretation of the model and assist practitioners in diagnosing the type of fatigue in the next phase.

3.3. Fatigue diagnosis

In this phase, one would determine which type of fatigue occurred. Since this framework focuses only on physical fatigue, there are two main types of fatigue that are possible (Cavuoto & Megahed, 2017): (a) whole body fatigue, and (b) localized muscle fatigue. Based on the parallel coordinates plot from the previous phase, one would identify the important features for prediction. If the features are derived from only one IMU (as in our first case in Section 4.1), one would conclude that the worker is experiencing localized muscle fatigue, near that IMU's location. Alternatively, if the features are derived only from the heart rate sensor (see Section 4.2), this implies that the worker is experiencing whole body fatigue. The last possibility would include features selected from one or more IMU and the heart rate sensor. In this case, the individual is experiencing a combination of whole-body fatigue (i.e. respiratory related) and localized fatigue. Based on the diagnosis, one can assign appropriate interventions in the next stage.

3.4. Fatigue recovery

From a management perspective, it is important to prescribe interventions that eliminate/reduce the safety hazards. In essence, "safety does not happen by accident" (Vries, Koster, & Stam, 2016) and thus, it is important to intervene to eliminate/mitigate the sources of fatigue. We recommend utilizing the safety design hierarchy (Manuele, 2005) from safety engineering. This hierarchy presents a structured approach for interventions, where practitioners should consider six actions in order of effectiveness. Since this is a well-known concept to safety professionals, we do not detail this further.

In our estimation, the fatigue diagnosis stage allows practitioners to directly pinpoint the hazard (i.e. type of fatigue). Practitioners can then prescribe interventions from a large number of options, including: (a) redesigning the task (which can eliminate the development of fatigue), (b) assigning rest breaks (which can reduce the level of fatigue before it reaches potentially dangerous levels), and (c) job rotation (where workers would essentially cycle between harder and easier jobs). The type of intervention assigned will depend on the resources available to safety practitioners and the constrains of their production processes. For this reason, we only recommend the adoption of the safety design hierarchy without providing a recommendation for the type of interventions to be assigned. The reader is referred to the survey of Lu et al. (2017) for a discussion of the type of interventions used by advanced manufacturing workers and safety professionals in combating physical fatigue at the workplace.

4. Case studies

To evaluate the performance of the proposed framework, we examine two case studies. The first case study involves a simulated manual material handling (MMH) task, and the second is a supply pick-up and insertion (SI) task. Both case studies replicate typical fatiguing manufacturing tasks (see the survey in Lu et al. (2017) for details) in a controlled lab environment in order to facilitate the data collection process. Since the data collection, data preprocessing and model construction steps are the same for the two tasks, we only explain them in detail in Sections 4.1.1 and 4.1.2.

4.1. Case study 1: manual material handling

4.1.1. Data collection, preprocessing and feature generation

Twenty four participants (9 females, 15 males; mean age 36.37 years with the standard deviation of 16.67 years) were recruited over a period of 11 months from the local community. Five of the participants were manufacturing workers, and the remainder represented a convenience sample of students with varying degrees of physical work experience. All participants reported that they were in good physical and mental health. In addition, they were screened by completing the Physical Activity Readiness Questionnaire (PAR-Q) (Thomas, Reading, & Shephard, 1992) at the start of the session to assess their eligibility to participate. They also provided informed consents at the start of the experiment. All study procedures were approved by the university's institutional review board (IRB).

Participants completed one three-hour experimental session for the simulated MMH task and another for the SI task. The order of the two experiments was randomized and participants had to complete the experiments in different days. The MMH task involved palletizing and transporting several weighted containers (see Fig. 3). Each participant was asked to perform the task at a set pace for three hours continuously (without breaks) to induce fatigue. Per the discussion in Section 3.1, four IMUs placed at the ankle, hip, wrist and torso, and a heart rate monitor on the chest were used for data collection. Furthermore, participants provided their subjective exertion (RPE) using the Borg Scale (Borg, 1998) every ten minutes.

The four step data cleaning procedure discussed in Section 3.1.2 was deployed for our case studies. After using the low pass filter for de-noising the IMU data, we used $RPE \geq 13$ as a cutoff for fatigue in step 2 per the analysis of Maman et al. (2017). Based on step 2, a total of nine participants were removed from

Fig. 3. A participant carrying out the MMH task.

the data for the following reasons: (a) three participants did not get fatigued by the end of the experiment; (b) three reported being fatigued within the first half an hour of the experiment (i.e. they may have been fatigued prior to conducting the experiment); (c) the IMUs failed to record data for two of the participants during the experiment; and (d) one of the participants deviated from the experimental protocol by taking two 10-minute bathroom breaks. As a result, we ended up with 15 participants whose data were deemed reliable for analysis. After synchronizing the data from the sensors in step 3, we removed the first 10 min of experimental data to avoid capturing the learning effect (Baghdadi et al., 2018a; Maman et al., 2017). Then, the % HRR was computed in step 4 as explained in the methodology section. After step 4, the jerk and posture profiles were generated based on the procedure of Baghdadi et al. (2018b) which was highlighted in Section 3.1.2.

To reduce the computational burden and to maintain a balanced dataset for training, we have only kept 20% of the data for each participant. These 20% corresponded to: (a) 10% (i.e. $10\% \times 180$ min $= 18$ min) at the beginning of the experiment, after the first 10 min are removed, where the participants are not fatigued, and (b) 10% at the end, where the participants are fatigued. The rationale for removing the 80% of the data is two-fold. First, the separation ensures that the differences between the fatigued and non-fatigued data for each participant are maximized, while the differences within each group are minimal. Second, based on Maman et al. (2017), we can assume that the size of the data can be decreased without losing much information related to fatigue detection. For each participant, we coded the response as 0 (for the first 18 min) and 1 for the latter 18 min to reflect the non-fatigued and fatigued states, respectively. Recall that our data cleaning procedure ensured that these values reflect the estimated RPEs by each participant.

Based on the discussion in Section 3.1.2, it is important to set the size of the time window prior to generating the features in Table 2. In our case studies, we have used a non-overlapping time window of 2 min. This means that each of the 18 min was divided into nine fractions of two-minute periods. The rationale for selecting two-minutes for the time window was mainly based on the observation that the average cycle time for MMH was approximately one minute. Therefore, each two-minute time interval is guaranteed to include at least one cycle of the task. Based on this decision, we generated the proposed features from each sensor for each two-minute time window. The reader can replicate our analysis by consulting our data and code (see the *Supplementary Materials* Section).

4.1.2. Model construction and validation

As a first step for feature selection, time series plots of all features were constructed to evaluate which features were virtually unchanged from the non-fatigued to fatigued states. Based on the visualizations, 15 (of the 55 candidate) features were dropped. The second step (where wrapper or embedded methods are used) of feature selection is applied after the training and test samples are generated using the leave p-participants out cross validation approach. Based on the discussion in Section 4.1.1, we had 15 participants with reliable data for this case study. Thus, $p = 2$ (i.e. $2/15 = 13\%$) was used for the leave p-participants-out cross validation approach to split the data into training and test sets. This resulted into 105 possible training/test sets $(15!/((15-2)! \times 2!) = 105)$, which we would evaluate to obtain an estimate of the variation in the performance of our analytical models.

Prior to deploying the analytical models, two additional tasks were carried out. First, the last step of variable selection was deployed using two popular methods: best subset selection and LASSO (refer to Section 3.1.3 for details). Second, to reduce the bias from model training and improve the performance of the predictive models bootstrap resampling with replacement was applied to the training data. The sample size for each bootstrap sample was $n = 234$, which was based on 13 participants $\times$ 18 samples per participant. For our analysis, we used 200 bootstrap samples (each having $n = 234$) based on the recommendation of Pattengale, Alipour, Bininda-Emonds, Moret, and Stamatakis (2009).

To develop the fatigue prediction models, several methods were applied during our preliminary analysis of the data. The models evaluated included: logistic regression, penalized logistic regression, decision trees (DT), naive Bayes (NB), k-nearest neighbors (kNN), support vector machines (SVM), and three ensemble models (random forest (RF), bagging, and boosting). Due to their relatively poor performance, DT, NB and kNN were eliminated. In addition, models using best subset selection typically had better prediction performance with less features than their LASSO counterparts. Therefore, our case study focused on using the best subset selection with the following five analytical models: (a) logistic regression, (b) SVM, (c) RF, (d) RF with bagging (hereafter bagging), and (e) RF with boosting (hereafter boosting). In addition, we compared these five models to the approach of Maman et al. (2017) since it was the only paper that considered multiple tasks in the context of occupational fatigue (see Table 2). To ensure that the comparison is fair, we considered two different variants of the penalized logistic regression approach with LASSO proposed in Maman et al. (2017). The first is utilizing their approach and features (on our data), and the second involves using their methodology with our features and data. In our estimation, this allows us to better evaluate whether our proposed method is superior to theirs. The reader should note that they did not consider model interpretation in their feature generation and thus we expect that our features are easier to interpret by practitioners.

4.1.3. Fatigue detection results

In Table 4, the predictive performance of our five models is compared with the two variants from Maman et al. (2016). The table shows the mean (and standard deviation in parentheses) for each of our four metrics. In addition, the average number of features selected by each model is also presented. The reported results are based on 105 constructed test datasets from the two-participants-out cross validation. For the first three numeric columns, a higher value is desired since it reflects a better prediction performance. The consistency column captures the aver-

Table 4
Mean performance and the corresponding standard deviation of the classification methods for fatigue detection in MMH task, (the recommended model is **in bold**).

Category	Model	Sensitivity	Specificity	Accuracy	G-mean	Consistency	# of Features
BSS	**Random Forest**	**0.879 (0.14)**	**0.879 (0.15)**	**0.879 (0.09)**	**0.869 (0.10)**	**0.152 (0.18)**	**5.352**
	Bagging	0.872 (0.13)	0.869 (0.15)	0.870 (0.09)	0.863 (0.10)	0.143 (0.17)	5.352
	Boosting	0.871 (0.13)	0.872 (0.15)	0.870 (0.08)	0.862 (0.10)	0.147 (0.17)	5.352
	Support Vector Machine	0.811 (0.18)	0.828 (0.17)	0.820 (0.11)	0.805 (0.13)	0.198 (0.19)	5.352
	Logistic Regression	0.790 (0.17)	0.766 (0.20)	0.778 (0.11)	0.758 (0.15)	0.227 (0.20)	5.352
LASSO	Penalized Logistic Regression*	0.802 (0.20)	0.916 (0.11)	0.859 (0.11)	0.846 (0.13)	0.175 (0.20)	18.943
	Penalized Logistic Regression	0.810 (0.13)	0.775 (0.17)	0.793 (0.08)	0.781 (0.09)	0.197 (0.16)	11.133

* Features used in the model are only those generated in Maman et al. (2017).

age absolute difference between the sensitivity and specificity for each model, evaluated on the 105 test datasets. It is noted that the smaller the consistency is, the similar performance in detecting fatigued and no-fatigued states simultaneously would be. Moreover, a smaller number of features facilitates the interpretation of the model, which is important in the fatigue identification and diagnosis phases.

Four main observations from Table 4 need to be highlighted. First, as expected from the preliminary analysis, the number of features selected with the best subset selection are much less than those selected by the LASSO model. This means that the usability of the analytical models with the BSS model is much higher than that with LASSO since practitioners' need to monitor and understand approximately five features (instead of 11 or 19). Second, the performance of all seven models is relatively high with an overall average accuracy greater than 0.77. Third, the performance of the three ensembles is better than the remaining models. Fourth, the penalized logistic regression of Maman et al. (2017) outperforms its variant with our features from a prediction perspective. However, this comes at the cost of adding eight features to the model (i.e. $\approx 70\%$ increase in the variables used). Based on these observations and this case study, one can conclude that our framework has shown higher detection performance (with less features) when compared to the approach in Maman et al. (2017).

The next logical research question is to examine how the prediction performance varies while limiting the number of sensors used. To evaluate this question, we utilize the random forest model since Table 4 showed that it had the highest mean accuracy, sensitivity, specificity and G-mean when compared to the other two ensembles. Table 5 reports the prediction results, when features are limited to those from one, two, three, four and all sensor combinations. Note that the values that are not shown in the table (e.g. ankle, hip, wrist and HR sensors) reflect scenarios when a prediction was not possible. This means that the main features that detected the fatigue were eliminated with the added constraints on which possible features to select from.

From the results in Table 5, one can see that the prediction performance does not vary significantly as the number of sensors' are changed. For example, the average accuracy varies from 0.855 to 0.880 (with a standard deviation $\approxeq 0.09$) as the number of sensors vary. This is only true if the torso IMU is included in the analysis. Based on this observation, we recommend only using the torso

Table 5
Mean performance and the corresponding standard deviation of the random forest model for fatigue detection using different sensor combinations for the MMH task (the recommended model is **in bold**).

# sensors	Sensor Combination					Sensitivity	Specificity	Accuracy	G-mean	Consistency
5	Ankle	Hip	Wrist	Torso	HR	0.879 (0.14)	0.879 (0.15)	0.879 (0.09)	0.869 (0.10)	0.152 (0.18)
4	Ankle	Hip	Wrist	Torso		0.883 (0.14)	0.878 (0.15)	0.880 (0.09)	0.871 (0.10)	0.148 (0.18)
	Ankle	Hip		Torso	HR	0.851 (0.16)	0.883 (0.13)	0.867 (0.10)	0.858 (0.11)	0.149 (0.16)
		Hip	Wrist	Torso	HR	0.883 (0.12)	0.872 (0.15)	0.877 (0.08)	0.870 (0.10)	0.147 (0.16)
	Ankle		Wrist	Torso	HR	0.877 (0.13)	0.873 (0.15)	0.875 (0.08)	0.867 (0.10)	0.146 (0.16)
	Ankle	Hip	Wrist		HR	–	–	–	–	–
3			Wrist	Torso	HR	0.880 (0.12)	0.874 (0.15)	0.877 (0.08)	0.869 (0.09)	0.142 (0.16)
	Ankle			Torso	HR	0.846 (0.15)	0.882 (0.13)	0.864 (0.09)	0.856 (0.10)	0.148 (0.16)
	Ankle	Hip		Torso		0.851 (0.16)	0.883 (0.13)	0.867 (0.10)	0.858 (0.11)	0.149 (0.16)
		Hip	Wrist	Torso		0.882 (0.12)	0.872 (0.15)	0.877 (0.08)	0.869 (0.10)	0.147 (0.16)
		Hip		Torso	HR	0.860 (0.16)	0.885 (0.14)	0.872 (0.10)	0.863 (0.11)	0.150 (0.17)
	Ankle		Wrist	Torso		0.877 (0.13)	0.873 (0.15)	0.875 (0.08)	0.867 (0.10)	0.146 (0.16)
	Ankle	Hip			HR	–	–	–	–	–
	Ankle	Hip	Wrist			–	–	–	–	–
	Ankle		Wrist		HR	–	–	–	–	–
		Hip	Wrist		HR	–	–	–	–	–
2			Wrist	Torso		0.880 (0.12)	0.874 (0.15)	0.877 (0.08)	0.869 (0.09)	0.142 (0.16)
	Ankle			Torso		0.846 (0.15)	0.882 (0.13)	0.864 (0.09)	0.856 (0.10)	0.148 (0.16)
		Hip		Torso		0.860 (0.16)	0.885 (0.14)	0.872 (0.10)	0.863 (0.11)	0.150 (0.17)
				Torso	HR	0.842 (0.16)	0.867 (0.14)	0.855 (0.10)	0.846 (0.11)	0.149 (0.17)
	Ankle	Hip				–	–	–	–	–
	Ankle				HR	–	–	–	–	–
	Ankle		Wrist			–	–	–	–	–
		Hip	Wrist			–	–	–	–	–
		Hip			HR	–	–	–	–	–
			Wrist		HR	–	–	–	–	–
1				**Torso**		**0.847 (0.16)**	**0.864 (0.14)**	**0.855 (0.10)**	**0.847 (0.11)**	**0.148 (0.16)**
	Ankle					–	–	–	–	–
		Hip				–	–	–	–	–
			Wrist			–	–	–	–	–
					HR	–	–	–	–	–

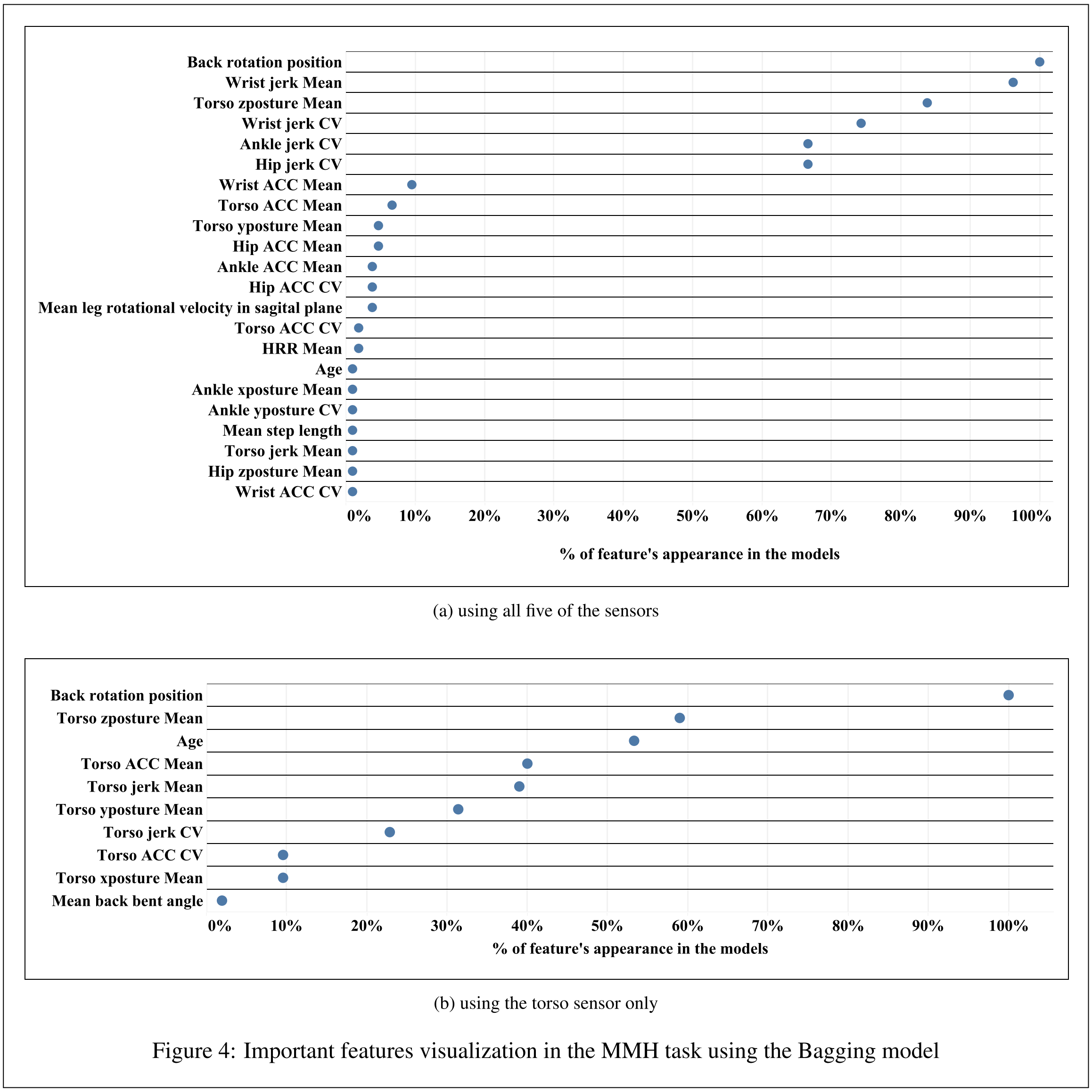

Fig. 4. Important features visualization in the MMH task using the Bagging model.

IMU sensor for detecting fatigue in manual material handling environments (that are similar to those analyzed in our case study). While the prediction performance is almost the same, the costs incurred by the firm are much lower, and the usability of the system by using only one sensor is significantly improved. This is an important practical takeaway, which has not been reported in previous studies investigating fatigue in MMH tasks (see the references in Table 1).

4.1.4. Fatigue identification results

A first step in understanding fatigue is to examine how frequently a feature is selected all of the 105 two-participants-out cross validation bagging model test sets. In this section, we limit our analysis to two cases: (a) when all five sensors are utilized, and (b) when only the torso sensor is used. The results for these analyses are shown in Fig. 4(a) and (b), respectively. From both figures, one can see that all three categories of features (i.e. statistical, biomechanical, and individual features) are selected in our models. For the five sensor case, one biomechanical feature (*mean back rotational position*, i.e. feature #53 in Table 2) and five statistical features appeared in more than 65% of the models. All other remaining features appeared in less than 10% of the models. On the other hand, age becomes a much more predictive factor if we only rely on the torso sensor. In that case, *mean back rotational position* is still selected in 100% of the models.

Once a list of predictive/important features is established, we then investigate how those features vary as the participant transition from the non-fatigued to fatigued states. As highlighted in Section 3.2, this analysis can be done visually using a parallel coordinates plot. Fig. 5 depicts this analysis (using the median model sorted by accuracy) for the five sensors and one sensor cases. Note that the lines graphed in these plots represent the average values per variable for each of the two participants in the test set examined by the median model.

From Fig. 5(a), one can see that all of the six features highlighted in Fig. 4(a) are present in the median model. It is interesting to note that only the wrist features exhibited a consistent pattern across both participants when examining the fatigued cases (**black line**) and the non-fatigued cases (**gray line**). Specifically, the *coefficient of variation for wrist jerk* tended to be higher, and the *mean wrist jerk* tended to be lower in the fatigued cases. For the remaining four features, there were not any consistent

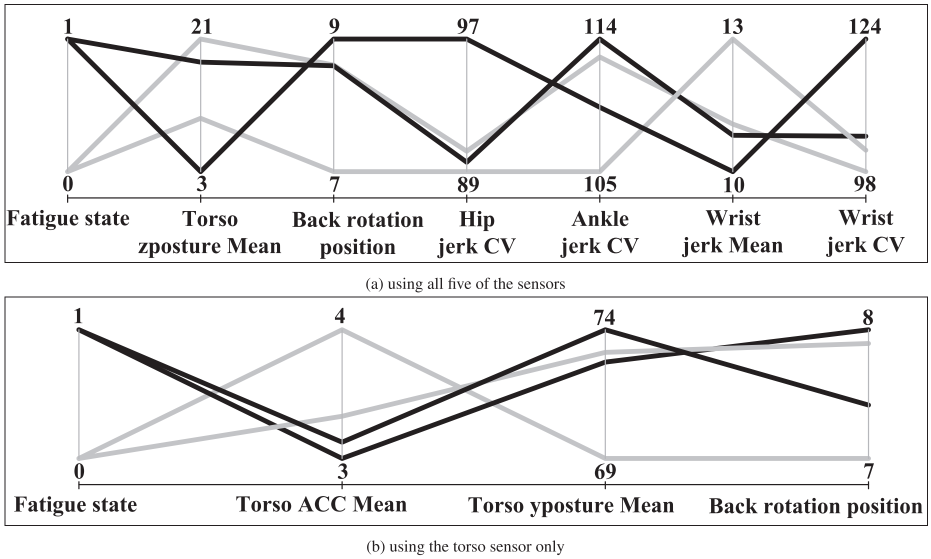

Figure 5: Features visualization for the median Bagging model in the MMH task

Fig. 5. Features visualization for the median Bagging model in the MMH task.

patterns for both test subjects. Similarly from Fig. 5(b), one can see that only the *torso ACC mean* feature showed a clear separation between the fatigued and non-fatigued states for both participants. We hypothesize that these two figures may provide justification for why the ensemble models outperformed the logistic regression models. Specifically, these plots may suggest interactive and non-linear effects that can be trained for and captured using the ensemble models.

4.1.5. Fatigue diagnosis results

From the fatigue identification results, one can conclude that the type of fatigue is localized at the back. This conclusion is supported by: (a) the prediction performance is almost unchanged (and high) when only the features from the torso sensor are used for prediction, and (b) the *mean back rotational position* was selected as an important feature in 100% of the models. This was the only feature that was selected in 100% of the models. Our results are consistent with findings in the ergonomics literature, which suggest that manual material handling may lead to a higher prevalence of back injuries (Mital, 2017).

4.2. Case study 2: supply pick up and insertion

4.2.1. Task description and data preparation

Similar to the task in Maman et al. (2017), we examined supply pickup and insertion task. The task involved walking while carrying supplies, and then bending forward to unscrew and fasten bolts at the supply box (destination). A snapshot of the experiment is provided in Fig. 6. The task's cycle time was set for two minutes to mimic the activity in Maman et al. (2017). By design, this activity should be less fatiguing than the MMH task of the first case study.

The mechanism used to collect and preprocess data is similar to that used in case study 1. The four step data cleaning procedure suggested in Section 3.1.2 resulted in having 13 participants (instead of 15 for the first case study) with reliable/clean data. Then, the sensor data were synchronized after removing the initial ten minutes of the experiment. After down sampling, the jerk, posture, and % HRR profiles were computed. Similar to the MMH task, the first 18 min of the data (after removing the learning period) were labeled as *not fatigued* and the last 10 min were marked as *fatigued*.

Fig. 6. Sensor placement on a participant for SI task.

From those two-eighteen minute periods, we generated the list of features in Table 2. Based on the visual feature selection procedure, 41 of those features were retained for further analysis. The leave-two-participants-out cross validation resulted in 78 training and test datasets. This is smaller than the datasets used in the first case study since the number of participants with reliable data was

Table 6
Mean performance and the corresponding standard deviation of the classification methods for fatigue detection in SI task, (the recommended model is **in bold**).

Category	Model	Sensitivity	Specificity	Accuracy	G-mean	Consistency	# of Features
	Random Forest	**0.876 (0.12)**	**0.918 (0.10)**	**0.897 (0.08)**	**0.892 (0.09)**	**0.100 (0.13)**	**6.346**
	Bagging	0.863 (0.12)	0.910 (0.10)	0.886 (0.08)	0.882 (0.09)	0.097 (0.13)	6.346
BSS	Boosting	0.868 (0.12)	0.893 (0.12)	0.880 (0.09)	0.875 (0.09)	0.118 (0.13)	6.346
	Support Vector Machine	0.728 (0.19)	0.847 (0.16)	0.787 (0.12)	0.773 (0.13)	0.226 (0.16)	6.346
	Logistic Regression	0.525 (0.28)	0.723 (0.21)	0.624 (0.12)	0.558 (0.19)	0.391 (0.27)	6.346
LASSO	Penalized Logistic Regression*	0.674 (0.19)	0.925 (0.15)	0.800 (0.14)	0.773 (0.15)	0.257 (0.23)	16.179
	Penalized Logistic Regression	0.748 (0.22)	0.824 (0.06)	0.786 (0.10)	0.775 (0.17)	0.151 (0.16)	20.868

* Features used in the model are only those generated in Maman et al. (2017).

smaller. Two hundred bootstrap samples with fixed sample size (11 participants × 18 samples per participant = 198) were used to evaluate the stability of proposed models.

To reduce the computational burden, we only examined the seven models analyzed in case study 1. This means that we did not examine whether the kNN, NB or decision trees performed adequately for this task. The results for using these seven models for fatigue detection are presented in the following subsection.

4.2.2. Fatigue detection results

The predictive performance of the seven models is summarized in Table 6. Similar to Table 4, this table shows the mean (and standard deviation in parentheses) for each of the four performance measures as well as the average number of features selected by each model. The reader should note the reported results are based on 78 constructed test datasets from the two-participants-out cross validation.

There are two main observations to be made pertaining to the results in Table 6. First, the number of features selected with the best subset selection are much less than those selected by the LASSO model. This means that the usability of the analytical models with the BSS model is much higher than that with LASSO since practitioners' need to monitor and understand approximately six features (instead of 16 or 21). Second, the prediction performance of the three ensembles is much higher than all other models. Note that the performance gap is much larger in this task than in the MMH task. Based on this case study, our framework has shown higher detection performance (with less features) when compared to competing models from the literature.

Next, we examine how the prediction performance varies while restricting the number of sensors used when performing SI task. To gage this question, we utilize the random forest model since Table 6 showed that it had the highest prediction performance. Table 7 shows the prediction results when features are limited to those from one, two, three, four and all sensor combinations. Similar to Table 5, the values, which are not shown reflect scenarios when a prediction was not possible.

From the results in Table 7, one can observe that the prediction performance does not vary significantly as the number of sensors' are changed. For instance, the average accuracy varies from 0.854 to 0.897 (with standard deviations ≅ 0.10) as the number of sensors vary. Note that this observation only holds if the heart

Table 7
Mean fatigue detection performance (and the corresponding standard deviation) of the random forest model using different sensor combinations for the SI task (the recommended approach is **in bold**).

# sensors	Sensor Combination					Sensitivity	Specificity	Accuracy	G-mean	Consistency
5	Ankle	Hip	Wrist	Torso	HR	0.876 (0.12)	0.918 (0.10)	0.897 (0.08)	0.892 (0.09)	0.100 (0.13)
4	Ankle	Hip	Wrist	Torso		0.863 (0.13)	0.911 (0.05)	0.887 (0.07)	0.884 (0.08)	0.097 (0.12)
	Ankle	Hip		Torso	HR	0.853 (0.13)	0.893 (0.12)	0.873 (0.09)	0.869 (0.10)	0.107 (0.13)
		Hip	Wrist	Torso	HR	0.853 (0.17)	0.911 (0.14)	0.882 (0.13)	0.866 (0.13)	0.121 (0.14)
	Ankle	Hip	Wrist		HR	0.834 (0.16)	0.921 (0.10)	0.877 (0.10)	0.870 (0.11)	0.132 (0.16)
	Ankle		Wrist	Torso	HR	0.826 (0.19)	0.955 (0.04)	0.890 (0.10)	0.882 (0.12)	0.138 (0.19)
3		Hip	Wrist	Torso		0.867 (0.15)	0.904 (0.12)	0.885 (0.12)	0.872 (0.12)	0.100 (0.09)
	Ankle	Hip			HR	0.856 (0.13)	0.887 (0.15)	0.871 (0.10)	0.865 (0.12)	0.117 (0.15)
		Hip	Wrist		HR	0.831 (0.15)	0.923 (0.06)	0.877 (0.08)	0.872 (0.09)	0.124 (0.14)
	Ankle			Torso	HR	0.825 (0.16)	0.935 (0.05)	0.880 (0.09)	0.874 (0.09)	0.131 (0.14)
			Wrist	Torso	HR	0.818 (0.18)	0.927 (0.12)	0.872 (0.13)	0.856 (0.14)	0.131 (0.15)
		Hip		Torso	HR	0.852 (0.17)	0.874 (0.17)	0.863 (0.13)	0.844 (0.14)	0.132 (0.16)
	Ankle		Wrist		HR	0.820 (0.19)	0.957 (0.03)	0.888 (0.10)	0.879 (0.12)	0.147 (0.19)
	Ankle	Hip	Wrist			–	–	–	–	–
	Ankle	Hip		Torso		–	–	–	–	–
	Ankle		Wrist	Torso		–	–	–	–	–
2				Torso	HR	0.823 (0.16)	0.896 (0.12)	0.859 (0.12)	0.844 (0.13)	0.109 (0.12)
	Ankle				HR	0.828 (0.16)	0.917 (0.07)	0.872 (0.10)	0.867 (0.10)	0.114 (0.13)
		Hip			HR	0.837 (0.15)	0.904 (0.08)	0.870 (0.08)	0.854 (0.14)	0.117 (0.14)
			Wrist		HR	0.818 (0.15)	0.920 (0.05)	0.869 (0.08)	0.863 (0.09)	0.123 (0.14)
		Hip		Torso		–	–	–	–	–
	Ankle		Wrist			–	–	–	–	–
	Ankle	Hip				–	–	–	–	–
	Ankle			Torso		–	–	–	–	–
		Hip	Wrist			–	–	–	–	–
			Wrist	Torso		–	–	–	–	–
1					**HR**	**0.820 (0.14)**	**0.889 (0.06)**	**0.854 (0.08)**	**0.850 (0.09)**	**0.102 (0.13)**
	Ankle					–	–	–	–	–
		Hip				–	–	–	–	–
			Wrist			–	–	–	–	–
				Torso		–	–	–	–	–

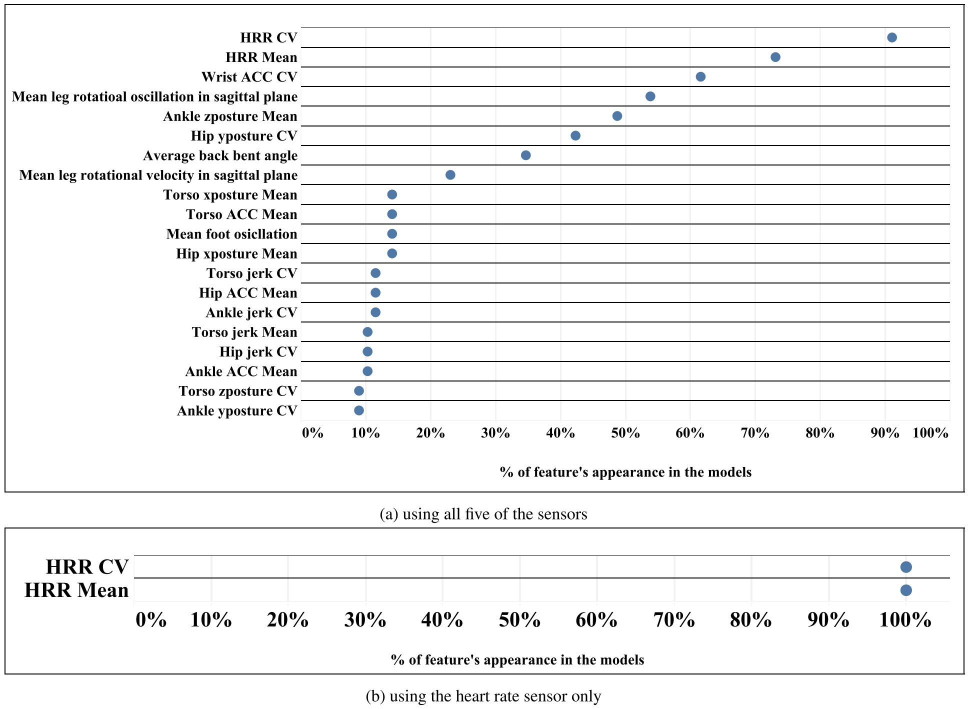

Fig. 7. Important features visualization in the SI task using the Random Forest model.

rate sensor is included in the analysis. Accordingly, using solely the heart rate sensor is appropriate for detecting fatigue in supply pick up and insertion environments (that are similar to those analyzed in our case study). Similar to the earlier case study, this is a novel contribution (showcasing that one sensor can present similar performance to multiple sensors with a much higher usability).

4.2.3. Fatigue identification results

As in case study 1, we follow the two-step approach for fatigue identification. First, we examine how frequently a feature is selected from all of the 78 two-participants-out cross validation random forest model test sets. We limit the analysis to two cases: (a) when all five sensors are utilized, and (b) when only the heart rate sensor is used. The results from these analyses are shown in Fig. 7(a) and (b), respectively. Neither cases included any individual features (which is different from the earlier case when age appeared in both). Only statistical features were selected in the one sensor model, which is perhaps not surprising since none of the biomechanical features can be generated if only the heart rate sensor is used. For the five sensor case, one biomechanical feature (*mean leg rotational oscillation in sagittal plane*), i.e. feature #51 in Table 2 and three statistical features appeared in more than 50% of the models. On the other hand, in the single sensor case, all the statistical features (*HRR CV, HRR Mean*) created using the heart sensor were selected in 100% of the models.

Second, we investigate how those features range as participants transition from the non-fatigued to fatigued states. Fig. 8 illustrates this analysis (using the median model sorted by accuracy) for: (a) the five sensors, and (b) the one sensor cases. Recall that the lines graphed in these plots represent the average values per variable for each of the two participants in the test set examined by the median model. The conclusion is similar to that of case study 1, where only one feature had different values for the non-fatigue (**gray line**) and fatigue cases (**black line**) across the two test participants. However, here, this effect is only observed for the one sensor case. Specifically, in Fig. 8(b), the mean HRR is higher in the fatigued state. This result makes sense since an increased heart rate is a fatigue symptom (see Cavuoto & Megahed, 2017 for more details).

4.2.4. Fatigue diagnosis results

From the fatigue identification results, one can conclude that the participants experience whole-body fatigue in the SI task. This conclusion is based on the ability to accurately detect the non-fatigue and fatigue states through the use of only the heart rate sensor. The elevated mean percent HRR shown for both participants in Fig. 8(b) supports this conclusion.

5. Discussion and conclusions

5.1. Summary of the main contributions

In this paper, we proposed an integrated framework for managing fatigue (and consequently changes in work performance) using minimally-intrusive wearable sensors. Based on the case studies in Section 4, this study makes four main contributions. First, we demonstrated the capability of using a unified modeling approach for managing physical fatigue in different occupational tasks/settings. The case studies show the ability to detect, identify, and diagnose fatigue in multiple occupationally-relevant settings. The ability to identify/diagnose fatigue through the use of wearable sensors has not been shown prior in the literature. Second, the insights from the *fatigue identification* phase of our framework can be used to inform sensor placement and selection. We demonstrated that the prediction performance using one sensor is equivalent to that of using all sensors for our two case studies. Third, we showed that the importance of different types of features (statistical summaries of the sensors' profiles, biomechanical features, and individual characteristics of workers) varies with different manufacturing tasks. Thus, researchers and practitioners should consider this finding when developing models for detecting/managing fatigue in other settings. Fourth, from an intelligent systems perspective, this study has presented a modified leave *p*-participants out cross validation approach (see Section 3.1.3) to account for the inherent au-

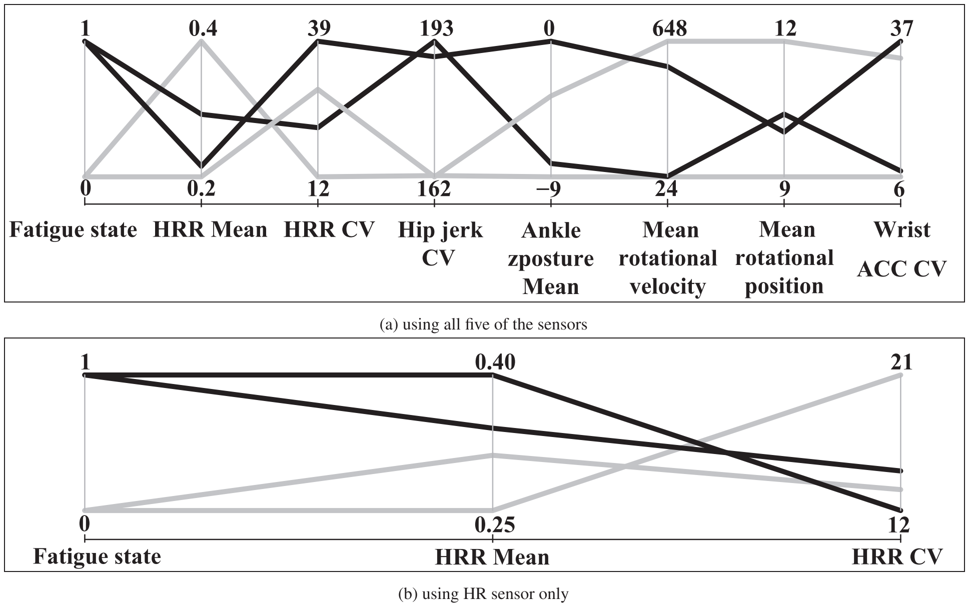

Fig. 8. Features visualization for the median Random Forest model in the SI task.

tocorrelation within each participant's experimental data. To capitalize on the advantages of the standard 10-fold cross validation approach, we have recommended leaving 10% of the participants data out. This corresponded to leaving $p = 2$ participants out in our case studies. While this required a large number of computational experiments, the standard errors of the performance metrics were smaller than those obtained when $p = 1$. For this reason, the large number of simulations was justified in our case. Furthermore, we suggest that researchers should examine $p > 1$ values in future studies.

5.2. Relevance to expert and intelligent systems research and practice

Our framework attempts to bridge the gaps between predictive and prescriptive analytics in the context of human performance modeling. The majority of current expert and intelligent systems research have focused on either the predictive (Khan, Schiøler, Kulahci, Zaki, & Rasmussen, 2019; Lu, Wang, & Yoon, 2019a; Maldonado, López, Jimenez-Molina, & Lira, 2020; Weng et al., 2017; Weng et al., 2018) or prescriptive (Chai & Ngai, 2019; de Leoni & Marrella, 2017; Rezaeiahari & Khasawneh, 2020; Sadic, de Sousa, & Crispim, 2018) modeling components. Perhaps, more importantly there is limited work that have considered the impact of "humans" (especially on an individual level) on the overall performance of a firm (Grosse et al., 2017). Our proposed framework provides a novel approach to detect, diagnose and intervene when fatigue occurs, which is a known precursor of poor performance as shown in the discussion in Section 2. Thus, our proposed framework establishes a successful precedent that can inform the development of more advanced "human-in-the-loop" systems, where the effect of human operators is both predicted/modeled and incorporated into prescriptive decision-making models.

From an expert systems design perspective, the sequential nature of our framework attempts to overcome the "black box" nature of many machine learning algorithms. We have shown that the sequential application of predictive models when combined with visual analytic tools can provide insights for prescriptive interventions. Furthermore, this study demonstrates that futuristic intelligent systems can capture in real-time the well-being of human operators in addition to the data typically captured on the equipment. This can allow for more dynamic operational interventions (e.g., work-rest scheduling models), where the distribution of work tasks between different human operators (and possibly robotic assistants) is optimized.

5.3. Relevance to "human performance" management practice

In our estimation, the proposed framework and the case study findings have significant implications for practitioners interested in managing/optimizing their workforce as a part of larger set of resources that include machines and supporting computational technologies. We have shown that changes in a worker's physical performance can be detected and modeled using wearable sensors. Utilizing the principles behind the technology adoption model (TAM), we have shown that fatigue associated specialized jobs can be detected using one sensor (without a loss in prediction performance). The emphasis on fatigue identification and diagnosis through visual analytical approaches allows practitioners to identify the risks, which are to be tackled through an appropriate intervention strategy. In essence, our framework can provide near real-time insights into the well-begin of shop-floor workers and their associated productivity levels. This information can be incorporated into the safety and productivity components of the SQDCM (safety, quality, delivery, cost, and morale) lean production effectiveness dashboard.

Our case-study findings have significant implications for manufacturing occupations, as they are likely to encourage the management to invest in data-driven manufacturing to develop better plans to prevent fatal and non-fatal occupational injury. The fatigue detection phase of the proposed framework can be used for work scheduling practice as well, since the scheduling approaches should incorporate the fatigue status of the workers. The reader is refereed to Mossa, Boenzi, Digiesi, Mummolo, and Romano (2016) for an example of how ergonomic risk can be incorporated in scheduling.

5.4. Limitations and suggestions for future research

There are a few limitation that may influence the interpretation of our results. First, the sample sizes are small as a result of time committed by each participant. Second, the participants for our two case studies varied in age and experience. Some of them represented a convenience sample of undergraduate and graduate students who may have a limited experience with manufacturing operations. Others were recruited from industry, and as such, are much more experienced/trained. Thus, our 10 min training window may not be sufficient for some participants, i.e. the baseline performance for the non-fatigued state may not reflect their true steady-state performance. Third, the fatigue detection models are based on the participants' perceived ratings of exertion. Different participants may have varying levels of pain tolerance. Thus, we implicitly assume that the aliasing of perception and fatigue will have the same effect on performance as fatigue alone. This assumption is reasonable based on the ergonomics literature. Specifically, Mehta and Cavuoto (2015, p. 94) state that "... muscle activation, perception of discomfort, and/or motivation, might have a greater contribution to fatigue development than peripheral factors". Fourth, the evaluation of our framework's performance was limited to focused lab experiments. Future studies should evaluate how this framework performs in the field.

In our estimation, there are three main streams of research that can capitalize on our framework and findings. First, studies should investigate how our framework can be extended to simultaneously monitor and manage fatigue for hundreds of workers. While our current prediction performance is excellent for an individual worker (and for typical predictive modeling applications in the literature), it will suffer from a high false alarm rate if implemented across the shop-floor. To alleviate this issue, future research should consider: (a) reducing the frequency of data collection, which would increase the average time (but not samples) between false alarms; and (b) controlling the false discovery rate (Benjamini & Hochberg, 1995), which is designed for testing multiple hypotheses. Second, there are several information systems, ethical and legal implications that arise from collecting workers' performance data. Policies that account for these implications are needed. Third, there is an excellent opportunity for optimization models that can optimize recovery (or alternatively minimize fatigue development) while meeting the demands of the production schedule and the resource constraints. Such models will benefit from the data-driven/real-time nature of our framework.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Credit authorship contribution statement

Zahra Sedighi Maman: Methodology, Software, Data curation, Formal analysis, Writing - original draft. **Ying-Ju Chen:** Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing. **Amir Baghdadi:** Investigation, Data curation, Visualization. **Seamus Lombardo:** Software, Visualization. **Lora A. Cavuoto:** Conceptualization, Methodology, Supervision, Funding acquisition, Writing - review & editing. **Fadel M. Megahed:** Conceptualization, Methodology, Supervision, Funding acquisition, Writing - review & editing.

Acknowledgments

This work was supported in part by ASSP Foundation Grant (titled "ASSIST: Advancing Safety Surveillance using Individualized Sensor Technology"), Ohio Supercomputer Center Grant PMIU 0138, and NSF CMMI-1635927. Dr. Megahed's work was also partially supported by the *Neil R. Anderson Endowed Assistant Professorship* at Miami University.

Supplementary material

In this study, the **R** programming language and MATLAB were used to generate the results. Our raw data, code and result files are available in the following repository: https://github.com/zahrame/FatigueManagement.github.io (which can be cited using the DOI: https://doi.org/10.5281/zenodo.3692767).

Supplementary material associated with this article can also be found, in the online version, at doi:10.1016/j.eswa.2020.113405

References

Afsar, P., Cortez, P., & Santos, H. (2015). Automatic visual detection of human behavior: A review from 2000 to 2014. *Expert Systems with Applications, 42*(20), 6935–6956.

Algamal, Z. Y., & Lee, M. H. (2015). Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. *Expert Systems with Applications, 42*(23), 9326–9332.

Argon, N. T., & Ziya, S. (2009). Priority assignment under imperfect information on customer type identities. *Manufacturing & Service Operations Management, 11*(4), 674–693.

Bächlin, M., Förster, K., & Tröster, G. (2009). Swimmaster: a wearable assistant for swimmer. In *Proceedings of the eleventh international conference on ubiquitous computing* (pp. 215–224). ACM.

Baghdadi, A., Cavuoto, L. A., & Crassidis, J. L. (2018a). Hip and trunk kinematics estimation in gait through Kalman filter using IMU data at the ankle. *IEEE Sensors Journal, PP*(99), 1–8. doi:10.1109/JSEN.2018.2817228.

Baghdadi, A., Cavuoto, L. A., Jones-Farmer, A., Rigdon, S. E., Esfahani, E. T., & Megahed, F. M. (2019). Monitoring worker fatigue using wearable devices: A case study to detect changes in gait parameters. *Journal of Quality Technology*, 1–25.

Baghdadi, A., Maman, Z. S., Lu, L., Cavuoto, L. A., & Megahed, F. M. (2017). Effects of task type, task duration, and age on body kinematics and subjective fatigue. In *Proceedings of the human factors and ergonomics society annual meeting: 61*. SAGE Publications Sage CA: Los Angeles, CA. 1040–1040

Baghdadi, A., Megahed, F. M., Esfahani, E. T., & Cavuoto, L. A. (2018b). A machine learning approach to detect changes in gait parameters following a fatiguing occupational task. *Ergonomics*, (just-accepted), 1–21. doi:10.1080/00140139.2018.1442936.

Bao, L. (2003). *Physical activity recognition from acceleration data under semi-naturalistic conditions*. Massachusetts Institute of Technology Ph.D. thesis..

Bao, L., & Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. In *Proceedings of the international conference on pervasive computing* (pp. 1–17). Springer.

Bavaresco, M. V., D'Oca, S., Ghisi, E., & Lamberts, R. (2019). Technological innovations to assess and include the human dimension in the building-performance loop: A review. *Energy and Buildings*, 109365.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence, 97*(1), 245–271.

Bonato, P., Ebenbichler, G. R., Roy, S. H., Lehr, S., Posch, M., Kollmitzer, J., & Della Croce, U. (2003). Muscle fatigue and fatigue-related biomechanical changes during a cyclic lifting task. *Spine, 28*(16), 1810–1820.

Borg, G. (1998). *Borg's perceived exertion and pain scales.*. Human kinetics.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Cai, X., Gong, J., Lu, Y., & Zhong, S. (2018). Recover overnight? Work interruption and worker productivity. *Management Science, Preprint*. doi:10.1287/mnsc.2017.2792.

Catapult Sports (2018). World leader in sports technology for elite sports. http://www.catapultsports.com/. Online; accessed 1-June-2018.

Cavuoto, L., & Megahed, F. (2017). Understanding fatigue: Implications for worker safety. *Professional Safety*, (December), 16–19.

Ceschi, A., Demerouti, E., Sartori, R., & Weller, J. (2017). Decision-making processes in the workplace: How exhaustion, lack of resources and job demands impair them and affect performance. *Frontiers in Psychology, 8*, 313.

Chai, J., & Ngai, E. W. (2019). Decision-making techniques in supplier selection: Recent accomplishments and what lies ahead. *Expert Systems with Applications*, 112903.

Charbonnier, S., Roy, R. N., Bonnet, S., & Campagne, A. (2016). Eeg index for control operatorsâ mental fatigue monitoring using interactions between brain regions. *Expert Systems with Applications, 52*, 91–98.

Chiang, L. H., Russell, E. L., & Braatz, R. D. (2000). *Fault detection and diagnosis in industrial systems*. Springer-Verlag.

Chow, D. H., Man, J. W., Holmes, A. D., & Evans, J. H. (2004). Postural and trunk muscle response to sudden release during stoop lifting tasks before and after fatigue of the trunk erector muscles. *Ergonomics, 47*(6), 607–624.

Côté, J. N., Mathieu, P. A., Levin, M. F., & Feldman, A. G. (2002). Movement reorganization to compensate for fatigue during sawing. *Experimental Brain Research, 146*(3), 394–398.

Dag, A., Topuz, K., Oztekin, A., Bulur, S., & Megahed, F. M. (2016). A probabilistic data-driven framework for scoring the preoperative recipient-donor heart transplant survival. *Decision Support Systems, 86*, 1–12.

Daugherty, P. R., & Wilson, H. J. (2018). *Human+ machine: Reimagining work in the age of AI*. Harvard Business Press.
Davidson, B. S., Madigan, M. L., & Nussbaum, M. A. (2004). Effects of lumbar extensor fatigue and fatigue rate on postural sway. *European Journal of Applied Physiology, 93*(1–2), 183–189.
Dhurasevic, M., & Jakobovic, D. (2018). A survey of dispatching rules for the dynamic unrelated machines environment. *Expert Systems with Applications, 113*, 555–569.
Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Proceedings of the International workshop on multiple classifier systems* (pp. 1–15). Springer.
Dolan, P., & Adams, M. (1998). Repetitive lifting tasks fatigue the back muscles and increase the bending moment acting on the lumbar spine. *Journal of Biomechanics, 31*(8), 713–721.
Ebenbichler, G. R., Bonato, P., Roy, S. H., Lehr, S., Posch, M., Kollmitzer, J., & Della, U. C. (2002). Reliability of emg time-frequency measures of fatigue during repetitive lifting.. *Medicine and Science in Sports and Exercise, 34*(8), 1316–1323.
Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap: Monographs on statistics and applied probability*: 57. New York and London: Chapman and Hall/CRC.
Eisman, E. M., Navarro, M., & Castro, J. L. (2016). A multi-agent conversational system with heterogeneous data sources access. *Expert Systems with Applications, 53*, 172–191.
Ferjani, A., Ammar, A., Pierreval, H., & Elkosantini, S. (2017). A simulation-optimization based heuristic for the online assignment of multi-skilled workers affected by fatigue in manufacturing systems. *Computers & Industrial Engineering, 112*, 663–674.
Fontes, E., Smirmaul, B., Nakamura, F., Pereira, G., Okano, A., Altimari, L., … Moraes, A. d., et al. (2010). The relationship between rating of perceived exertion and muscle activity during exhaustive constant-load cycling. *International Journal of Sports Medicine*.
Foster, C. (1998). Monitoring training in athletes with reference to overtraining syndrome. *Medicine and Science in Sports and Exercise, 30*, 1164–1168.
Fukuda, K., Straus, S. E., Hickie, I., Sharpe, M. C., Dobbins, J. G., & Komaroff, A. (1994). The chronic fatigue syndrome: A comprehensive approach to its definition and study. *Annals of Internal Medicine, 121*(12), 953–959.
Ghaderyan, P., Abbasi, A., & Saber, S. (2018). A new algorithm for kinematic analysis of handwriting data; towards a reliable handwriting-based tool for early detection of alzheimer's disease. *Expert Systems with Applications, 114*, 428–440.
Grosse, E. H., Glock, C. H., & Neumann, W. P. (2017). Human factors in order picking: A content analysis of the literature. *International Journal of Production Research, 55*(5), 1260–1276.
Hallemans, A., Beccu, S., Van Loock, K., Ortibus, E., Truijen, S., & Aerts, P. (2009). Visual deprivation leads to gait adaptations that are age-and context-specific: I. step-time parameters. *Gait & Posture, 30*(1), 55–59.
Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. Elsevier.
Helbostad, J. L., Leirfall, S., Moe-Nilssen, R., & Sletvold, O. (2007). Physical fatigue affects gait characteristics in older persons. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 62*(9), 1010–1015.
Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*: 398. John Wiley & Sons.
Iskander, J., Hossny, M., & Nahavandi, S. (2018). A review on ocular biomechanic models for assessing visual fatigue in virtual reality. *IEEE Access, Preprint*(99). doi:10.1109/ACCESS.2018.2815663. 1–1
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*: 112. Springer.
Kajimoto, O. (2008). In Y. Watanabe, B. Evengård, B. H. Natelson, L. A. Jason, & H. Kuratsune (Eds.), *Development of a method of evaluation of fatigue and its economic impacts* (pp. 33–46)). Tokyo: Springer Japan.
Karg, M., Kühnlenz, K., Buss, M., Seiberl, W., Tusker, F., Schmeelk, M., & Schwirtz, A. (2008). Expression and automatic recognition of exhaustion in natural walking. In K. Blashki (Ed.), *Proceedings of the international conference interfaces and human computer interaction* (pp. 165–172).
Karg, M., Venture, G., Hoey, J., & Kulic, D. (2014). Human movement analysis as a measure for fatigue: A hidden Markov-based approach. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 22*(3), 470–481.
Kavanagh, J. J., Morrison, S., & Barrett, R. S. (2006). Lumbar and cervical erector spinae fatigue elicit compensatory postural responses to assist in maintaining head stability during walking. *Journal of Applied Physiology, 101*(4), 1118–1126.
Kent-Braun, J. A., Ng, A. V., Doyle, J. W., & Towse, T. F. (2002). Human skeletal muscle responses vary with age and gender during fatigue due to incremental isometric exercise. *Journal of Applied Physiology, 93*(5), 1813–1823.
Khan, A. R., Schiøler, H., Kulahci, M., Zaki, M., & Rasmussen, P. W. (2019). Predictive manufacturing: A classification strategy to predict product failures. *Expert Systems with Applications*.
King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis, 9*(2), 137–163.
Koesdwiady, A., Soua, R., Karray, F., & Kamel, M. S. (2017). Recent trends in driver safety monitoring systems: state of the art and challenges. *IEEE Transactions on Vehicular Technology, 66*(6), 4550–4563.
Kong, X. T., Luo, H., Huang, G. Q., & Yang, X. (2018). Industrial wearable system: The human-centric empowering technology in industry 4.0. *Journal of Intelligent Manufacturing*, 1–17.
Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering, 160*, 3–24.
Kumar, S., & Mital, A. (2017). *Electromyography in ergonomics*. Routledge.
Larivière, C., Gagnon, D., & Loisel, P. (2000). The comparison of trunk muscles emg activation between subjects with and without chronic low back pain during flexion–extension and lateral bending tasks. *Journal of electromyography and kinesiology, 10*(2), 79–91.
Lee, M., Roan, M., Smith, B., & Lockhart, T. E. (2009). Gait analysis to classify external load conditions using linear discriminant analysis. *Human Movement Science, 28*(2), 226–235.
de Leoni, M., & Marrella, A. (2017). Aligning real process executions and prescriptive process models through automated planning. *Expert Systems with Applications, 82*, 162–183.
Levenson, A. (2017). Workplace fatigue is a systems problem.. *Consulting Psychology Journal: Practice and Research, 69*(2), 130.
Loriol, M. (2017). A sociological stance on fatigue and tiredness: social inequalities, norms and representations. *Neurophysiologie Clinique/Clinical Neurophysiology, 47*(2), 87–94.
Lu, H., Wang, H., & Yoon, S. W. (2019a). A dynamic gradient boosting machine using genetic optimizer for practical breast cancer prognosis. *Expert Systems with Applications, 116*, 340–350.
Lu, L., Megahed, F. M., & Cavuoto, L. A. (2019b). Interventions to mitigate fatigue induced by physical work: A systematic review of research quality and levels of evidence for intervention efficacy. *Human Factors*. 0018720819876141
Lu, L., Megahed, F. M., Sesek, R. F., & Cavuoto, L. A. (2017). A survey of the prevalence of fatigue, its precursors and individual coping mechanisms among us manufacturing workers. *Applied Ergonomics, 65*, 139–151.
Maldonado, S., López, J., Jimenez-Molina, A., & Lira, H. (2020). Simultaneous feature selection and heterogeneity control for svm classification: An application to mental workload assessment. *Expert Systems with Applications, 143*, 112988.
Maman, Z. S., Baghdadi, A., Megahed, F., & Cavuoto, L. (2016). Monitoring and change point estimation of normal (in-control) and fatigued (out-of-control) state in workers. In *Proceedings of the ASME 2016 international design engineering technical conferences and computers and information in engineering conference*. American Society of Mechanical Engineers. doi:10.1115/DETC2016-60487. pp. V003T11A011–V003T11A011
Maman, Z. S., Lu, L., Megahed, F. M., & Cavuoto, L. A. (2019). A DMAIC perspective on physical fatigue management. *Professional Safety, 64*(6), 26–27.
Maman, Z. S., Yazdi, M. A. A., Cavuoto, L. A., & Megahed, F. M. (2017). A data-driven approach to modeling physical fatigue in the workplace using wearable sensors. *Applied Ergonomics, 65*, 515–529.
Manuele, F. A. (2005). Risk assessment & hierarchies of control. *Professional Safety, 50*(5), 33–39.
Marangunić, N., & Granić, A. (2015). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society, 14*(1), 81–95.
Mehta, R. K., & Cavuoto, L. A. (2015). The effects of obesity, age, and relative workload levels on handgrip endurance. *Applied ergonomics, 46*, 91–95.
Mital, A. (2017). *Guide to manual materials handling*. CRC Press.
Montenegro, J. L. Z., da Costa, C. A., & da Rosa Righi, R. (2019). Survey of conversational agents in health. *Expert Systems with Applications*.
Moon, J., Kwon, Y., Park, J., & Yoon, W. C. (2019). Detecting user attention to video segments using interval eeg features. *Expert Systems with Applications, 115*, 578–592.
Mossa, G., Boenzi, F., Digiesi, S., Mummolo, G., & Romano, V. (2016). Productivity and ergonomic risk in human based production systems: A job-rotation scheduling model. *International Journal of Production Economics, 171*, 471–477.
Naranjo-Flores, A. A., & Ramírez-Cárdenas, E. (2014). In J. L. García-Alcaraz, A. A. Maldonado-Macías, & G. Cortes-Robles (Eds.), *Human factors and ergonomics for lean manufacturing applications* (pp. 281–299)). Cham: Springer International Publishing.
National Science Foundation (2019). Future of work at the human-technology frontier: Core research (FW-HTF). https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505620&org=NSF. Online; last accessed 5-January-2020.
Ødegaard, F., & Roos, P. (2014). Measuring the contribution of workers' health and psychosocial work-environment on production efficiency. *Production and Operations Management, 23*(12), 2191–2208.
Oneto, L., Navarin, N., Donini, M., & Anguita, D. (2018). Emerging trends in machine learning: Beyond conventional methods and data.. *Esann*.
Pacaux-Lemoine, M.-P., Berdal, Q., Enjalbert, S., & Trentesaux, D. (2018). Towards human-based industrial cyber-physical systems. In *Proceedings of the 2018 IEEE industrial cyber-physical systems (icps)* (pp. 615–620). IEEE.
Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R., Moret, B. M., & Stamatakis, A. (2009). How many bootstrap replicates are necessary? In *Proceedings of the annual international conference on research in computational molecular biology* (pp. 184–200). Springer.
Pirttikangas, S., Fujinami, K., & Nakajima, T. (2006). Feature selection and activity recognition from wearable sensors. In *Proceedings of the international symposium on ubiquitious computing systems* (pp. 516–527). Springer.
Quagliarella, L., Sasanelli, N., & Belgiovine, G. (2008). An interactive fall and loss of consciousness detector system. *Gait & Posture, 28*(4), 699–702.
Rea, B. (2018). Ai, robotics, and automation: Putting humans in the loop. Deloitte. Dbriefs Webcast. https://www2.deloitte.com/us/en/pages/dbriefs-webcasts/events/october/2018/dbriefs-ai-robotics-automation-putting-humans-in-loop.html. Online. Last accessed 5-January-2020.
Rescio, G., Leone, A., & Siciliano, P. (2018). Supervised machine learning scheme for electromyography-based pre-fall detection system. *Expert Systems with Applications, 100*, 95–105.

Rezaeiahari, M., & Khasawneh, M. T. (2020). Simulation optimization approach for patient scheduling at destination medical centers. *Expert Systems with Applications, 140*, 112881.
Ricci, J. A., Chee, E., Lorandeau, A. L., & Berger, J. (2007). Fatigue in the us workforce: Prevalence and implications for lost productive work time. *Journal of Occupational and Environmental Medicine, 49*(1), 1–10.
Ryu, J., & Kim, D.-H. (2017). Real-time gait subphase detection using an emg signal graph matching (esgm) algorithm based on emg signals. *Expert Systems with Applications, 85*, 357–365.
Sadic, S., de Sousa, J. P., & Crispim, J. A. (2018). A two-phase milp approach to integrate order, customer and manufacturer characteristics into dynamic manufacturing network formation and operational planning. *Expert Systems with Applications, 96*, 462–478.
Saraiva, R., Perkusich, M., Silva, L., Almeida, H., Siebra, C., & Perkusich, A. (2016). Early diagnosis of gastrointestinal cancer by using case-based and rule-based reasoning. *Expert Systems with Applications, 61*, 192–202.
Schall, M. C., Sesek, R. F., & Cavuoto, L. A. (2018). Barriers to the adoption of wearable sensors in the workplace: A survey of occupational safety and health professionals. *Human Factors*. Online First
Schapire, R. E. (2003). The boosting approach to machine learning: An overview. In *Nonlinear estimation and classification* (pp. 149–171). Springer.
Strohrmann, C., Harms, H., Kappeler-Setz, C., & Troster, G. (2012). Monitoring kinematic changes with fatigue in running using body-worn sensors. *IEEE Transactions on Information Technology in Biomedicine, 16*(5), 983–990.
The White House (2016). Advanced manufacturing: A snapshot of priority technology areas across the federal government, product of the subcomittee for advanced manufacturing of the national science and technology council, executive office of the president of the United States. https://www.whitehouse.gov/sites/whitehouse.gov/files/images/Blog/NSTC%20SAM%20technology%20areas%20snapshot.pdf. Online; last accessed 6-April-2018.
Thomas, S., Reading, J., & Shephard, R. J. (1992). Revision of the physical activity readiness questionnaire (par-q).. *Canadian Journal of Sport Sciences, 17*(4), 338–345.
Venugopal, G., Navaneethakrishna, M., & Ramakrishnan, S. (2014). Extraction and analysis of multiple time window features associated with muscle fatigue conditions using semg signals. *Expert Systems with Applications, 41*(6), 2652–2659.
Vries, J., Koster, R., & Stam, D. (2016). Safety does not happen by accident: Antecedents to a safer warehouse. *Production and Operations Management, 25*(8), 1377–1390.
Waldeck, N. (2014). Advanced manufacturing technologies and workforce development. *Advanced Manufacturing Technologies and Workforce Development*. Taylor & Francis.
Wang, L. (2016). Machine learning in big data. *International Journal of Advances in Applied Sciences, 4*(4), 117–123.
Wang, Y., Jiang, X., Lee, L. H., Chew, E. P., & Tan, K. C. (2017). Tree based searching approaches for integrated vehicle dispatching and container allocation in a transshipment hub. *Expert Systems with Applications, 74*, 139–150.
Weng, B., Ahmed, M. A., & Megahed, F. M. (2017). Stock market one-day ahead movement prediction using disparate data sources. *Expert Systems with Applications, 79*, 153–163.
Weng, B., Lu, L., Wang, X., Megahed, F. M., & Martinez, W. (2018). Predicting short--term stock prices using ensemble methods and online data sources. *Expert Systems with Applications, 112*, 258–273.
Willson, J. D., & Kernozek, T. W. (1999). Plantar loading and cadence alterations with fatigue. *Medicine and Science in Sports and Exercise, 31*(12), 1828–1833.
Wojcik, L. A., Nussbaum, M. A., Lin, D., Shibata, P. A., & Madigan, M. L. (2011). Age and gender moderate the effects of localized muscle fatigue on lower extremity joint torques used during quiet stance. *Human Movement Science, 30*(3), 574–583.
Yakowicz, W. (2016). Robots are coming for dirty, dangerous, and dull jobs. Inc., http://www.inc.com/will-yakowicz/robots-replacing-dirty-dangerous-jobs-is-good.html. Online; accessed 6-April-2018.
Yoshino, K., Motoshige, T., Araki, T., & Matsuoka, K. (2004). Effect of prolonged free-walking fatigue on gait and physiological rhythm. *Journal of Biomechanics, 37*(8), 1271–1280.
Young, J. G., Trudeau, M. B., Odell, D., Marinelli, K., & Dennerlein, J. T. (2013). Wrist and shoulder posture and muscle activity during touch-screen tablet use: Effects of usage configuration, tablet type, and interacting hand. *Work, 45*(1), 59–71.
Yun, X., Bachmann, E. R., Moore, H., & Calusdian, J. (2007). Self-contained position tracking of human movement using small inertial/magnetic sensor modules. In *Proceedings of the 2007 IEEE international conference on robotics and automation,* (pp. 2526–2533). IEEE.
Yung, M. (2016). Fatigue at the workplace: Measurement and temporal development. PhD Dissertation, Department of Kinesiology, University of Waterloo, http://hdl.handle.net/10012/10119. Online; accessed 6-April-2018.
Yung, M., Bigelow, P. L., Hastings, D. M., & Wells, R. P. (2014). Detecting within-and between-day manifestations of neuromuscular fatigue at work: An exploratory study. *Ergonomics, 57*(10), 1562–1573.
Yung, M., Lang, A. E., Stobart, J., Kociolek, A. M., Milosavljevic, S., & Trask, C. (2017). The combined fatigue effects of sequential exposure to seated whole body vibration and physical, mental, or concurrent work demands. *PloS One, 12*(12), e0188468.
Zanzotto, F. M. (2019). Human-in-the-loop artificial intelligence. *Journal of Artificial Intelligence Research, 64*, 243–252.
Zhang, J., Lockhart, T. E., & Soangra, R. (2014). Classifying lower extremity muscle fatigue during walking using machine learning and inertial sensors. *Annals of Biomedical Engineering, 42*(3), 600–612.
Zhao, C., Zheng, C., Zhao, M., Tu, Y., & Liu, J. (2011). Multivariate autoregressive models and kernel learning algorithms for classifying driving mental fatigue based on electroencephalographic. *Expert Systems with Applications, 38*(3), 1859–1865.