Development of a Manual Handling Task Recognition and Holistic Fatigue Estimation System

Victor Chan

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Doctorate of Philosophy in Human Kinetics

School of Human Kinetics
Faculty of Health Sciences
University of Ottawa

Table of Contents

Acknowledgements

I would first like to acknowledge my supervisor, Ryan Graham. Throughout my degree, you have always returned shockingly prompt feedback, yet somehow never lacked thoroughness and insightfulness. You have always shared thoughtful suggestions catered to each of our individual interests, while never being overbearing. And most importantly, you have always provided me with generous support and encouragement, whilst entrusting me with the intellectual freedom to research and develop whatever I wanted. You have been an exceptional supervisor, and I am grateful to have you as a mentor.

Thank you to all my lab mates for the amazing times we've shared in and outside of the laboratory. Whether it be going for a bike ride or crushing card games, sharing drinks or some fresh air, enjoying great music or meals, you have helped remind me, time and time again that, while this degree was important, it certainly did not have to define me. Beyond this, I also need to thank you for sharing your invaluable academic, technological, and bureaucratic wisdom that have undoubtedly saved me countless hours of pain and suffering. A special shout-out to those that have been there since the start: Matt Mavor, Kristen Beange, and Mohammad Akhavanfar. Additionally, a big thank you to Di Kang for your help with the long, gruelling data collection sessions. You are all seriously special, to say the least!

Thank you to my friends who have kept me grounded through the years – you know who you are. The countless hours we have shared through badminton, snowboarding, skateboarding, camping, and bike rides have helped me immensely through this degree. I am grateful for my partner, Bridget Shanks, for your endless love and support. Thank you for encouraging me on all my pursuits, and I look forward to the next chapter in our lives together. And to my dog, Toklo: You have put up with my madness more than anyone and yet still show me unconditional love. I could not have done it without all of you.

And of course, a huge thank you to my mom, Angelina, my dad, Pun Chiu, and my sister, Claudia, for your unwavering support. Although I can sometimes tell you may just be pretending to understand what I have been working on, it means a lot to me that you do your best to stay on up to date on my projects. It is through your care and teachings that I am where I am today, so this achievement is just as much yours as is it mine. Lastly, a thank you to my funding sources: NSERC, OGS, Mitacs, uOttawa, and CRE-MSD.

Co-Authorship

This dissertation includes one published, and three un-published manuscripts. The authorship is as follows:

Chapter 3: **Chan, V. C.**, Ross, G. B., Clouthier, A. L., Fischer, S. L., & Graham, R. B. (2022). The role of machine learning in the primary prevention of work-related musculoskeletal disorders: A scoping review. *Applied Ergonomics*, *98*, 103574.

Chapter 4: **Chan, V. C.**, Graham, R. "*Raw Versus Fused Inertial Data for Machine Learning-Based Recognition of Manual Handling Activities"*. Submitted to IEEE Transactions on Industrial Informatics.

Chapter 5: **Chan, V. C.**, Graham, R. *"Sensor data required for automated activity recognition of manual handling tasks"* Submitted to Human Factors.

Chapter 6: **Chan, V. C.**, Graham, R. "*Continuous Fatigue Estimation During Manual Handling Tasks*". In writing, with plans to submit to Applied Ergonomics.

All items in this thesis were filed in a US provisional patent application:

Chan, V. C., Graham, R. (2024). *"A System and Method for Continuous Fatigue Modelling for Manual Handling in Occupational Settings"* U.S. Patent Application No. 63/668,935.

For each manuscript, Victor C. Chan completed the tasks of conceptualization and design of the research, acquisition of the data, analysis and interpretation of the data, writing the manuscripts, and critical revision of the manuscripts before publication.

Abstract

The overarching goal of this dissertation was to contribute to the body of research on mitigating work-related musculoskeletal disorders (WMSDs) and occupational accident risk by identifying opportunities to leverage machine learning (ML) techniques, then develop an effective intervention strategy. A scoping review on the role of ML in the primary prevention of WMSD was conducted and identified fatigue modelling during physical labour as a notable and novel intervention strategy. Some shortcomings of existing efforts include the dichotomization of fatigue, the over-reliance on self-reported fatigue as a ground truth, and the inability to automatically adapt estimations to various tasks. Thus, this doctoral research aimed to build off these shortcomings and develop the primary components of an automatic system to adapt to the performance of various manual handling tasks using minimal wearable sensors, then estimate holistic fatigue as a continuous variable. To achieve this, four research objectives were formulated:

1) Perform a scoping review to document how ML techniques have been used to improve efforts to mitigate WMSD risk and identify areas for further research and development.
2) Develop human activity recognition (HAR) models that utilized either raw (i.e., linear acceleration, angular velocity, magnetometer) or fused (i.e., orientation and position) IMU data to classify manual handling tasks.
3) Perform a sensitivity analysis of the quantity and placement of IMUs for classification of the manual handling tasks.
4) Develop task-specific ML models to estimate continuous fatigue levels during the performance of the six manual handling tasks, where fatigue was quantified holistically using a composite index.

The scoping review returned 130 English studies published since 1990 regarding the use of ML techniques and were classified as contributions to the six steps of WMSD prevention (van der Beek et al., 2017). ML techniques were found to contribute most commonly to the development of interventions (48 studies; step 4), where a notable intervention approach was fatigue modelling during physical work.

For studies 2 – 4, a simulated, fatiguing manual handling protocol was conducted, where participants performed 5 min manual handling sets of pulling, lifting, carrying, and pushing a crate,

followed by walking and standing. Between each set, they performed a fatigue assessment, then continued to perform manual handling sets and fatigue assessment without breaks until a stopping criterion was reached.

In the second study, HAR models were trained to classify the manual handling tasks with strong performances (accuracy ≥ 94.39%). However, HAR models performed significantly (yet marginally) better when using raw versus fused IMU data ($p < 0.01$). The third study developed 36 HAR models, with 18 sensor combinations and two ML model architectures to provide a sensitivity analysis. A greater number of IMUs lead to greater classification performances, while using one IMU achieved similar classification performance compared to using six (i.e., T8 and 6 IMU accuracy = 93.17% and 94.26%, respectively). Marginally better performances were observed with the deeper neural network (four convolutional and two long short-term memory layers) compared to the shallower neural network (three dense layers) until 3 or more IMUs were involved, after which the shallower neural network was equal or better.

In the fourth study, a holistic fatigue composite index (FCI) comprised of 6 fatigue variables was computed for all participants and reflected an increase in fatigue throughout the manual handling protocol. Then, task-specific ML regression models were trained to estimate the FCI using kinematic data from 1 or 3 IMUs (combinations informed by study 3), with and without heart rate. On average, using only the T8 IMU and heart rate data allowed for estimation of the FCI with small errors (RMSE = 0.06, MAE = 0.05) and moderate correlation ($r = 0.59$) across all tasks and participants.

The contributions of this dissertation include a thorough scoping review, knowledge surrounding the development of HAR models, and a novel approach to estimate task-specific holistic fatigue. This work represents the primary components of an unobtrusive and inexpensive system that can inform workers and employers of near real-time fatigue progression during the performance of manual handling work tasks, to ultimately mitigate WMSD and occupational accident risk.

List of Tables

List of Figures

List of Equations

Definition of Terms

3D	Three dimensional
AD	Anterior deltoid
ANN	Artificial neural network
BB	Biceps brachii
BN	Bayesian network
CART	Classification and regression tree
CNN	Convolutional neural network
COP	Centre of pressure
DeepConvLSTM	Deep convolutional long short-term memory recurrent neural network
DT	Decision tree
ECG	Electrocardiogram
EKG	Electrocardiogram
EMG	Electromyography
FC	Fully connected artificial neural network
FCI	Fatigue composite index
FFT	Fast Fourier transform
FN	False negative
FP	False positive
HCA	Hierarchical cluster analysis
HMM	Hidden Markov model
HR	Heart rate
HRinst	Instantaneous heart rate
HRV	Heart rate variability
IMU	Inertial measurement unit
***k*-NN**	*k*-nearest neighbors
LASSO	Least absolute shrinkage and selection operator
LDA	Linear discriminant analysis
LES	Lumbar erector spinae
LFHFratio	Low frequency high frequency ratio
LSTM	Long short-term memory
MAE	Mean absolute error
MDF	Median frequency
ML	Machine learning
MLP	Multilayer perceptron
MLS	Maximal lift strength
MMH	Manual materials handling
NB	Naïve bayes
NLP	Natural language processing
NN	Neural network
NNI	Normal-to-normal interval
PAR-Q+	Physical activity readiness questionnaire
PCA	Principal component analysis
PCHIP	Piecewise cubic Hermite interpolating polynomial
PD	Posterior deltoid
QDA	Quadratic discriminant analysis
REBA	Rapid entire body assessment
RF	Random forest
RGB	Red green blue
RGB-D	Red green blue with depth

RMS	Root mean square
RMSE	Root mean squared error
RPE	Rating of perceived exertion
RULA	Rapid upper limb assessment
sEMG	Surface electromyography
SGD	Stochastic gradient descent
SGTB	Stochastic gradient tree boosting
SVM	Support vector machine
TB	Triceps brachii
TN	True negative
TP	True positive
VAS	Visual analog scale
VL	Vastus lateralis
VMO	Vastus medialis obliquus
WMSD	Work-related musculoskeletal disorders

For definition of the machine learning terminology, please refer to Appendix A. Machine learning glossary.

*Note that **activity** and **task** have been used interchangeably throughout this dissertation, both referring to an action that have a discrete start and end during the manual handling protocol. The tasks in this work were stand, pull, lift, carry, push, and walk.*

Chapter 1. Introduction

Work-related musculoskeletal disorders (WMSD) account for over a third of all lost workdays and led to $54 billion in costs in the United States of America in 2014 (Kang et al., 2014). Occupational accidents are similarly pervasive in the European Union, with over 2.5 million incidences per year since 2010 (Raenović, 2023). As WMSDs and occupational accidents are immensely detrimental to workers' quality of life and employers' productivity, it is desirable to understand the etiology, then develop and implement interventions and controls to mitigate this risk. Dramatic increases in computing power and data collection capacity have enabled the widespread development of machine learning (ML) models to perform complex analyses and enable intricate systems for use in a variety of fields (Rodríguez et al., 2022). A popular application of ML is the development of interventions to mitigate WMSD and occupational accident risk; it was determined through a scoping review that this was the primary area in which ML has been employed to mitigate WMSD risk (Victor C.H. Chan et al., 2022). A notable method was the use of ML to classify or estimate fatigue status during manual work, because it can be used as a proxy to indirectly predict WMSDs and/or occupational accident risk (Maman et al., 2017).

Fatigue arising from physically-demanding work tasks (e.g., manual handling) can alter movement behaviours (e.g., movement kinematics, affordances in movement solutions; Côté et al., 2005), potentially increasing the risk of developing work-related musculoskeletal disorders (M. F. Antwi-Afari et al., 2017). Fatigue can also increase the risk for occupational accidents (i.e., acute trauma; Namian et al., 2021), potentially through a decline in attention/awareness (Swaen et al., 2003), reduction in force generation capacity (Place et al., 2010), and/or change in movement control (e.g., reduction in kinesthetic awareness, changes in movement variability, affordances in movement solutions; Gribble and Hertel, 2004). Thus, recognizing and subsequently preventing the development of excessive fatigue in the workplace is imperative to mitigate the risk for WMSDs and occupational accidents. However, there is currently limited capability to unobtrusively monitor holistic fatigue progression during manual handling work. Current methods are difficult to implement at work because they are limited to subjective self-reports (Abdel-Malek et al., 2021), strength assessments to measure decline in force generation capacity (V C H Chan et al., 2020), endurance assessments to quantify the duration of force generation (Mehta & Agnew, 2012), analyzing electromyography (EMG) signals (Shair et al., 2017), and spectral analyses of electroencephalography (EEG) to evaluate mental fatigue (Aryal et al., 2017). The ability to unobtrusively monitor the progression of fatigue due to physical exertion in the workplace would enable the optimization of work-rest ratios, scheduling of timely breaks or rotations, and provision of near real-time feedback regarding movement quality, all of which can help mitigate WMSD and accident risk (V C H Chan et al., 2020). Thus, estimating holistic fatigue status solely using data from accessible, wearable devices is desirable because these data can be collected unobtrusively and automatically. For this purpose, kinematic data can be collected using a variety of methods, including marker-based motion capture (Karg et al., 2014), markerless motion capture with (RGB-D) or without (RGB) depth information (Haque et al., 2016), and inertial measurement units (IMUs; Chan et al., 2020). Wearable devices, such as IMUs (Mavor et al., 2020; Robert-Lachaine et al., 2017) and heart rate (HR) monitors (Maman et al., 2017), are often preferred because they can be placed on or underneath clothing/equipment and are relatively inexpensive to implement. However, various challenges exist in developing a system to estimate fatigue during manual handling work, such as the accommodation of different tasks and estimating fatigue as a continuous quantity.

The estimation of fatigue using kinematic data is relatively novel and has primarily relied on the use of ML models to classify a discrete number of states, typically "not fatigued" versus "fatigued" (e.g., Baghdadi et al., 2018; Escobar-Linero et al., 2022; Karvekar et al., 2021; Marotta et al., 2021). Although this simplification is necessary as a starting point, it does not completely capture the nature of fatigue progression because fatigue is a continuous process instead of finite, discrete states (Enoka & Duchateau, 2016). Dichotomization of fatigue also limits the preventative capabilities of a fatigue modelling system. Several important studies have estimated fatigue as a continuous variable. For one, Karg et al. (2014) examined repetitive squats, trained person-specific models for seven individuals, utilized whole-body, optoelectronic kinematic data from 35 marker locations, and estimated subjective ratings using a Hidden Markov model. Maman et al. (2017) recruited eight individuals who donned four IMUs and an HR monitor during the performance of three occupational tasks. Using least absolute shrinkage and selection operator (LASSO) to select task-specific input features, they detected fatigue and estimated subjective ratings using penalized logistic and multiple linear regression models, respectively. Gholami et al. (2020) recruited five runners who wore eight textile strain sensors in their garments to measure kinematics, then used a random forest algorithm to estimate rating of perceived exertion. More recently, Lee et al. (2024) used HR critical power to estimate whole-body fatigue perception in 33 workers during construction using linear regression. While various approaches in these pioneering studies have been proposed to model fatigue as a continuous variable, several limitations still exist.

First, fatigue models driven by kinematics are highly task-specific, yet the proposed methods are not capable of automatically adapting to various tasks that may be performed in real-world contexts thus far. Instead, fatigue models have been presented for one, repetitive task at a time. Second, when fatigue is estimated as a continuous variable, models have primarily estimated self-reported ratings of fatigue as the ground truth. While self-reports can be valid and useful (Christodoulou, 2017), they may be inaccurate without a calibration procedure (Abdel-Malek et al., 2021) and only represent one of many components of fatigability (Enoka & Duchateau, 2016). Lastly, more data may be warranted to improve fatigue estimations and reduce the quantity of input data required. In this dissertation, these aforementioned shortcomings and gaps in existing work were identified through a scoping review. These were then used to guide the development of a novel fatigue modelling approach comprised of human activity recognition (HAR), followed by ML-based estimation (i.e., regression) of a composite index comprised of variables known to be associated with different components of fatigability (Enoka & Duchateau, 2016).

HAR is a data-driven method (e.g., kinematic data from IMUs) to automatically recognize what activity/task is being performed (such as walking or sitting; Nurwulan and Selamaj, 2021), or the technique with which a task was performed (such as stoop versus squat lifting; Hlucny and Novak, 2020) in near real-time. HAR has traditionally been achieved using ML techniques, involving shallow or deep learning approaches (Victor C.H. Chan et al., 2022). Knowing which tasks are being performed and their durations is an important intermediary step in many occupational injury prevention interventions (Victor C.H. Chan et al., 2022). In the context of fatigue modelling, HAR can be used to automatically inform which task-specific fatigue models to employ, as fatigue-related changes in kinematics are task-dependent. However, the methods to develop, train, and implement HAR models using kinematic data from IMUs for manual handling tasks have been inconsistent in literature. Specifically, data from IMU have been used with (Bances et al., 2022; Del Rosario et al., 2019) and without (Khatun et al., 2022, 2023; Mavor et al., 2023) sensor fusion processing to obtain sensor orientation and position, while the quantity and placement of the IMUs have been highly variable, even for similar tasks (e.g., for manual handling HAR, three, eight, and seventeen IMUs have been used on various locations; Bastani et al., 2016; Lee et al., 2020; Yan et al., 2021). Additional research is needed to determine best practices to develop highly accurate HAR models, especially for use with fatigue modelling applications.

Various approaches have been used for fatigue modelling, including detection (classification of fatigue states) and estimation (regression of a continuous fatigue variable). Statistical approaches can be used identify features that are associated with fatigue progression (Bonato et al., 2003; V C H Chan et al., 2020; Maman et al., 2017) and are useful for feature selection purposes. Rule/logic-based methods, particularly methods involving fuzzy logic, are suitable for fatigue detection because of their ability to deal with stochastic and ambiguous biological signals (W. Wang et al., 2020). ML-based approaches are popular for fatigue modelling (either detection or estimation) because they are suitable to model non-linear interactions between numerous variables (Victor C.H. Chan et al., 2022), where deep learning can be used end-to-end to automatically extract features (Lecun et al., 2015). Common approaches are to train ML models using data from numerous, representative individuals and then estimate fatigue level on unseen individuals (i.e., person-independent model; Gholami et al., 2020), or train models and make estimations using data from each separate person (i.e., person-specific models; Karg et al., 2014). While person-specific models can lead to better performance, the need for training data from each new participant represents a barrier to model generalizability and practical application; thus, person-independent models are preferred for real-world implementation (Scheurer et al., 2020). This dissertation employed a person-independent modelling approach to estimate a composite index comprised of various variables relevant to worker fatigue.

Fatigue has been traditionally difficult to define due to the various mechanisms involved and the task dependency of the primary mechanism(s) (Barry & Enoka, 2007). In this dissertation, it has been operationalized as: "…a disabling symptom in which physical and cognitive function is limited by interactions between performance fatigability and perceived fatigability" (Enoka & Duchateau, 2016). This definition accommodates the wide scope of conditions ascribed to fatigue, where the attributes of performance and perceived fatigability should be assessed using outcome variables relevant to the context of interest (Enoka & Duchateau, 2016; Kluger et al., 2013). In the context of fatiguing manual labor, numerous manifestations of fatigue are relevant and not limited to perceived changes in fatigue-related sensations. Workers also often experience a decline in force generation capacity (Seo et al., 2015). They may report increases in heart rate (Hwang et al., 2016) and changes in heart rate variability (HRV; Tran et al., 2009), which provide (in)direct insight into cardiovascular output and autonomic control of heart activity, respectively. Fatigue at the muscle level may alter muscle activity in a manner that can be observed as a leftward shift in EMG frequency content (Ament & Verkerke, 2009; Kallenberg et al., 2007). Jerk, the derivative of acceleration, may also increase as a symptom of fatigue (L. Zhang et al., 2019). As numerous variables that represent components of fatigability have been identified as relevant to manual handling work, this dissertation established a ground truth represented by a composite index of variables that capture the aforementioned fatigue-related changes.

1.1 Thesis statement

The overarching goal of this dissertation was to contribute to the body of research on mitigating WMSD and occupational accident risk by identifying opportunities, addressing gaps in knowledge, and developing an effective intervention strategy. Through an extensive scoping review, key opportunities for leveraging ML techniques were uncovered, guiding the creation of a novel approach to continuously estimate holistic fatigue during various manual handling tasks. To ensure effectiveness, several aspects were critically examined: comparing data processing methods for HAR (Chapter 4), sensitivity analyses to determine the optimal quantity and placement of IMUs (Chapter 5), the introduction of a fatigue composite index (FCI) which integrated self-reported fatigue with five other objective measures of performance and perceived fatiguability (Chapter 6), and the development of task-specific models to estimate the FCI continuously using a minimal number of body-worn sensors (Chapter 6). The eventual goal is to create a system by integrating these components as a cascading ensemble and deploy them for mobile inference. The envisioned use of this system was to unobtrusively monitor workers' fatigue by collecting kinematic data, employing HAR models to classify tasks, and estimating continuous fatigue levels, all in near real-time. When a dramatic increase and/or excessive fatigue is detected, the system can trigger alerts for workers and/or employers, enabling timely interventions such as recommendations for breaks, job rotations, or adjustments in technique, ultimately enhancing safety and reducing the risk of WMSDs and accidents.

1.2 Statement of objectives

1) Conduct a scoping review to determine how ML techniques have been used for primary WMSD prevention in the steps outlined by van der Beek et al. (2017).

    Secondary objective: Identify areas where ML can be effectively leveraged to mitigate WMSD and occupational accident risk.

2) Develop ML models to utilize data from wearable IMUs to automatically recognize manual handling tasks.

    Secondary objective: Compare the performance and training time between models driven by raw IMU signals versus features derived using sensor fusion processing.

3) Perform sensitivity analyses on the quantity and placement of IMUs for automatic recognition of six MMH tasks.

    Secondary objective: Compare the performance of using a shallower (3 hidden layers) versus a deeper (6 hidden layers) artificial neural network on HAR performance.

4) Develop task-specific ML models to estimate continuous fatigue levels during the performance of the six MMH tasks.

    Secondary objective: Develop a fatigue composite index (FCI) to represent a holistic ground truth of fatigue.

Chapter 2. Review of Literature

2.1 Defining fatigue

Fatigue is difficult to define due to the mechanisms involved and the task dependency of the primary mechanism(s) (Barry & Enoka, 2007). Considering the focus on WMSD risk, fatigue was operationally defined as: "…a disabling symptom in which physical and cognitive function is limited by interactions between performance fatigability and perceived fatigability" (Enoka & Duchateau, 2016). To expand, performance fatigability was defined as "…the decline in an objective measure of performance over a discrete period of time" (Kluger et al., 2013) and "…depends on the contractile capabilities of the involved muscles and the capacity of the nervous system to provide an adequate activation signal for the prescribed task" (Enoka & Duchateau, 2016). Perceived fatigability was defined as "…changes in the sensations that regulate the integrity of the performer" (Kluger et al., 2013), and "…is derived from the sensations that regulate the integrity of the performer based on the maintenance of homeostasis and the psychological state of the individual" (Enoka & Duchateau, 2016). Under this definition, fatigue should not be discussed with an adjective (e.g., muscle, central, peripheral, etc.) to avoid suggesting that a specific area of fatigue occurs in isolation (Enoka & Duchateau, 2016).

This definition was selected for several reasons. First, fatigue is considered as one entity, comprised of physical and cognitive function. This is appropriate because the causes of physical and cognitive fatigue are inseparable; for example, afferent signals from muscle nerve fibres influence both spinal motor neuron input (performance fatigability) as well as pain perception (perceived fatigability) (Amann, 2012; Martin et al., 2008). Moreover, this definition emphasizes the interaction between performance and perceived fatigability, important because the primary mechanism(s) causing fatigue are task dependent and may interact with other factors (Barry & Enoka, 2007; Enoka & Duchateau, 2016; Martin et al., 2008). For example, even within one performance metric, maximal voluntary contraction (MVC), the primary mechanism causing decline in contractile force depends on the task and context (Barry & Enoka, 2007; Taylor & Gandevia, 2008; Westerblad et al., 2010). Lastly, this definition decompartmentalizes the knowledge garnered across various disciplines (Enoka & Duchateau, 2016), as it accommodates the wide scope of conditions that have been ascribed to fatigue and acknowledges their interactions to affect performance and perceived fatigability (Kluger et al., 2013). Under this definition, attributes of performance and/or perceived fatigability can be assessed as outcome variables, where assessments should be ecologically valid and relevant to the context(s) of interest (Enoka & Duchateau, 2016).

2.2 Effects of fatigue

Because fatigue changes the sensations that regulate the integrity of the performer (Enoka & Duchateau, 2016), individuals can perceive their level of fatigue. Individuals can rate these fatigue-related sensations which, when tracked using a subjective self-report such as visual analog scale (VAS), increase with fatigue (Aaronson et al., 1999). Another effect of fatigue is the reduction in muscle contractile capability, either decreasing the maximal contractile force and/or causing failure in prolonged, submaximal contractions. This change results from decreased excitation-contraction coupling efficiency (Ament & Verkerke, 2009) and are measured using (MVC) assessments (Zemková et al., 2019).

Fatigue also affects HR, which has been linked with both performance and perceived fatigability. For example, with a constant workload, instantaneous HR increases with physically-induced fatigue (Hwang et al., 2016), representing a change in contractile function. However, HR variability (HRV) using spectral analyses can also provide insight into cognitive fatigue. It is shown that the low-frequency (LF) content of HRV increases with fatigue (Schmitt et al., 2013; Tran et al., 2009), and provides insight into "global" fatigue level by providing an indirect evaluation of the autonomic control of heart activity (Schmitt et al., 2015).

Fatigue-related changes at the muscle level include a decrease in action potential (AP) propagation velocity along the sarcolemmae (Juel, 1988), "muscle wisdom" (Barry & Enoka, 2007), and/or a decrease in motor neuron firing frequency due to intracellular ADP accumulation (Metzger, 1996). These effects can be directly measured using EMG, as fatigue causes a leftward shift in the EMG frequency content (i.e., increase and decrease in lower and higher frequencies, respectively) (Ament & Verkerke, 2009; Kallenberg et al., 2007); mean (MNF) and median frequency (MDF) are commonly used spectral variables to observe this leftward shift (Georgakis et al., 2003; Ortega-Auriol et al., 2018; Phinyomark et al., 2012). Overall, fatigue-related changes in EMG have been well-documented and have been offered as part of an objective definition of fatigue (Enoka & Duchateau, 2008; Kallenberg et al., 2007).

Lastly, fatigue also alters the kinematics of work tasks in ways that can reduce productivity and/or increase the risk for developing MSDs or accidents. Isometrically, fatigue from prolonged standing can adversely affect standing postural control (Gribble & Hertel, 2004; Paillard, 2012), while fatigue from prolonged, seated computer work is associated with increased trunk flexion angle, internal spinal load, and perceived discomfort (Jia & Nussbaum, 2018). Dynamically, fatigue leads carpenters to perform nailing, sawing, and screwing with reduced speed (Hammarskjöld & Harms-Ringdahl, 1992), while it also increases trunk range of motion during walking load carriage (Qu & Yeo, 2011). During repetitive lifting, fatigue alters angular displacements at the knee, hip, trunk and elbow (Bonato et al., 2003; Dolan & Adams, 1998; Resnick, 1996; Sparto et al., 1997), and increases trunk angular velocity (Trafimow et al., 1993). Furthermore, fatigue affects movement variability, which allows for flexibility and adaptability in the face of fatigue (Gaudez et al., 2016). Fatigue can also decrease dynamic stability through various detriments to individuals' ability to respond to perturbations (e.g., decrease proprioception; H. M. Lee et al., 2003). As a result, fatigue has been shown to decrease the dynamic stability of the spine after fatiguing repetitive lifting (Asgari et al., 2017; Granata & Gottipati, 2008). Considering the definition of fatigue selected, variables that represent subjective fatigue, maximum force generation capacity, HR(V), and muscle activity should all be considered when attempting to represent fatigue objectively and holistically. The less direct relationship between kinematics and fatigue can be leveraged to provide unobtrusive insight into fatigue status (V C H Chan et al., 2020).

2.3 Biomechanical measurements of fatigue

Although numerous instruments and methods have been employed to study fatigue, the focus of this subchapter is on commonly-used, objective, biomechanical fatigue assessment methods only, which primarily quantify performance fatigability: MVC assessments, motion capture, and EMG.

2.3.1 Maximal voluntary contraction (MVC)

For decades, MVC assessments were widely used because an early definition of fatigue was "…a reduction in the force-generating capacity of the neuromuscular system that occurs during sustained activity" (Bigland Ritchie et al., 1983). Thus, quantification of the force-generating capacity was imperative to quantify the reduction. Today, MVC assessments are still used to assess performance fatigability. Although methodologies vary, MVC assessments are often isometric, and the maximal force or torque is measured using load cells (EMG may also be collected; Chapter 2.3.3). Experiments tend to allow three attempts per MVC trial, where the peak or average peak of the attempts are recorded. Participants are often asked to ramp up the force they produce and hold for 2-3 s to minimize the effect of momentum and to ensure that the force output plateaus (Brutsaert et al., 2003). Various modalities are used, depending on the body region(s) of interest. Common modalities include: assessing lumbar extensor strength using a roman chair (Herrmann et al., 2006), assessing elbow flexor strength using a Biodex or force transducer (Abdel-Malek et al., 2021; Gauche et al., 2009; Oda & Kida, 2001), measuring shoulder flexion and/or abduction torque with a Biodex or torque transducer (Gerdle et al., 1993; Hagberg, 1981; Looft & Frey-Law, 2020; Mehta & Agnew, 2012; Yassierli et al., 2007), leg press to assess leg strength (Rube & Secher, 1991), and knee extensions to assess quadriceps strength (Brutsaert et al., 2003; Hunter et al., 2002). The main benefits of MVCs are that force-generation capacity can be directly assessed, the rate of change of MVC force can be assessed, and the force, torque and/or EMG measurements can be used for normalization (Enoka & Duchateau, 2016). A limitation of MVC assessments is that fatigue can be induced by performing them (Abdel-Malek et al., 2021), which can be a confounding factor in experiments. Furthermore, MVCs cannot be conducted concurrently with the task(s) of interest; motion capture and EMG are better suited for this purpose.

2.3.2 Motion capture

Motion capture systems measure kinematics by quantifying the coordinate location and orientation of bodies/segments and their time derivatives (Hamill et al., 2014). Due to the abundance of motion capture systems that can be used to collect these data, this subchapter focused on three widely-used types of systems: optoelectronic, markerless, and inertial-based. Optoelectronic motion capture systems use charge-coupled device (or infrared) cameras to record the position of markers affixed to a moving subject at high sampling frequencies of 100-500 fps (Robertson et al., 2014). At least three, non-collinear markers are needed on each segment to define its' local coordinate system (LCS) during calibration (Hamill et al., 2014). Optoelectronic systems are considered the gold-standard for motion capture because of their (constantly improving) measurement accuracies: 0.2mm and 2.0mm error for static postures and dynamic movements, respectively (Eichelberger et al., 2016; Merriaux et al., 2017). However, markerless systems are gaining attention because they are easier to implement, are less affected by skin motion artefacts, and can capture numerous individuals concurrently in a larger capture volume (Colyer et al., 2018). Markerless motion capture systems can estimate 3D kinematics by using cameras (with and without depth sensors) and an ML-based discriminative approach (Colyer et al., 2018). A discriminative approach uses image and depth data to directly infer the shape, pose, and location of the body using ML models learned using large databases of body parts, joints, and pose examples (Colyer et al., 2018). These systems are still in development and are limited because their validity has yet to be established for some tasks (Colyer et al., 2018) and are difficult to validate against marker-based motion capture systems due to their development using non-markered subjects. Inertial-based systems rely on numerous IMUs to track the segments they are attached to. An IMU is typically comprised of an accelerometer to measure linear acceleration and a gyroscope to measure angular velocity (Tickoo & Iyer, 2017); a magnetometer is often included to provide an external frame of reference for drift correction (Wittmann et al., 2019). IMUs can provide 6-degree-of-freedom segment tracking by defining segment coordinate axes, applying a sensor fusion algorithm (e.g., Madgwick and Kalman filters), integrating angular velocity, and double-integrating linear acceleration signals (Schepers et al., 2018; Tickoo & Iyer, 2017). Although processing is needed to minimize the drift amplified by the integrations (Schepers et al., 2018), inertial-based systems are preferred when occlusion may be an issue. However, they can be compromised by strong magnetic disturbances and may be inconvenient to wear depending on the task/activity of interest.

The principles to analyze the kinematic data from the different sensors are the same (Hamill et al., 2014). For studying fatigue, kinematic data can be used to calculate kinematic variables or be used with force data for kinetic analyses. Joint angles between adjacent segments are commonly calculated using the Cardan-Euler method, where the orientation of one segment's LCS relative to another is represented by three successive rotations about unique axes, often a sagittal-frontal-transverse rotation sequence (Hamill et al., 2014). These joint angles can be used to compute descriptive statistics, but also used for movement variability (e.g., goal equivalent manifold; Sedighi & Nussbaum, 2019), coordination analyses (e.g., continuous relative phase; Chan et al., 2020; Smallman et al., 2013; Zehr et al., 2018), or local dynamic stability (e.g., Lyapunov exponent; Granata & Gottipati, 2008). If internal loading variables are of interest, they can be estimated using ground reaction force data and inverse dynamics in a linked-segment model (Kazemi et al., 2021). Kinematics alone has been used to indirectly detect (Baghdadi et al., 2018) or model fatigue progression (Chapter 2.4; Maman et al., 2017). In these examples, ML models were trained to detect fatigue from the kinematics of an ankle-worn IMU during walking (Baghdadi et al., 2018), or estimate continuous fatigue status using four body-worn IMUs during MMH tasks (Maman et al., 2017). Although motion capture is a minimally invasive method to monitor fatigue, it requires validation.

2.3.3 Electromyography (EMG)

EMG measures motor unit APs as muscle electrical activity using electrodes (Kamen, 2014). The acquired signal is termed EMG activity and is commonly recorded with surface electrodes placed in different configurations: monopolar, bipolar, or double differential (Kamen, 2014). In all configurations, one electrode is typically placed on an electrically neutral site (e.g., a bony prominence). In the monopolar arrangement, another electrode is placed on the muscle of interest. Although appropriate for some isometric contractions, monopolar arrangements are less stable and have poor selectivity because of interference from surrounding activities; thus, they should not be used for dynamic contractions (Kamen, 2014). Bipolar arrangements are more stable with better selectivity because two electrodes are placed on a muscle and a single differential amplifier can attenuate signals common to the two electrodes using common-mode rejection (Kamen, 2014). Three inputs with a double differential amplifier may offer even better selectivity

(Guerrero et al., 2016). Surface EMG is relatively non-invasive but has downsides: the sensors can act as antennas for environmental radio frequencies; have trouble isolating smaller muscles from crosstalk; and cannot measure muscles beyond 20mm below the skin surface (Kamen, 2014). EMG electrodes can also be wireless and include on-board IMUs (e.g., Delsys Trigno Avanti).

EMG is frequently used to augment biomechanical models by accounting for muscle co-contraction (Cholewicki & McGill, 1994) and provide insight into contraction coordination (Gorelick et al., 2003). Many analysis methods driven solely using EMG have been employed to study the effects of fatigue and model the change in fatigue. An abundance of techniques exists and are continually being developed; commonly used methods are summarized in Appendix G. These analysis methods have been used in biomechanical and physiological research to study fatigue. As fatigue develops, EMG amplitude tends to increase if force production is relatively constant (Bigland-Ritchie et al., 1986; Kallenberg et al., 2007), and EMG frequency shifts to the left (i.e., decreases; Ament & Verkerke, 2009; Kallenberg et al., 2007). Entropy has been found to decline with the development of fatigue (Rampichini et al., 2020; Xie et al., 2010). Fractal analysis of fatigue has shown that the area under the multi-fractal spectrum of EMG can increase (G. Wang et al., 2007). Lastly, recurrence quantification analysis shows that percentage of determinism increases with the progression of fatigue (Ito & Hotta, 2012).

2.4 Fatigue modelling using kinematics

Fatigue modelling driven solely by kinematic data (i.e., from IMUs or cameras) has the benefit of being inobtrusive, requiring minimal instrumentation compared to other established methods. Several approaches have been employed to estimate fatigue status during manual, occupational tasks: statistical, state classification, and forecasting. Chan et al. (V C H Chan et al., 2020) used data from two IMUs (one each on the T8 vertebrae and pelvis), a composite index comprised of ten variables for each spine flexion-extension repetition was computed. This index was statistically expressed as a deviation (z-score) from each participant's typical, unfatigued movements at baseline. A strength of this method is that it is computationally inexpensive and has been found to be strongly correlated ($r \geq |0.91|$) to two established fatigue measures: fatigue visual analog scale (VAS) and MVCs (V C H Chan et al., 2020). However, to implement such a method, an unfatigued baseline needs to be established for each person. A state classification approach can sidestep this limitation by relying on pre-trained ML models that classify the state of fatigue, typically two states (e.g., unfatigued versus fatigued). Baghdadi et al. (Baghdadi et al., 2018) used one IMU on the ankle and trained a support vector machine (SVM) to classify participants as fatigued or unfatigued during walking, defined as being in a state pre- or post-performance of repetitive, MMH tasks. The authors achieved an accuracy of 90%, demonstrating that one IMU can be sufficient to detect fatigue during walking. A strength of this approach is the high classification accuracy, but no insight into continuous fatigue status is gained – an important consideration because fatigue is a continuous process (Enoka & Duchateau, 2016). Some work has quantified continuous fatigue, often referred to as fatigue estimation or fatigue forecasting. Maman et al. (2017) employed a multiple linear regression model to estimate fatigue level using five IMUs and a HR monitor. When the MMH task was known, the model was able to estimate the continuous, subjective fatigue level with a mean absolute error (MAE) of 2.16 on the Borg 6 – 20 scale (Maman et al., 2017). Using the same data, Hajifar et al. (2021) utilized autoregression models and further reduced the MAE to < 1.24. While the current state of fatigue modelling is rapidly improving, several limitations still exist.

One limitation with current fatigue modelling frameworks is that they focus on the classification of a discrete number of fatigue states (Victor C.H. Chan et al., 2022). Another limitation is that current frameworks are incapable of adapting to different tasks being performed; they are either designed for one movement type (Baghdadi et al., 2018), or tasks must be manually annotated prior to fatigue modelling (Maman et al., 2017). Third, many sensors (i.e., ≥ 5 sensors) are still being used, which may be a barrier for implementation and adherence in the real-world contexts. Future work should aim to employ fewer sensors, with the sensor selection guided by optimization (Clouthier et al., 2020). Lastly, more efforts should be focused on the improvement of continuous fatigue estimation, such as reducing the error in the estimates, and by including other fatigue measures to holistically represent the ground truth (e.g., force generation capacity, HR variables, EMG, etc.). Forecasting fatigue may be more efficacious for injury and accident mitigation compared to state classification because information about the progression of fatigue (e.g., the rate of change) can be used to provide warnings before excessive fatigue occurs (Bose et al., 2019).

2.5 Machine learning-based time series classification and regression

Machine learning (ML) refers to the use of algorithms to optimize a performance criterion using training data and/or past experience (Alpaydin, 2020). Using hyperparameters specified by the researchers, ML algorithms use computer programs to "learn" from existing data and develop models that reveal underlying structure (unsupervised learning) or make predictions on discrete or continuous output variable(s) in unseen data (supervised learning) (Victor C.H. Chan et al., 2022). Supervised learning is a category of ML algorithms which aim to learn a mapping from the input to an output whose correct values (i.e., the labels) are provided by a supervisor (e.g., the researcher) (Alpaydin, 2020). Supervised ML algorithms aim to reduce the sum of the costs, which is the error between the predicted and actual outcome (Duda et al., 2001). Supervised ML algorithms are often used to classify samples into two (i.e., binary classification) or more (multiclass classification) categories, or to perform regression, where the output is a continuous variable. The versatility of ML algorithms allows them to be useful for kinematic-based fatigue modelling. Specifically, classification can be used to classify the state of fatigue for the individual (e.g., fatigued vs. non-fatigued; (Baghdadi et al., 2018)), while regression can be used to estimate the fatigue level of the individual (i.e., fatigue forecasting) as a continuous variable (e.g., subjective fatigue appraisal; Maman et al., 2017). Additionally, ML is effective for HAR (Ordóñez & Roggen, 2016), which can be used as an intermediary step in a fatigue modelling framework that automatically accommodates different tasks. In all cases, the ML architectures employed need to be suitable for time series data.

Human movement data is typically comprised of high-dimensional, continuous, time series signals which are difficult to model due to unique properties, including an explicit time dependency and being non-stationary (Längkvist et al., 2014). To ensure time dependencies are accounted for, models should either include past input data or have memory of past inputs (Längkvist et al., 2014). Additionally, the characteristics (i.e., mean, variance, and frequency) of non-stationary signals change over time; thus, models should either be robust to non-stationary data or pre-processing steps (e.g., detrending) should be conducted (Bagnall et al., 2017). Various ML algorithms are effective for modelling time series data, such as *k*-nearest neighbours (*k*-NN) with dynamic time warping (DTW), SVM with quadratic kernel, and rotation forest (Bagnall et al., 2017). In recent years, artificial neural networks (ANNs) have gained more attention for modelling time series data because they can automatically extract features (Victor C.H. Chan et al., 2022). Two forms of ANNs are especially effective for modelling time series data: recurrent neural networks (RNNs) and convolutional neural networks (CNNs). RNNs account for time dependencies because the hidden units have connections from both current and previous time frames (Längkvist et al., 2014), where long short-term memory (LSTM) cells are often used to bridge larger time intervals without the risk of exploding or vanishing gradients during training (Hochreiter & Schmidhuber, 1997). CNNs are suitable for high dimensional time series data because the filters (i.e., convolution kernels) can extract discriminative features, and the learned filters remain invariant across time (Fawaz et al., 2019). As such, human movement scientists have successfully performed fatigue state classification using an LSTM ANN (J. Wang et al., 2021) and an ANN with LSTM and convolutional layers to perform HAR (Clouthier et al., 2020; Ordóñez & Roggen, 2016). Little work has been done on forecasting continuous fatigue, and state-of-the-art ANN architectures (e.g., with LSTM and convolutional layers) have not been employed for this purpose.

Chapter 3. Study 1: The role of machine learning in the primary prevention of work-related musculoskeletal disorders: A scoping review

Published as:

> Chan, V. C., Ross, G. B., Clouthier, A. L., Fischer, S. L., & Graham, R. B. (2022). The role of machine learning in the primary prevention of work-related musculoskeletal disorders: A scoping review. *Applied Ergonomics, 98*, 103574.

Abstract

To determine the applications of machine learning (ML) techniques used for the primary prevention of work-related musculoskeletal disorders (WMSDs), a scoping review was conducted using seven literature databases. Of the 4,639 initial results, 130 primary research studies were deemed relevant for inclusion. Studies were reviewed and classified as a contribution to one of six steps within the primary WMSD prevention research framework by van der Beek et al., (2017). ML techniques provided the greatest contributions to the development of interventions (48 studies), followed by risk factor identification (33 studies), underlying mechanisms (29 studies), incidence of WMSDs (14 studies), evaluation of interventions (6 studies), and implementation of effective interventions (0 studies). Nearly a quarter (23.8%) of all included studies were published in 2020. These findings provide insight into the breadth of ML techniques used for primary WMSD prevention and can help identify areas for future research and development.

3.1 Introduction

In the United States of America (USA), over 600,000 work-related musculoskeletal disorders (WMSDs) accounted for a third of all lost work days and led to $54 billion in costs in 2014 (Kang et al., 2014). Affected workers are subjected to quality of life detriments and performance setbacks, as WMSDs impair bodily structures such as muscles, tendons, ligaments, cartilage, bones, joints, and/or nerves (Goes et al., 2020; Smith et al., 2020). WMSDs develop gradually due to a progression of cumulative damage (Gallagher & Schall, 2017; McGill, 1997), but do not typically include injuries resulting from an acute or instantaneous event such as slips, trips, or falls (Hales & Bernard, 1996). WMSDs can be caused and/or aggravated by various work activities and constitute a major problem globally (de Kok et al., 2019; Patel et al., 2017). Thus, primary WMSD prevention is a leading focus for researchers, practitioners, and organizations alike.

In order to strengthen preventative efforts, van der Beek et al. (2017) presented a framework for WMSD prevention research aimed are integrating relevant research disciplines. They identified six steps for WMSD prevention research: 1) determine incidence and severity; 2) identify risk factors; 3) understand the underlying mechanisms; 4) develop intervention(s); 5) evaluate intervention(s); and 6) implement effective intervention(s). The development of this framework was motivated by the limited success of preventative efforts to lessen the WMSD burden (van der Beek et al., 2017). Concurrently, advances in relevant research disciplines related to wearable sensing technologies (see recent reviews by Lim and D'Souza, 2020; Stefana et al., 2021) and computer science rooted in machine learning (ML) techniques may be particularly useful to address emerging research questions within the six WMSD prevention themes above. While recent reviews (Lim & D'Souza, 2020; Stefana et al., 2021) summarize the use of wearable sensing technologies for WMSD prevention, this scoping review summarizes how ML techniques are being applied to address WMSD prevention efforts to identify opportunities for continued advancement.

ML is an essential branch of artificial intelligence with increasing use in many industries due to technological advancements that have increased data collection volume and improved processing capacity. ML refers to the use of algorithms to optimize a performance criterion using training data and/or past experience (Alpaydin, 2020). Using hyperparameters specified by the researchers, ML algorithms use computer programs to "learn" from existing data and develop models that reveal underlying structure (unsupervised learning) or make predictions on discrete or continuous output variable(s) in unseen data (supervised learning). Machine learning algorithms are widely applicable for regression, classification, clustering, and reinforcement learning purposes. Unsupervised learning can discover underlying patterns in datasets, such as *k*-means clustering to identify subgroups (Reme et al., 2012). Reinforcement learning is used to determine an optimal action sequence (i.e., policy) to maximize return, and has been implemented to provide catered knowledge and feedback in a virtual posture training environment for workers (Akanmu et al., 2020). Most commonly, supervised learning can classify/make estimations on new cases, such as *k*-

nearest neighbours (*k*-NN) classifier to automatically recognize movement (Adaskevicius, 2014). ML can model non-linear interactions between numerous variables and is argued to be better suited for understanding the complex etiology of WMSDs and subsequently, preventing their occurrence (Ruddy et al., 2019). Although ML techniques often trade-off interpretability for predictive performance (Molnar, 2020), these techniques have begun to advance primary WMSD prevention efforts (Gross et al., 2020).

ML techniques have potential to help prevent WMSDs, as recent reviews have described the utility of ML techniques to address research questions tangential to WMSD prevention. For example, Halilaj et al. (2018) summarized how ML can be used to inform human movement biomechanics; where other groups have reviewed ML applications closely related to WMSD prevention, such as: artificial intelligence for injury risk assessment and performance prediction in team sports (Claudino et al., 2019); fuzzy decision support systems for musculoskeletal disorder diagnoses (Farzandipour et al., 2018); analysis of occupational accidents (Sarkar and Maiti, 2020); and textual injury surveillance analyses (Vallmuur, 2015). Select reviews have also focused on wearable technologies (Lim & D'Souza, 2020; Stefana et al., 2021) and electromyography (EMG) based muscle fatigue detection (Karthick et al., 2018) which overlap with the use of ML techniques in WMSD prevention. However, no review has robustly scoped the WMSD prevention literature specifically to summarize the state-of-the-art regarding the use of ML techniques to inform WMSD prevention with reference to the six WMSD prevention research themes (van der Beek et al., 2017). Summarizing the state-of-the-art will help practitioners understand what ML techniques are currently being employed and for what purpose. Such knowledge is also essential to help researchers and practitioners make decisions about where and how to use ML techniques to inform WMSD prevention efforts (research or practice). To this end, the objective of this work was to survey how ML techniques have been used in primary WMSD prevention using a scoping review methodology (Arksey & O'Malley, 2005; Levac et al., 2010). The literature search strategy, results, and discussion were guided, synthesized, and organized respectively using the six-step theoretical WMSD prevention framework by van der Beek et al. (2017) shown in Fig. 3.1.

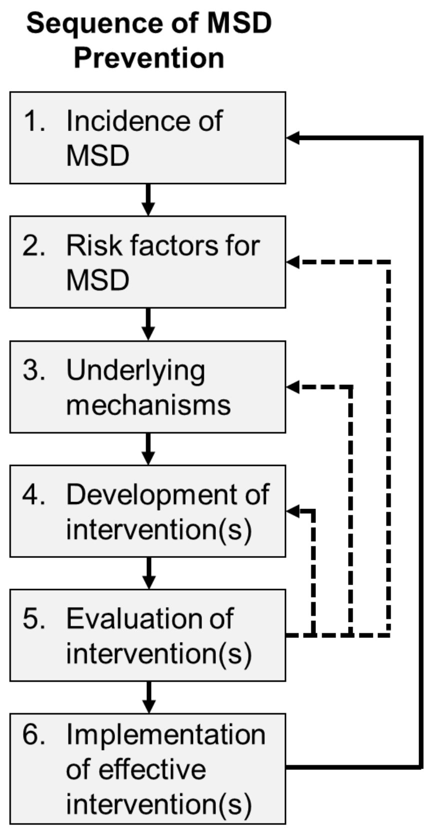


Fig. 3.1. The repeated sequence of work-related musculoskeletal disorder (MSD) prevention framework (van der Beek et al., 2017). The steps in this framework are: 1) determining the extent and severity of MSD using epidemiological data; 2) identifying risk factors using epidemiological data and cross-sectional studies; 3) determining the mechanisms and pathways for MSD development; 4) developing an intervention that targets known risk factors and mechanisms; 5) evaluating the efficacy and effectiveness of the intervention; and 6) implementing the intervention(s) while considering the process, cost-effectiveness, and fidelity.

3.2 Methods

This research employed the scoping review methodological framework presented by Arksey and O'Malley (2005) with recommendations by Levac et al. (2010), and the results were presented according to the PRISMA reporting guidelines. The six-step theoretical WMSD prevention framework by van der Beek et al. (2017) was selected to guide the formulation of the literature search methodology and organize the results and discussion because the multi-disciplinary considerations used to develop this framework accurately represent the nature of primary WMSD prevention. The ML terminology used throughout this review is presented in Appendix A.

3.2.1 Research question

How have ML techniques been used for primary WMSD prevention in the steps outlined by van der Beek et al., (2017)?

3.2.2 Literature search methodology

The literature search strategy was constructed around three concepts: machine learning, musculoskeletal disorders, and work. The search strategy was initially developed using the Medline (Ovid) database, guided using the WMSD prevention research framework (van der Beek et al., 2017) in conjunction with the health sciences and engineering librarians at the University of Ottawa. The search strategy was translated to six additional databases: Embase (Ovid), The Cochrane Central Register of Controlled Trials (CENTRAL; Ovid), PsycInfo (Ovid), IEEE Xplore, SPORTDiscus, and Scopus. These databases were selected to ensure a comprehensive search of health sciences, human kinetics, ergonomics, engineering, and computer science domains. Subject headings were used with the “explode” command wherever possible (which also search the associated sub-headings and keywords). The search strategies are presented in Appendix B. The results were screened and reviewed for inclusion using the following inclusion criteria:

- Published in English on or after January 1st, 1990
- Full-length, peer-reviewed, published journal article
    - Dissertations, conference abstracts, conference papers/proceedings, technical reports, patents, short communications, and supplements were excluded
- Original research
    - Book chapters/sections and reviews (of any type) were excluded
- Machine learning technique(s)[1] were used to develop their model and/or system, and/or to produce the final results

[1] Techniques were considered to be machine learning if they were identified in the scikit-learn documentation (Pedregosa et al., 2011).

    - Studies which applied dimensionality reduction techniques only were excluded
- The research represents a direct effort towards the prevention of WMSDs, as defined using the WMSD prevention framework (van der Beek et al., 2017)
    - The development/augmentation of biomechanical methods used to study WMSD etiology were included
    - Studies investigating traumatic and acute injuries due to accidents only (e.g., falls, struck-by, vehicle collision) were excluded; studies were included if WMSDs were investigated alongside traumatic and acute injuries

Failure to meet all the aforementioned criteria led to exclusion. The results were imported to Covidence research review management software (Covidence, Victoria, Australia) to identify duplicates and to conduct the title & abstract screening and full-text review. Title & abstract screening was conducted by two researchers (VCHC & GBR) where they had a screening agreement >90% on 50 randomly selected studies from the search results. Full-text review was conducted by the same researchers where agreement for inclusion was required for a study to progress to extraction. In the case of disagreements, the remaining authors were consulted.

Once the studies for inclusion were solidified, their reference lists and citing studies were searched using Google Scholar to manually identify additional relevant studies that may have been missed. Specifically, duplicates were removed amongst the referenced and citing studies, then subjected to title and abstract screening using the same inclusion criteria. If a study proceeded beyond the initial screening, these manually-identified studies were imported into Covidence and subjected to the previously-described full-text review by the same researchers.

3.2.3 Data extraction

Data extraction was conducted using Microsoft Excel (Microsoft, Washington, USA), where study author(s), year, title, journal, country, purpose(s), sample, data, ML technique(s) used, best technique(s), application, and outcomes were recorded. Furthermore, the availability of code and data in each study were indicated to determine if studies were reproducible. The country of each publication was determined by the majority of (co-)authors' affiliated organization country. If co-authors were equally distributed between multiple countries, the country of the corresponding author was used (this author never had multiple affiliated countries). To ensure conciseness, the study purposes were extracted in a manner to focus on the use of ML. For instance, if ML was only used to achieve one of four study objectives, only this objective would be extracted.

The ML techniques/algorithm(s) used in each study were extracted in the manner reported by the respective authors. However, specific information regarding their development (e.g., the selection of the hyperparameters) have been omitted. This was done to ensure conciseness while still providing a summary relevant to the study objective. Additionally, for counting purposes, similar ML algorithms were considered together. This decision primarily affected decision tree (DT) algorithms and artificial neural networks (ANNs): a variety of DT algorithms (classification and regression tree, multivariate adaptive regression splines, etc.) and ensemble learning methods (random forest, gradient boosting, etc.) exist; and ANNs vary in complexity, architecture, and purposes (recurrent neural network, convolutional neural network, self-organizing map, etc.). Although grouping related algorithms was necessary for counting, there are limitations to the conclusions that can be inferred from these data (discussed in Chapter 3.4.7). For full details on the models reported, readers are referred directly to each study.

The best ML technique(s) in each study was identified by the authors of each respective study. The performance metric(s) used to define the best techniques in each study (e.g., area under the receiver-operator curve, accuracy, F1-score, sensitivity, specificity, negative predictive value, precision, or computational speed) varied due to differing study purposes, degree of class imbalance, application of ML, validation/testing method (i.e., hold-out tests or cross validation), or number/type of outputs. Some studies identified more than one best ML technique, in which case multiple best techniques were recorded. This occurred because different models were best suited to different stages in a multi-stage framework, or ≥ 2 models showed comparable performance for the same purpose. Some studies employed only one ML technique; this technique was reported as best.

The application(s) of ML techniques in each study were classified into categories: binary classification, if they predicted a dichotomous response variable (e.g., fatigued or non-fatigued); multiclass classification, if they predicted one of three or more finite, discrete classes (e.g., activity classification); regression, if they estimated/modelled continuous response variable(s) (e.g., estimating grip force in newtons); clustering, if they computed similarity between samples to reveal underlying structure or patterns in a dataset (e.g., investigating subgroups in a sample); and reinforcement learning, if they learned optimal policies to maximize return (e.g., determining the optimal feedback content and delivery for virtual worker training). A study received multiple categorizations if ML techniques were used for multiple, different applications.

3.2.4 Data synthesis

The WMSD prevention research framework (van der Beek et al., 2017) was used to synthesize the data from the included studies. During data extraction, the primary contribution of each publication was assessed and classified into one of the six steps of the WMSD prevention research framework using the following criteria:

Step 1. Incidence and severity of WMSD: Determined the extent, severity, or impact of WMSDs in the population of interest using descriptive epidemiological data (e.g., from worker compensation companies).

Step 2. Risk factors for WMSD: Used epidemiological observations to gain insight into and/or identify potential work-related risk factors.

Step 3. Underlying mechanisms: Unravelled the underlying mechanisms and pathways for the association of work-related physical risk factors and WMSD in the workplace and/or laboratory. Studies that developed/augmented biomechanical methods for investigating WMSD etiology were also classified as this step.

Step 4. Development of intervention(s): Developed and introduced an intervention to reduce WMSD incidence based on previously-identified underlying mechanisms (i.e., reducing a possible risk factor).

Studies which introduced unobtrusive and/or automated ergonomic assessments were also classified as this step.

Step 5. Evaluation of intervention(s): Evaluated the effectiveness and/or practical feasibility of preventative interventions, with either efficacy studies under well-controlled circumstances or effectiveness studies in real working-life situations.

Step 6. Implementation of effective intervention(s): Studied the large-scale intervention implementation in the working society, with focus on the implementation process, cost-effectiveness, and fidelity.

Each study was assigned to one step only; if contributions from a study extended beyond the scope of one step, the study was classified based on the primary contribution presented (e.g., the primary purpose or research question).

3.3 Results

3.3.1 Overall results

The searches were conducted on December 3rd, 2020 and returned a total of 4,627 results from Medline (580 results), Embase (1,095 results), CENTRAL (64 results), PsycInfo (164 results), IEEE Xplore (13 results), SPORTDiscus (45 results), and Scopus (2,666 results). A total of 1,426 studies were identified as duplicates and 3,079 studies were excluded during title & abstract screening or full-text review, resulting in the inclusion of 122 studies. Manual searching of the reference lists and citing studies of these 122 initial studies occurred on June 16th, 2021. This manual searching identified an additional 12 studies, 8 of which were included, resulting in a total of 130 studies. The results of the literature search, screening process, and study classification are reported using the PRISMA guidelines and presented in Fig. 3.2.

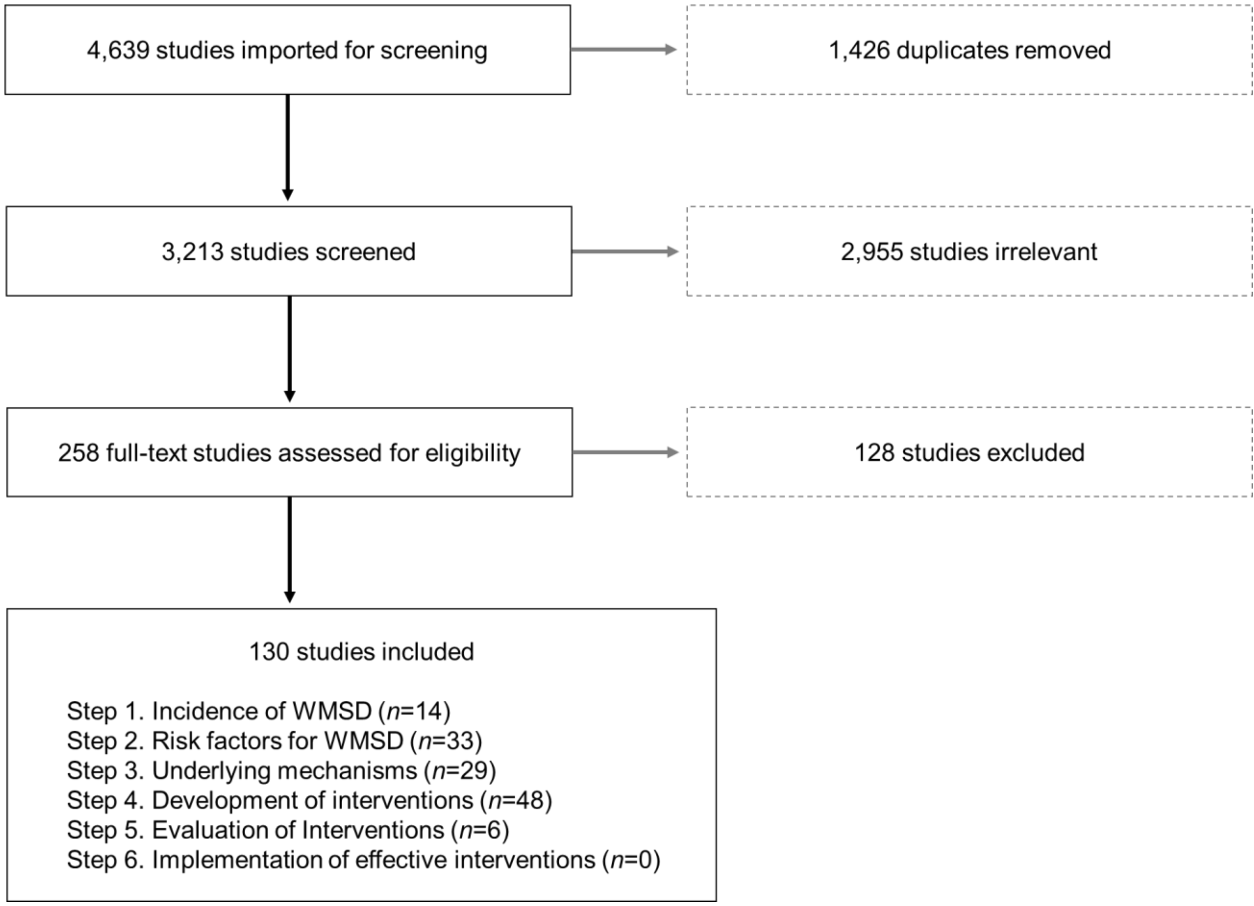


Fig. 3.2. Results of the literature search, screening process, and study classification presented according to the PRISMA reporting guidelines.

3.3.1.1 Publications by year

The number of included studies published each year from January 1st, 1990 – June 16th, 2021 is presented in Fig. 3.3. The earliest included study was published in 1996 and a total of 7 included studies were published before or in 2000. In the following decade (2001-2010), 19 included studies were published. Of the remaining 104 included studies published in 2011-2021, 31 were published in 2020, accounting for 23.8% of all included studies in this review.

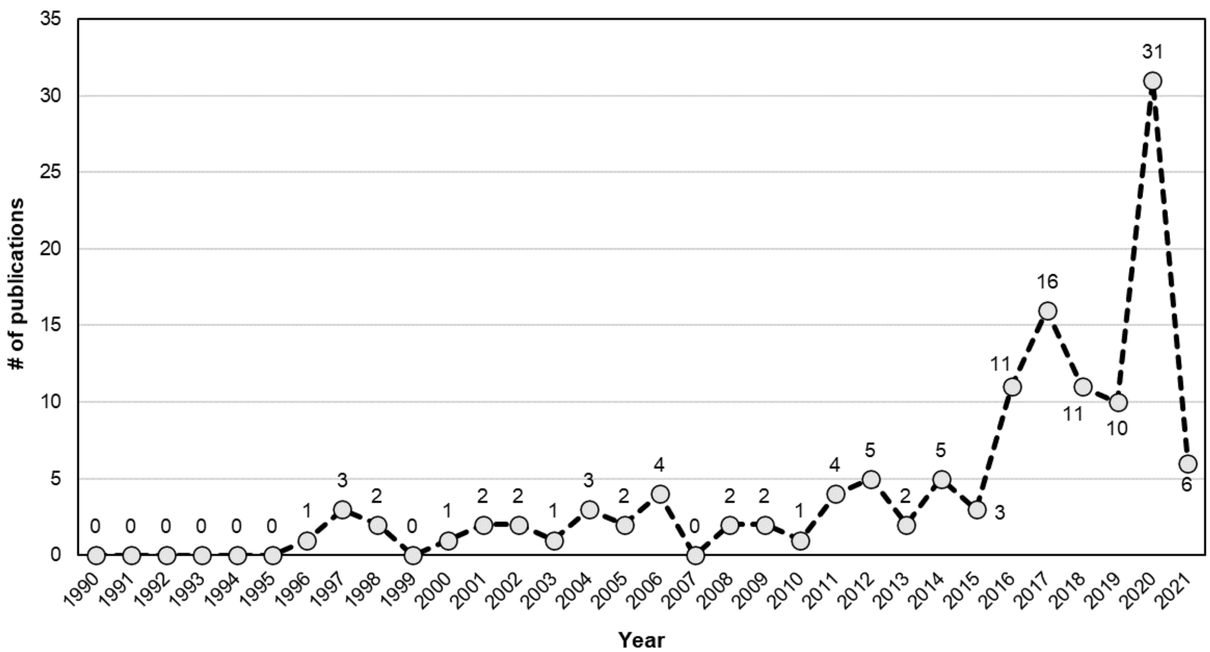


Fig. 3.3. The number of included studies published each year from January 1st, 1990 – June 16th, 2021.

3.3.1.2 Publications by country

The number of publications from each country are presented in Fig. 3.4. The USA produced the largest number of relevant publications (59), accounting for 45.4% of all included studies. In total, publications from 27 countries were included.

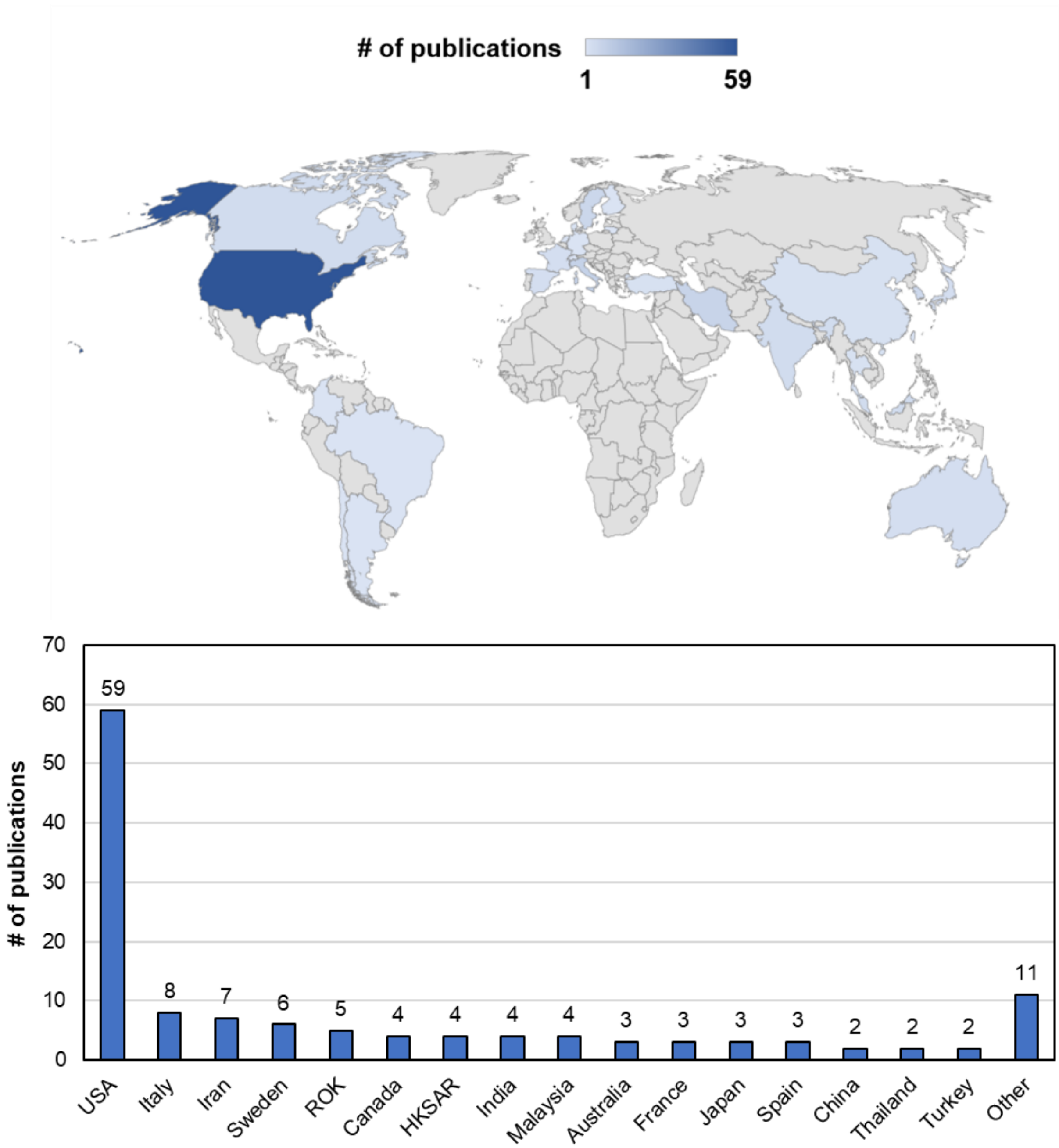


Fig. 3.4. The number of publications from each country, presented on a world map (top) and bar graph (bottom). USA = United States of America; ROK = Republic of Korea; HKSAR = Hong Kong Special Administrative Region; Other = countries with 1 publication: Argentina, Brazil, Chile, Colombia, Denmark, Finland, Germany, Lithuania, Netherlands, Switzerland, and Taiwan.

3.3.1.3 Machine learning techniques employed

The most commonly employed ML techniques and the number of times they were identified as being the "best" in a study (as identified by the authors of the respective studies) are presented in Fig. 3.5. ANN was the most commonly employed ML technique (used in 43.1% of studies) and selected as the best technique for 34.6% of studies. Types of ANNs employed include feedforward (multilayer perceptron), recurrent (long short-term memory; LSTM), convolutional (hourglass), radial basis function, self-organizing map, and fuzzy neural networks.

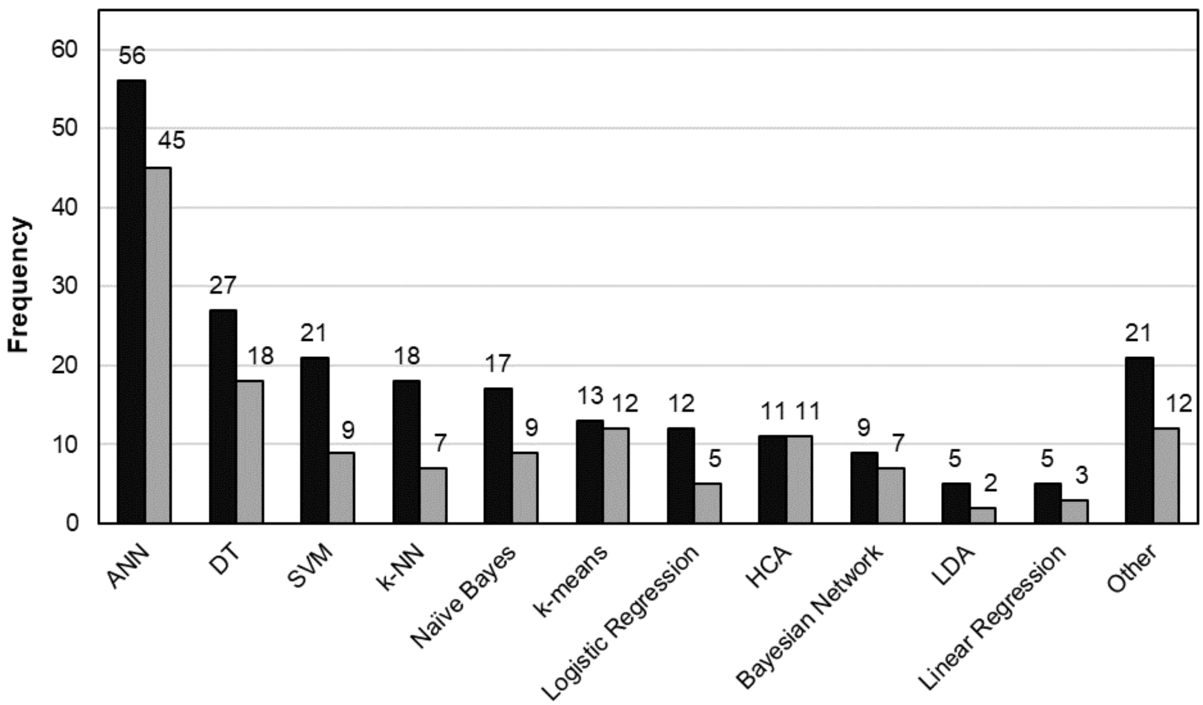


Fig. 3.5. Across the 130 included studies, the frequency of machine learning (ML) techniques used (black bars) and how often they were the best technique in a study (grey bars; evaluated by respective authors). ANN = artificial neural network; DT = decision tree and ensembles; SVM = support vector machine; *k*-NN = *k*-nearest neighbors; naïve Bayes = naïve Bayes classifier; *k*-means = *k*-means clustering; HCA = hierarchical cluster analysis; LDA = linear discriminant analysis, Other = ML techniques used in less than 4 studies.

3.3.1.4 Applications of machine learning

The included studies applied ML techniques most commonly for multiclass classification (36.2%), followed by regression (26.2%), binary classification (23.1%), clustering (18.5%), and reinforcement learning (0.8%). In total, more than half (59.2%) the studies employed ML for classification. Six studies used ML techniques for two different applications (none greater than two). The types of ML applications used in this review are presented in Fig. 3.6.

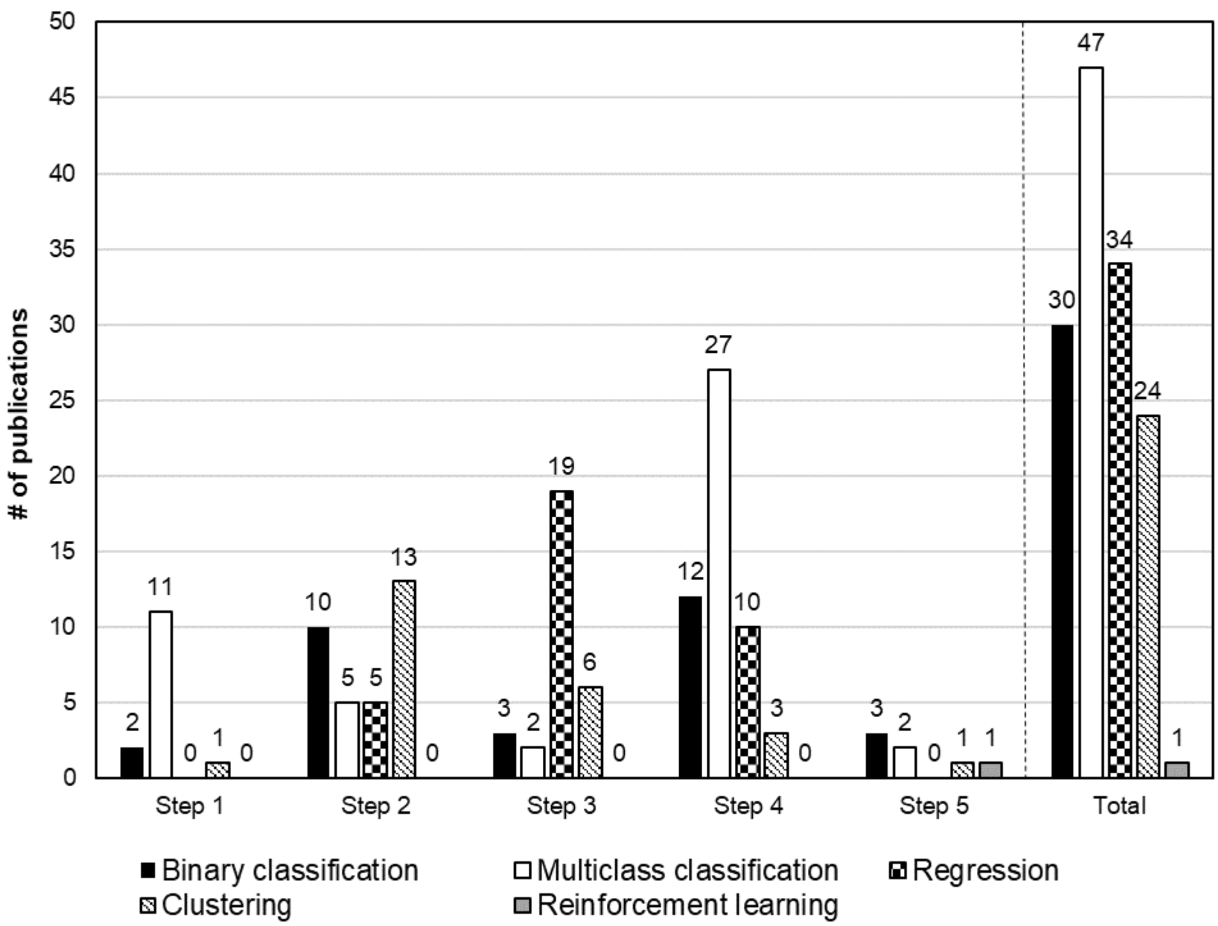


Fig. 3.6. The types of machine learning (ML) applications across the included studies presented by the step of the work-related musculoskeletal disorder (WMSD) framework (van der Beek et al., 2017) alongside the totals across all steps. Step 6 (implementation of effective interventions) was not depicted as no studies were classified to this step. Note that total number of applications were greater than the number of included studies (130) because six studies utilized ML techniques for two applications. Step 1 = incidence and severity of WMSD; step 2 = risk factors for WMSD; step 3 = underlying mechanisms; step 4 = development of interventions; step 5 = evaluation of interventions.

3.3.2 Synthesis of Results

The 130 studies identified were classified as one of the six steps of the WMSD prevention research framework (van der Beek et al., 2017; Fig. 3.1), summarized in this subchapter.

3.3.2.1 Incidence and severity of WMSD (Step 1)

Fourteen studies (10.8%) primarily contributed to determining the incidence and severity of WMSDs, presented in Appendix C1. The majority of these studies (9 of 14) were published in 2015 or later, and 12 of 14 studies were from the USA. Naïve Bayes classifier (NB) was the most commonly employed ML technique in this step, evaluated as the best technique in 7 of 9 studies in which it was used. In 11 of 14 studies in this step, ML techniques were used for multiclass classification. The sample sizes ranged from 2210 – 1,200,000 (mean = 113,534.4, SD = 314,213.3).

3.3.2.2 Risk factors for WMSD (Step 2)

Thirty-three studies (25.4%) primarily contributed to determining risk factors for WMSDs, presented in Appendix C2. The majority of these studies (17 of 33) were published in 2015 or later, and the largest proportion (10 of 33) were from the USA. ANN was the most commonly employed ML technique in this step (11 studies). Hierarchical cluster analysis (HCA) was most frequently evaluated as the best ML technique (9 studies). No studies in this step employed both ANN and HCA. The most common type of application for ML in this step was clustering (13 of 33), followed by binary classification (10 of 33). The sample sizes ranged from 40 - 403,560 (mean = 20,333.6, SD = 72,185.3).

3.3.2.3 Underlying mechanisms (Step 3)

Twenty-nine studies (22.3%) primarily contributed to determining the underlying mechanisms of WMSDs, presented in Appendix C3. The majority of these studies (16 of 29) were published in 2016 or later, and the largest proportion (13 of 29) were from the USA. ANN was the most commonly employed ML technique in this step, evaluated as the best technique in 18 of 20 studies in which it was used. In 19 of 29 studies in this step, ML techniques were used for regression. The sample sizes ranged from 1 – 204 (mean = 24.7, SD = 38.1).

3.3.2.4 Development of interventions (Step 4)

Forty-eight studies (36.9%) primarily contributed to the development of WMSD prevention intervention(s), presented in Appendix C4. The majority of these studies (42 of 48) were published in 2016 or later, and the largest proportion (23 of 48) were from the USA. ANN was the most commonly employed ML technique in this step, evaluated as the best technique in 19 of 24 studies in which it was used. In 27 studies in this step, ML techniques were used for multiclass classification. The sample sizes ranged from 1 – 12,985 (mean = 307.0, SD = 1,870.6).

3.3.2.5 Evaluation of interventions (Step 5)

Six studies (4.6%) primarily contributed to the evaluation of WMSD prevention intervention(s), presented in Appendix C5. A third of these studies (2 of 6) were published in 2020, and the largest proportion (2 of 6) were from Thailand. Bayesian network and hidden Markov model (HMM) were the most commonly employed ML techniques in this step, both evaluated as the best technique in the 2 studies in which they were used. The most common type of application for ML in this step was binary classification (3 of 6), followed by multiclass classification (2 of 6). The sample sizes ranged from 1 – 28 (mean = 13.2, SD = 9.2).

3.3.2.6 Implementation of effective interventions (Step 6)

No studies employing ML techniques primarily contributed to the study of effective implementation of WMSD prevention intervention(s).

3.4 Discussion

This scoping review was the first to investigate the role of ML techniques across all steps of primary WMSD prevention. Since 1990, 130 English, peer-reviewed, original journal articles have employed ML techniques to contribute to a component within WMSDs prevention, where 80.0% of these studies were published within the past decade. This trend in ML use was not surprising considering the findings of related reviews. Halilaj et al. (2018) reviewed ML for human movement biomechanics and observed an exponential increase in publications per year from 1995-2017; similarly, Sarkar and Maiti (2020) observed a dramatic increase in publications from 1995-2019 when studying ML for occupational accident analysis. Thus, ML techniques are being used with increasing frequency for occupational health and safety purposes in an attempt to extend analyses beyond traditional statistical approaches, and primary WMSD prevention is not an exception. The studies in this review were most frequently conducted by researchers affiliated with American institutions, aligned with observations from Sarkar and Maiti (2020). Two explanations for this finding are speculated: the exclusion of non-English studies heavily biases countries with English as a primary language; and, the USA was recently evaluated as the most productive scientific research country (Crew, 2019), where second- and third-ranked countries do not consider English a primary language.

The most frequently employed ML techniques in the current review were ANN and DT-based models. Similar findings were observed in other related reviews (Claudino et al., 2019; Halilaj et al., 2018; Sarkar & Maiti, 2020). The common use of ANNs may be explained by the suitability of certain architectures for time series data, common in biomechanical research for WMSD prevention. Time series data can be challenging to model because of their explicit time dependency and non-stationarity (Längkvist et al., 2014; Leise, 2017). Recurrent neural networks, for example, can leverage short-term time dependencies due to connections between hidden units (Längkvist et al., 2014), where LSTM cells further extend their ability by capturing long-term time dependencies (Hochreiter & Schmidhuber, 1997). Another explanation for the frequent use of ANNs and DT-based models is the variety of architectures and models that afford flexibility so that they can be used for various classification and regression problems. It follows that classification and regression applications were the most common throughout this review. This was primarily driven by the applicability of classification models to autocode claims and injury reports (step 1), classify the risk level of jobs (step 2), and classify fatigue, risk, or strenuous tasks being performed (step 4). Regression was mostly used to augment existing biomechanical methods to study WMSD etiology (step 3), but also to perform ergonomic assessments and model fatigue (step 4). For these purposes, the continuous outputs of kinematics, electromyography (EMG), ground reaction forces, internal loading, inputs for ergonomics assessments, and fatigue were estimated. Clustering was the third most common application of ML techniques, where it was used as a method to explore potential risk factors (step 2) and identify biomechanical patterns (step 3). Overall, problems in WMSD prevention research lends themselves to the application of ML techniques, specifically ANN and DT-based methods; however, the importance of model interpretability should be considered when employing these models in practice.

3.4.1 Incidence and severity of WMSD (Step 1)

Understanding the incidence and severity of WMSD is a key initial step to launching prevention efforts. However, navigating through and coding compensation claims data can be challenging and time consuming. This review found that the most common application of ML within the incidence and severity theme was to autocode worker compensation claims, representing a classification problem. Studies aimed to reduce the manual labour in quantifying injury incidences and in determining injury types, causes, and severity from large databases. The input data were primarily free-text injury narratives from worker compensation insurance claim reports, sampled from general industry workers. Naïve Bayes classification was the most commonly used ML technique to autocode injury narratives into causation groups because it can do so efficiently and accurately without manual intervention, while being highly interpretable and providing scores on prediction confidence to flag records for manual review (Lehto et al., 2009; H. R. Marucci-Wellman et al., 2017). A notable exception was the use of latent Dirichlet allocation on documents from social media websites to examine occupational injury under-reporting (Min et al., 2019). WMSDs were studied alongside many injury types, where the predominant injuries examined were acute/traumatic in nature (i.e., accidents). This speaks to the versatility of the methods used to investigate WMSD incidence and severity; the ML techniques have been applied to code injury types using various coding systems, such as the Bureau of Labor Statistics Occupational Injury and Illness Classification system in the USA.

3.4.2 Risk factors for WMSD (Step 2)

To effectively mitigate WMSD risk, it is important to understand which factors are most closely linked to risk. This current review identified three distinct ML-based approaches used to study WMSD risk factors from the primary literature. First, unsupervised clustering algorithms were used to retrospectively identify subgroups and potential WMSD risk factors. Another approach used ML techniques to identify predictors of physical WMSD risk factor magnitude (i.e., whole body vibration; Atal et al., 2020), approached with either clustering algorithms or regression models. Third, ML-based models were developed to classify industrial jobs as low- or high-risk for WMSDs, representing a classification problem. Studies targeting WMSD risk factor identification used diverse types of input data: questionnaires and interviews, injury records, radiographs, physiological measurements, biomechanical measurements (kinematic, kinetic, and EMG data), and ergonomic assessments of physical exposure (e.g., rapid upper limb assessment; RULA). Many working populations were also studied, often in isolation: manual material handlers; nursing, dental, and military personnel; ophthalmic and plastic surgeons; crane and dumpster operators; and general, meat processing, dairy parlour, semiconductor, computer, metal casting, agro-manufacturing, construction, energy, infrastructure, and mining workers. Despite the wide range of approaches used in this step, interpretable ML models are still valued to provide insight on the contribution of each predictor to WMSD risk. Although ANNs were commonly used to predict the risk-level for industrial jobs, cluster analyses using HCA are easily interpretable.

HCA was most frequently evaluated as the best algorithm to identify subgroups and explore potential WMSD risk factors. To identify potential risk factors, some studies followed the use of HCA or *k*-means clustering with traditional statistical approaches (e.g., analysis of variance; Leijon et al., 2006) to determine significant differences between clusters for WMSD incidences and input variables. DT-based models, such as chi-square automatic interaction detector (CHAID; Connaboy et al., 2018; Sivak-Callcott et al., 2011), were used in a similar manner without the need for additional statistical analyses. In general, this type of framework (using unsupervised learning) is useful for identifying subgroups and assessing their relationship with WMSD risk. Presented in Fig. 3.7, the prevalence of WMSDs can be compared between clusters identified using unsupervised learning. If one or more clusters significantly differ in WMSD prevalence, then other features can be compared between these clusters to identify potential risk factors (where further research is required to establish causality). The benefit of employing unsupervised learning algorithms for this purpose is the exploratory capability, where domain knowledge of potential risk factors specific to a population is not imperative. As such, some studies input many variables (e.g., 4 physical exposure assessments, 6 psychosocial variables, 10 individual variables, and a questionnaire; Márquez Gómez, 2020). It is important to note that such approaches can identify factors merely associated with WMSD occurrence; however, epidemiological studies (e.g., cohort, observational) remain necessary to make causal connections. Still, the capacity to explore associations between variables and subgroups makes unsupervised ML a useful tool to study WMSD risk factors in novel populations.

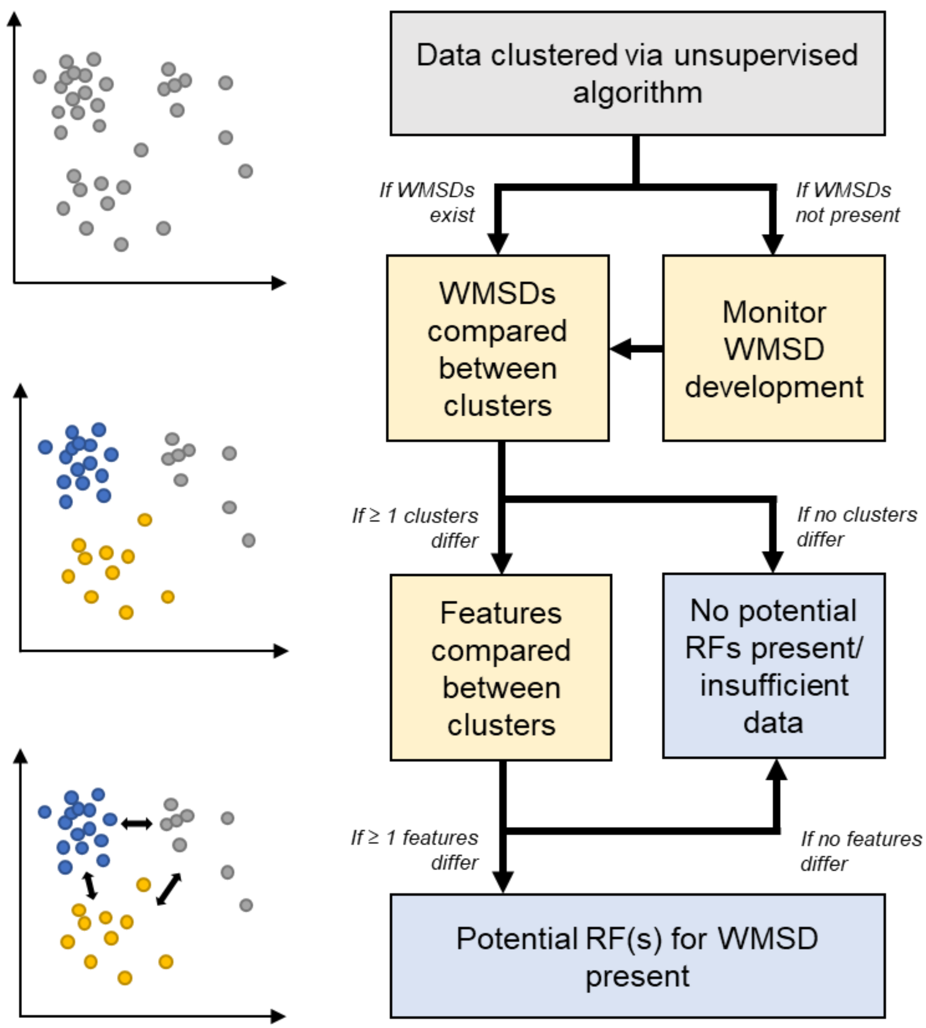


Fig. 3.7. A simplified, general framework to identify potential work-related musculoskeletal disorder (WMSD) risk factors (RFs) using unsupervised clustering algorithms. The label (e.g., WMSD status) of each data point (i.e., individual) is not provided to the clustering algorithms; instead, they identify natural or intrinsic clusters within the data guided by criterion specified by researchers. If the number of individuals with a WMSD differ significantly between clusters, then features can be compared between clusters to determine if they are potential RFs. Further research is required to establish casual relationships between these potential risk factors and WMSDs.

3.4.3 Underlying mechanisms (Step 3)

Understanding the mechanistic pathway between exposure and tissue-level injury is an important pursuit. Although some studies used ML techniques to identify patterns associated with WMSD risk factors (Baghdadi et al., 2021; Fujimaki & Mitsuya, 2002), the majority of these studies employed ML techniques to augment existing biomechanical methods used for studying WMSD etiology. One approach was to use easily-attainable measures and personal variables to estimate the output of traditionally-used biomechanical measurement tools, aiming to reduce the reliance on expensive, specialized equipment. The estimated outputs included EMG responses (Karwowski et al., 2006), grip and pinch strength (Sung et al., 2014; M. Wang et al., 2021), and ground reaction force (Muller et al., 2020). An additional noteworthy method is markerless motion capture, where RGB or RGB-D (depth) images from camera(s) are used to estimate 3D kinematics (Mehrizi et al., 2017, 2018, 2019), grip exertions (Asadi et al., 2020), and mechanical energy expenditure (Kong et al., 2018) during work tasks. This is accomplished using ML-based discriminative approaches to computer vision (Fig. 3.8; Colyer et al., 2018), and these methods greatly increase the accessibility of measuring 3D kinematics for studying WMSD etiology.

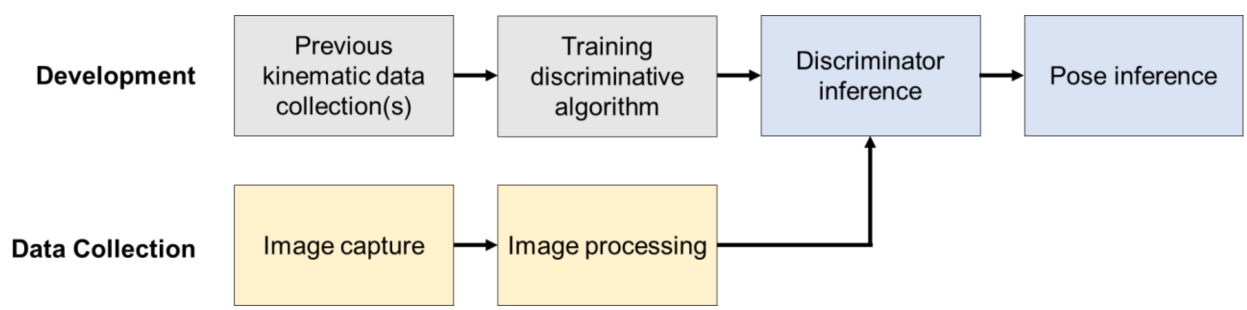


Fig. 3.8. Discriminative approaches for markerless motion capture are enabled by machine learning (ML), where image data are used to directly infer the shape, pose, and location (i.e., parameters) of the body in 3D space. Discriminative approaches do not require human body models, allowing quick processing by relying on discriminators trained using large databases of body parts, joints, or pose examples. Adapted from Colyer et al. (2018).

Another common application for ML models to help understand underlying mechanistic pathways was to estimate relevant biomechanical variables without reliance on musculoskeletal modeling (e.g., linked segment and finite element models), with the aim to reduce labour during setup and data processing. The biomechanical variables estimated were 3D joint kinematics (angle and angular velocity and acceleration) and kinetics (peak, mean, and cumulative moment, compression, and shear) mostly in the spine during manual handling tasks. These studies compared results obtained using ML-based methods to biomechanical models, demonstrating sufficient accuracy and insignificant differences. The studies in step 3 primarily posed regression problems, where ANNs were the most commonly used ML technique in this step. This was likely due to its applicability to discriminative computer vision as well as mapping the non-linear relationships between input kinematics to output biomechanical features in complex, whole-body movements. Although ANNs are considered "black box" models, interpretability may less crucial for problems in this step because the outputs are not directly involved in high-stakes decision making (Rudin & Radin, 2019). Overall, ML can aid in the study of the underlying mechanisms of WMSD by modeling the mechanical processes contributing to their development, increasing the accessibility of commonly-used biomechanical tools, and estimating the outputs of complex musculoskeletal models.

3.4.4 Development of interventions (Step 4)

Knowing when to intervene is a crucial step in WMSD prevention. To aid in knowing when to intervene, ML was frequently used for (semi)automated systems to permit real-time WMSD risk monitoring. This was the most common use for ML across all WMSD prevention steps, likely because this area is rapidly gaining research interest, with and without ML. Numerous unobtrusive, body-worn devices were used for this purpose, such as inertial measurement units (IMUs), smartphones, insole pressure systems, heart rate monitors, respiration monitors, EMG, electrocardiograms, and force-sensing chairs. Furthermore, several systems were developed using video cameras with and without depth sensors, allowing for easy workplace implementation. Worker risk was evaluated in several manners: automatically recognizing pre-defined "safe" or "unsafe" postures and quantifying the time spent in them; detecting the performance of tasks that may contribute to WMSD development (e.g., lifting, patient handling); quantifying the inputs needed for ergonomic assessments (e.g., RULA); and estimating/classifying workers' fatigue state. These problems mostly posed classification problems, with the exception of quantifying ergonomic assessment inputs and continuous fatigue estimation where regression was required. These systems were developed for manual handling, construction, laboratory, and seated work, and aim to help organizations encourage timely breaks, optimize work-rest ratios, and job rotations. Although relatively high accuracies were achieved throughout, these real-time monitoring systems are still in development and more research may be required to improve recognition accuracies (e.g., through the use of personalized models; Chapter 3.4.8), evaluate their effects, and integrate these tools into a larger system.

Another type of WMSD preventive measure developed using ML techniques were pre-employment screens to assess worker risk (i.e., do not provide real-time risk estimation). These approaches estimated workers' WMSD risk based on their movement strategies, anthropometrics, ergonomic assessments, questionnaires (e.g., Nordic Musculoskeletal questionnaire), or a combination of the above. Variables used to represent a WMSD risk (i.e., the outputs from their models) included load exposure level, strength capacity, and injury risk indices (developed using retrospective data). The aim of these interventions was to identify workers who may require ancillary training to mitigate WMSD risk. Considering the goals of these interventions and the use of the output (primarily for feedback), the interpretability of the ML models may not imperative, and favouring performance and speed for real-time use over interpretability (i.e., use of ANNs and DT-based models) may be justified. Because DT-based models, support vector machine (SVM), and *k*-NN were each used in 13 or more studies, this step was the most diverse in ML use across the framework for WMSD prevention research (van der Beek et al., 2017). This is likely attributable to growing research interest in recent years.

3.4.5 Evaluation of interventions (Step 5)

Interventions should be efficacious and effective, measured through thorough evaluation. ML was mainly applied for classification to support evaluation: assessing usability and acceptability of interventions with user feedback (Akanmu et al., 2020; Dingwell, 2006); developing a scoring system to rank intervention effectiveness (Ghasemi & Mahdavi, 2020); evaluating techniques to classify WMSD risk factor exposures (Samani et al., 2013); and quantifying the change in kinematic and muscular load risk factors post-intervention (Thanathornwong et al., 2014; Thanathornwong & Suebnukarn, 2015). The data used include questionnaires, interviews, rapid entire body assessments, cervical spine kinematics measured using accelerometers, and estimated muscular load data. Although most ML techniques were integrated into the interventions in this step, Bayesian network was implemented to understand the contribution of each factor to the acceptability of WMSD prevention software by Arbelaez Garces et al. (2016). Although the performance of the ML models used throughout the other steps this review were often evaluated using hold-out testing or cross-validation, only six included studies had a primary purpose to evaluate the effectiveness and/or practical feasibility of an intervention. Overall, ML techniques have not been used extensively in this area of WMSD prevention research and more opportunities may exist, especially for analyzing the results of intervention evaluation.

3.4.6 Implementation of effective interventions (Step 6)

Implementing effective and sustainable interventions can be challenging. This current review did not identify any studies leveraging ML to support implementation. This is in-line with other related reviews, where few studies are found for this purpose even without ML (Du et al., 2019). The lack of studies may be partially explained by the concentration of ML techniques used in the earlier steps of the WMSD prevention research framework (i.e., 1-4), most of which were published recently. Thus, more research employing ML techniques to study intervention implementation may arise after successful development and evaluation. Nonetheless, topics relevant to the implementation of WMSD preventative interventions represents an opportunity for further research.

3.4.7 Limitations

The findings from this review should be interpreted with consideration of the limitations. Although the subject headings (with explode command) and keywords used were carefully selected to capture as much of the relevant literature as possible, the literature search strategy had limitations. Limiting studies to English only meant that important studies in different languages were likely missed, considering China and Taiwan published the second and third most studies in a related review of ML for occupational accident analysis (Sarkar & Maiti, 2020). Furthermore, by excluding conference papers, additional relevant work may not be considered. Lastly, regression analyses (e.g., logistic, linear, multinomial) were not specifically searched as keywords. Although they are widely-used in ML, these keywords primarily returned studies employing regression analyses for traditional statistical approaches, thereby immensely inflating the number of search results. However, studies employing regression analyses indexed as machine learning were still found because the exploded machine learning subject heading was used wherever possible. In the end, a comprehensive search strategy was employed and these decisions were justified to ensure feasibility of conducting the scoping review (Levac et al., 2010) whilst including as much relevant work as possible.

Limitations also exist in the interpretations that can be drawn from the tallies of ML techniques used across each step of the framework for WMSD prevention. Because similar algorithms were grouped together (e.g., ANN and DT-based models), these tallies do not reflect the large variation and heterogeneity in the complexity of these algorithms. These groupings were necessary to ensure counting could be performed without making separate categories for each specific model architecture. Readers should not infer which ML techniques are "best" or superior based on these tallies, as all algorithms tend to perform equally well when averaged over all problems, and thus, are problem dependent (Adam et al., 2019). Instead, these data highlight what is frequently used and can provide insight into the types and frequency of problems that are faced in WMSD prevention. Readers are encouraged to read study summaries in tables 1-5 to gain a better understanding of why a ML model may have been suitable for each problem and are referred directly to the studies for full details on their development.

3.4.8 Future directions

Considering the trend in publications per year, the progress and use of ML for WMSD prevention research will likely continue to increase at a rapid pace; therefore, several suggestions are posed for future research. First, the development and application of interpretable ML models should be considered wherever possible. This can be incorporated in experiments by training and evaluating multiple ML models for comparison. The dichotomy between performance and interpretability of ML models may be false (Rudin & Radin, 2019), and other interpretable models may perform equally as well or better and should not be overlooked. Second, studies should look to include code and allow access to data wherever possible (even if through request only) to improve the reproducibility of ML studies. Although the majority of studies provided thorough explanations of the algorithms or provided pseudo-code, only 13 and 22 studies in this review provided code and data, respectively. Third, the development of real-time worker risk monitoring systems was the most popular area of research in this review. To further improve accuracy for the intermediary step of activity recognition (classifying task, posture, fatigue level, etc.), researchers should look to implement personalized ML models. This can be achieved via the use of hybrid models utilizing training data from known users and the unknown user (e.g., transfer learning; Hernandez et al., 2020), and the use of similarity weighting (Ferrari et al., 2020). For the state-of-the-art of human activity recognition methods, readers are referred to Yadav et al., (2021).

Another area for advancement is to continue looking for opportunities to leverage the capabilities of reinforcement learning, which are suitable for problems requiring goal-direct learning and decision making (Sutton & Barto, 2018). Across 130 reviewed studies, only one study employed reinforcement learning, which demonstrated promise for applications in virtual training environments (Akanmu et al., 2020). Lastly, the largest gaps in research for ML applications in WMSD prevention research were in steps 5 and 6, particularly in intervention implementation where no studies were found. In the field of implementation science, behaviour change interventions are being developed using ML-based systems that gain insight into behaviours to predict what interventions will work best given the context (Michie et al., 2017). Related approaches have also been implemented for safety behaviours surrounding high-risk occupational accidents (Gao et al., 2019, 2020; Goh et al., 2018; Guo et al., 2020), and could be extrapolated to the context of WMSD prevention.

3.5 Conclusion

Primary prevention is a cost-effective method to mitigate the risk of debilitating WMSDs. Although the use of ML techniques is relatively new for this purpose, they have enhanced WMSD prevention efforts across most stages of WMSD prevention (van der Beek et al., 2017) due to the diversity and versatility of ML algorithms available to learn complex interactions. The works surveyed indicate that many ML techniques have been implemented because a single, superior algorithm does not exist. Instead, the model with the greatest performance for each problem was determined by evaluating numerous models or selected based on background knowledge and/or the nature of the data. Since 1990, ML has reduced the labour in processing injury reports, while enabling exploration of large datasets to identify potential biomechanical, social, and psychological risk factors. ML has greatly improved the accessibility of biomechanical analyses used to study WMSD etiology, and many interventions have been developed to monitor real-time WMSD risk or predict which workers may sustain a WMSD. Some of the greatest opportunities for ML techniques are present in intervention evaluation and implementation research. The wealth of data made available by technological advancements invites the use of ML in WMSD prevention, and the vastly improved processing capacities ensure that they will continue to pave the way for future advancements.

Acknowledgements

The authors would like to acknowledge Nigèle Langlois of the University of Ottawa Health Sciences Library for her assistance in the development of the literature search strategies.

Chapter 4. Study 2: Raw Versus Fused Inertial Data for Machine Learning-Based Recognition of Manual Handling Activities

Submitted for publication as:

> **Chan, V. C.**, Graham, R. "*Raw Versus Fused Inertial Data for Machine Learning-Based Recognition of Manual Handling Activities*". Submitted to IEEE Transactions on Industrial Informatics.

Abstract

Human activity recognition (HAR) is an intermediary step in many occupational injury prevention interventions, where inertial measurement units (IMUs) are commonly used. However, it is unknown if using raw IMU outputs (RAW) versus processed sensor orientation and position data (FUSION) affect HAR performance. This study compared artificial neural networks trained with RAW versus FUSION for recognizing six manual handling activities. Cross-validation was used to evaluate performance and training time on a dataset comprised of over 52,000 activities performed sequentially by 24 participants until excessive fatigue occurred. The results showed marginally lower loss (μ_{RAW} = 0.1676, μ_{FUSION} = 0.1893, $p < 0.01$), greater accuracy (μ_{RAW} = 94.93%, μ_{FUSION} = 94.39%, $p < 0.01$), and greater weighted averaged F1-score (μ_{RAW} = 94.96%, μ_{FUSION} = 94.41%, $p < 0.01$) in RAW versus FUSION, with similar training times (p = 0.13). Comparisons between approaches, activity-specific classification performances, and recommendations for use are discussed.

4.1 Introduction

In 2022, an estimated 2.3 million occupational injuries and accidents occurred in the United States alone, leading to a median of 10 days away per incident (U.S. Bureau of Labor Statistics, 2023). These detrimental incidents often result from over-exposure to repetitive, strenuous movements that lead to excessive fatigue. Various methods exist to reduce exposure to occupational injury and accident risk factors, such as ergonomic assessments (Massiris Fernández et al., 2020), data-driven job rotations (Padula et al., 2017), and computing tradespace metrics (Mavor et al., 2023). However, many of these methods require knowing what activities are being performed by a worker, how frequently they are performed, and for what durations.

Human activity recognition (HAR) is a data-driven method to automatically recognize what activities are being performed that can be implemented unobtrusively in the field.

Typically requiring the collection of human movement data and classification algorithms, HAR is a growing method to classify movements for a wide variety of contexts, including occupational settings where the resulting information can be an important intermediary step for injury prevention systems and interventions (Islam et al., 2022). Machine learning techniques are often used for this purpose because of their ability to model non-linear relationships between variables in biological systems, such as human movement data and the activities being performed (Islam et al., 2022). Artificial neural networks (ANNs), in particular, can be used end-to-end to minimize the need for feature selection and engineering allowing them to be used for a variety of data sources with minimal processing (Murad & Pyun, 2017). Furthermore, ML models can be employed to make predictions quickly enabling HAR in near real-time. Thus, it is desirable to improve HAR to improve injury and accident mitigation systems.

4.1.1 State-of-the-Art

To obtain the necessary data to train and implement these models for HAR, numerous technologies have been employed, including (but not limited to) optical motion capture (Clouthier et al., 2020), instrumented insoles (D'arco et al., 2022), and inertial measurement units (IMUs) (Mavor et al., 2023). The benefits of using IMUs for this purpose are that they are inexpensive, can be worn above or under clothing/equipment, can be easily employed in real-world situations, and require comparatively less setup and processing compared to other technologies.

Modern IMUs are often comprised of three micro-electromechanical components: tri-axial accelerometer, gyroscope, and magnetometer, measuring three-dimensional linear acceleration, angular velocity, and magnetic field, respectively. On their own, the raw signals are not directly useful for human movement analysis, requiring sensor fusion to obtain orientation and position. Examples of common sensor fusion algorithms include Kalman, Madgwick, and complementary (Nazarahari & Rouhani, 2021), while other approaches perform multi-sensor fusion using machine learning (Qiu et al., 2022). Despite the vast array of sensor fusion algorithms, at a higher level, they have similar objectives when obtaining orientation and position: initiate the sensor orientation using gravity, perform dead reckoning using accelerometer or gyroscope data, and correct for bias using the other remaining signals. Broadly, methods vary in how the gain between signals are modulated and how they are fused (Nazarahari & Rouhani, 2021). In the context of using IMUs for HAR, sensor fusion to obtain sensor orientation and position has been applied (Bances et al., 2022; Del Rosario et al., 2019) and forgone (Khatun et al., 2022, 2023; Mavor et al., 2023), though both typically lead to satisfactory classification performances. While special attention has been focused on other processing methods (e.g., transforming time series data to images; Ramanujam et al., 2021) little discussion or rationale has been presented justifying the use or omission of sensor fusion processing prior to performing HAR. Thus, investigation on the differences in methods is warranted.

4.1.2 Knowledge Gap

The use of IMU data without (RAW) and with (FUSION) sensor fusion processing has not been directly compared in research and thus, it is unknown if and how beneficial performing the additional step of sensor fusion is for HAR performance and training time.

4.1.3 Objectives

The objectives of the current study are as follows:

1) Compile a large kinematic dataset of individuals performing a repetitive manual handling protocol comprised of walking, standing, and pulling, lifting, carrying, and pushing a crate until exhaustion.

2) Train fully connected ANNs to automatically recognize the manual handling task performed using IMU data without (RAW) and with (FUSION) sensor fusion processing to obtain sensor orientation and position.

3) Evaluate the performance (loss, accuracy, weighted average F1-score, and training time) of the ANNs on the RAW and FUSION datasets using *k*-fold cross-validation.

4) Statistically compare the performance metrics between the ANNs trained using RAW versus FUSION datasets using Student's *t*-tests.

It was hypothesized that the approaches would lead to different classification performance and training times because sensor fusion introduces new variables and reduces the number of features (i.e., signals), respectively.

4.1.4 Contributions

- Provide a performance comparison between RAW and FUSION IMU data for HAR. Methods were kept consistent between approaches for direct comparisons.
- Compile a large kinematic dataset of manual handling activities, performed until fatigue. The amount of data exceeds prior work; together with the induction of fatigue, large inter- and intra-individual kinematic variability was introduced for this investigation (Jimale & Mohd Noor, 2023). Exposure to variability can make HAR models more robust to unseen data and, in conjunction with *k*-fold cross-validation, strengthen the confidence in performance estimates (Jimale & Mohd Noor, 2023).
- Provision of evidence to inform researchers and developers on what data processing approach may be suitable for their IMU-based HAR application.

4.2 Participant Recruitment

This research was approved by the Research Ethics Board at the University of Ottawa. Twenty-four participants from the Ottawa area were recruited for this study via convenience sampling. All participants were cleared to participate in physical activity, screened using the physical activity readiness questionnaire (PAR-Q+; Appendix D) (Warburton et al., 2021). To reduce the risk for injury resulting from overexertion and/or accidents, participants must not have had any injuries or musculoskeletal disorders in the 12 months prior to their session. Lastly, participants were required to be 16 years of age or older so that they could provide informed consent.

4.3 Experimental Protocol

A diagram of the study workflow is presented in Fig. 4.1. Upon arrival at the University of Ottawa Movement Biomechanics and Analytics Laboratory, participants were asked to read over the research consent form (Appendix E), before receiving a verbal explanation and an opportunity to ask questions. Next, they completed the PAR-Q+, then signed the informed consent form if they were eligible and agreed to participate.

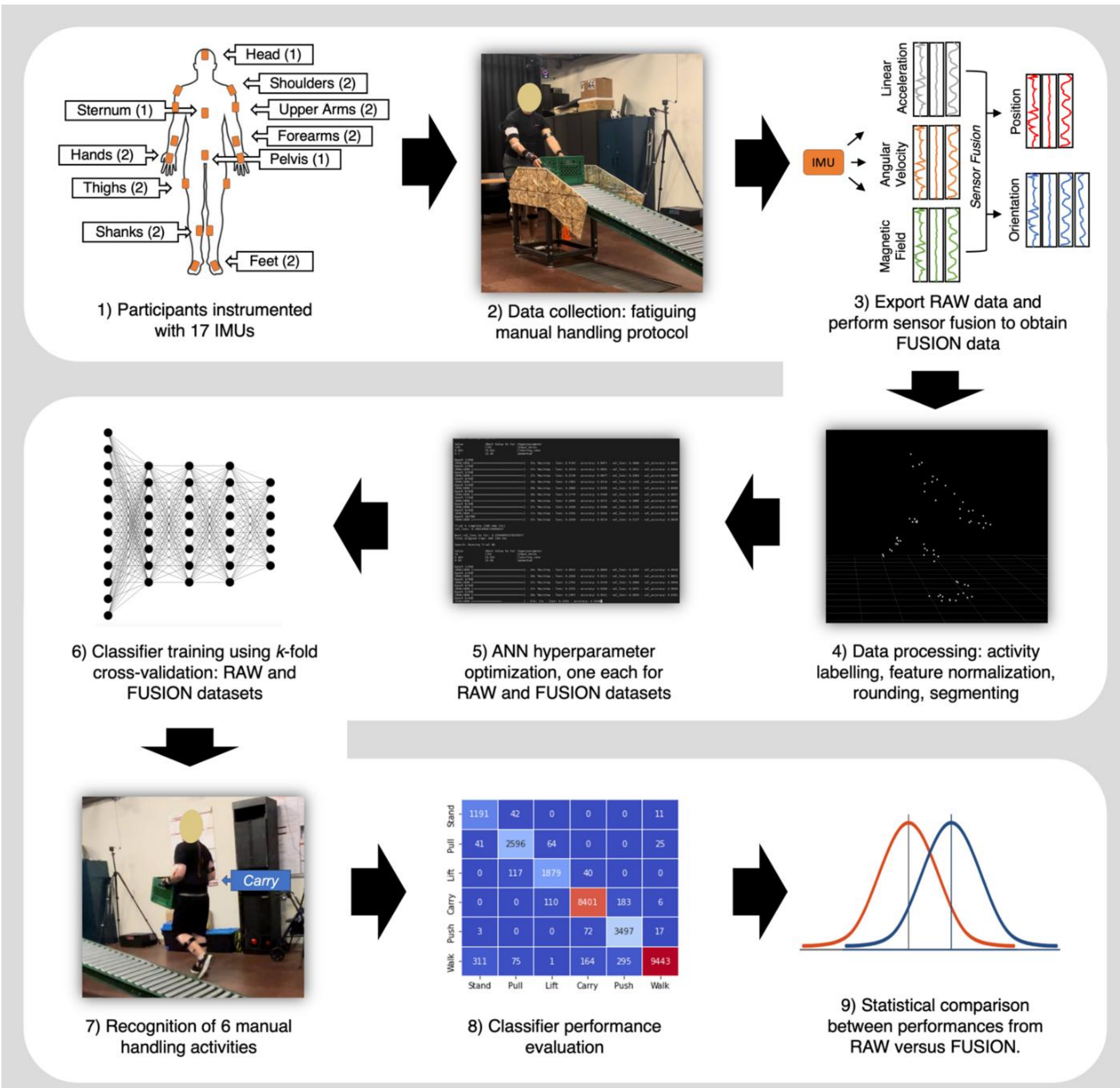


Fig. 4.1. Diagram of the study workflow. IMU = inertial measurement unit; ANN = artificial neural network.

4.3.1 Instrumentation & Calibration

The participants changed into athletic clothing and footwear. Then, an Xsens MVN Link inertial motion capture system (Xsens, Enschede, Netherlands) was worn to capture whole body kinematics. This system consisted of 17 IMUs secured/worn by the participants using neoprene bands and a tight-fitting Xsens shirt: one on the head, sternum, and pelvis; and bilaterally on the shoulders, upper arms, forearms, hands, thighs, shanks, and feet. To calibrate the IMU system, participants were asked to stand in a neutral posture, walk for 5 s forward, turn around, and return to the starting location and position. The kinematic data were collected using Xsens MVN Analyze 2022.0.2 software (Xsens, Enschede, Netherlands), sampled at 120 Hz. These data were collected as part of a larger study where other data were concurrently collected; only the relevant instrumentation have been presented.

4.3.2 Manual Handling Protocol

This protocol was designed to allow common manual handling activities to be performed in a repetitive manner. A fatiguing protocol was selected for this purpose because it elicits a large number of repetitions of the activities of interest and because fatigue is a perturbation that can elicit individual-specific changes in movement behaviours (Cortes et al., 2014).

Participants received an explanation of the protocol and a practice opportunity for familiarization. The manual handling protocol was a circuit where participants were required to pull, lift, carry, and push a crate (8.1 kg; 30 cm tall × 33 × 33 cm; handles for good coupling), followed by walking. A diagram of the manual handling protocol is presented in Fig. 4.2. Participants were asked to push and pull the crate using both arms and were instructed to follow a pace of approximately 15 s per round (i.e., one sequence of pulling, lifting, carrying, pushing, and walking), enforced via a metronome. Aside from these constraints, participants were free to perform the activities with self-selected technique. One set was comprised of 20 consecutive rounds (lasting 5 min), with a fatigue assessment before and between each movement set (lasting approximately 2 min). Prior to each set, participants stood upright at position A) facing C) in Fig.

10, while the heading and position were reset in Xsens MVN Analyze; then, participants performed a stomp to indicate the start of each trial.

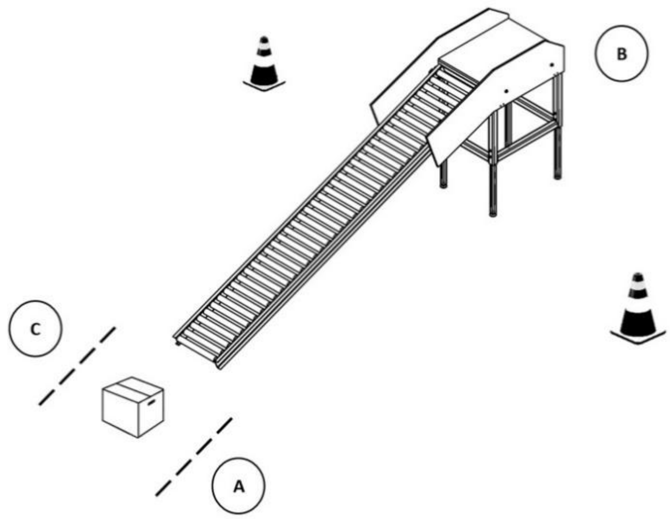


Fig. 4.2. The manual handling protocol. Participants were instructed to start at A) behind the line (0.5 m away from the crate), descend, and pull the crate towards themselves. They lifted the crate, turned, and carried it to B) by walking around the pylon (approx. 3 m). At B), they released the crate onto the low-friction table (0.75 m tall) and pushed it across and down the roller conveyor to the initial position between A) and C). Then, participants walked around the pylon to C), where they performed the same sequence except on the other side of the roller conveyor. Performing all activities from A) to C), or C) to A), was considered one round (i.e., one sequence of pull, lift, carry, push, and walk).

The fatigue assessment was comprised of a subjective fatigue rating using a 10 cm visual analog scale (VAS) (Chiarotto et al., 2019) and a maximal lift strength assessment (Zemková et al., 2019). To measure their fatigue VAS, individuals were asked to make a mark along the 10 cm VAS on a piece of paper to holistically indicate their perceived fatigue. The fatigue VAS was anchored on each end by the following statements: “No fatigue at all”, and “The most fatigue I’ve ever experienced in my life, and I cannot continue”. Participants were always blinded to their previous response on the fatigue VAS. The maximal lift strength assessment was comprised of three, maximal effort static deadlifts, with approx. 5 s break between attempts. The lift handle on an s-type load cell was set to knee-height, and participants were instructed to "ramp" up the force on the handle over 3 s. These data were collected using custom LabVIEW software (National Instruments, Austin, TX, USA), sampled at 100 Hz. Participants never knew the results of their maximal lift strength assessments and were instructed to maintain the same technique throughout the study to the best of their abilities. Once completed, they immediately began the next set of fatiguing manual handling protocol. They were asked to perform as many sets of the fatiguing manual handling protocol as possible until any of these conditions were met: their VAS became ≥ 9 cm; their maximal lift strength became < 70% of their baseline value; they could not maintain pace (15 s per round) for 3 consecutive rounds; or they wished to stop. The data from the fatigue assessments were only used to determine when to stop data collections due to excessive fatigue. Once the study ceased, the Xsens suit was removed from the participants. Due to the nature of the protocol (i.e., participants continuing until a stopping condition was met), the length of each data collection session varied but were always completed within the same day.

4.4 Data Processing

All IMU data were collected on a Dell desktop (Dell Inc., Round Rock, USA) while maximal lift force was collected on a Dell laptop, both running Windows 10 (Microsoft, Redmond, USA). Data processing was conducted using an Acer desktop (Acer, New Taipei City, Taiwan) running Windows 11 and an Apple M2 Mac Mini desktop (Apple Inc., Cupertino, USA) running Sonoma 14.

4.4.1 IMU Data

Each Xsens IMU contains an accelerometer, gyroscope, and magnetometer, which quantify three-dimensional linear acceleration, angular velocity, and magnetic field, respectively. These data were processed in Xsens MVN Analyze using the high-definition reprocessing tool which employs sensor fusion to obtain sensor orientation and position (Roetenberg et al., 2013; Schepers et al., 2018). For sensor orientation, this tool integrated the angular velocity data obtained from the gyroscope (i.e., via dead reckoning). Then, the gravitational and magnetic components from the accelerometer and magnetometer were used to stabilize the signal and minimize the effects of integration drift. For sensor position, the accelerometer signals were double integrated to obtain position. Gravity was compensated by applying the estimated sensor orientation then subtracting gravitational acceleration. The results from these sensor fusion algorithms have been validated against optical motion capture systems for whole-body movements (Mavor et al., 2020; Robert-Lachaine et al., 2017).

Raw linear acceleration, angular velocity, and magnetic field signals from the accelerometer, gyroscope, and magnetometer, respectively, from all 17 IMUs were exported. Additionally, fused sensor data were exported: sensor orientation (in quaternions) and position. In total, 16 signals per IMU were exported: 9 RAW and 7 FUSION.

4.4.2 Activity Labelling

The primary researcher labelled all movements as being one of six manual handling activities (i.e., stand, pull, lift, carry, push, and walk) by visually observing the simulated markers computed from the Xsens data in Mokka software (Biomechanical ToolKit, Capitale-Nationale, Canada). Significant movements that were not one of the aforementioned activities were labelled as “other”; an example of a movement labelled as “other” was knocking over a pylon then stopping to fix it. The start point of each set was defined as the moment the stomp was completed in the Xsens data. All data prior to the completion of the stomp were discarded.

4.4.3 Data Preparation

A single matrix was created using data from all participants comprised of 272 features (16 signals × 17 IMUs). The dataset was constructed by concatenating the data from all movement sets, across all participants. The activity label was included for each frame of data, describing that frame to be part of one of 7 activities; then, all frames with the activity label “other” were removed. Feature normalization was then conducted on a feature-by-feature basis by subtracting the mean and dividing by the standard deviation of all data frames across participants and movements in the corresponding subset (Clouthier et al., 2020; Mavor et al., 2023). Lastly, all data were rounded to four decimal places.

The RAW and FUSION datasets were subsets of the entire dataset. RAW only included the raw IMU signals (9 signals × 17 IMUs = 153 features) while FUSION only included the fused, sensor orientation and position signals (7 signals × 17 IMUs = 119 features).

A sliding window approach was used to divide the data into segments containing an equal number of data frames (i.e., 60 or 120), where the step size was always 1/4 of the window size (Clouthier et al., 2020; Mavor et al., 2023). Each segment was assigned a label according to the activity performed for the majority of that segment. The dimension of the input data segments for the ANN was the sliding window size by the number of features.

4.5 Machine Learning-Based Activity Recognition

Three Nvidia Titan RTX GPUs (Nvidia, Santa Clara, USA) with 24 GB of GDDR6 memory each were used to perform hyperparameter optimization and *k*-fold cross-validation using a distributed mirrored strategy. The GPUs were installed on a Lambda Labs server (Lambda Labs, San Jose, USA) powered by two Intel Xeon Gold 6248 CPUs @ 2.50 GHz (Intel, Santa Clara, USA), 384 GB of DDR4 RAM, and running Ubuntu 18.04. All machine learning development was conducted using Python 3.10.12 (Van Rossum & Drake, 2009) and TensorFlow 2.12 (Abadi et al., 2016).

4.5.1 Fully Connected Artificial Neural Network

A fully connected ANN was selected because it can be used end-to-end to automatically extract features and strong performances for IMU-based HAR have been shown (Mavor et al., 2023). Despite other architectures achieving stronger classification performances (Khatun et al., 2022; Mavor et al., 2023), the performance differences were not large while this ANN avoids excessive complexity. This can result in faster training times while requiring less memory (Mavor et al., 2023).

The ANN was comprised of an input layer, three fully connected dense hidden layers with rectified linear unit (ReLU) activation functions, and a dense output layer with a Softmax activation function (Fig. 4.3). A drop-out operator with a probability set to $p = 0.5$ was implemented on the output of the dense layers for regularization. The reshaped IMU data segments were the ANN inputs, and the corresponding ground truth labels were used to compute loss (i.e., error). The dense hidden layers consisted of hidden units (i.e., neurons or nodes), which computed an element of the layer's output. The input into each hidden layer was the output of the previous layer. A set of randomly initialized parameters were used to control the transformations for each hidden unit and layer. Feed-forward propagation through the ANN was used to obtain a vector representing the probability of the input belonging to each class. Prediction loss was computed using the categorical cross-entropy function. During training, the gradients of the cost function with respect to the parameters were computed using backpropagation, where stochastic gradient descent (SGD) was used to update the parameters to minimize loss. This was performed iteratively until a stopping criteria was reached.

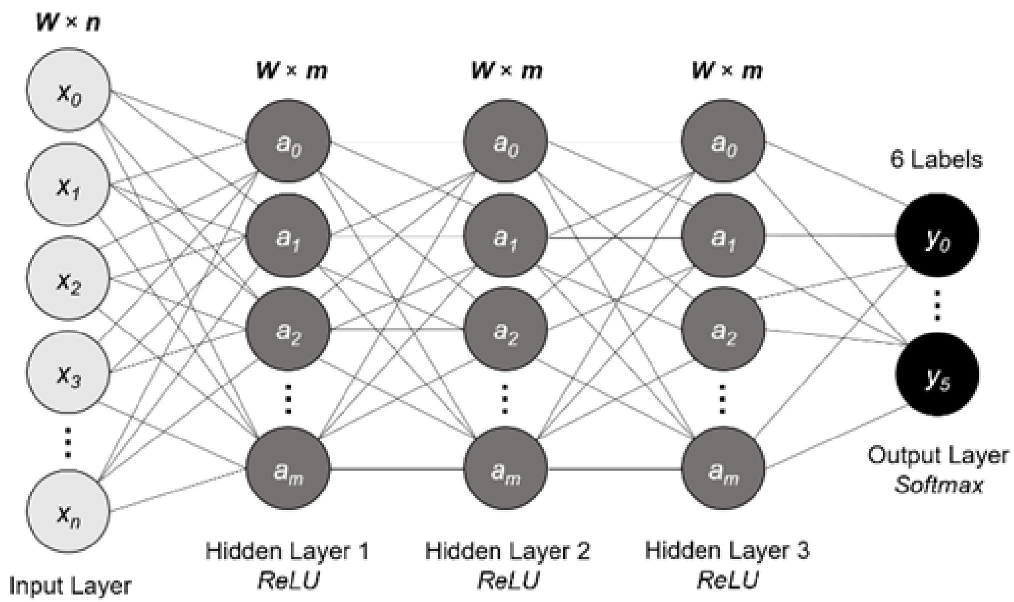


Fig. 4.3. Fully connected artificial neural network architecture. The input features (n) were either 9 (RAW) or 7 (FUSION) signals per IMU. W = sliding window size; m = number of hidden units; x = input units; a = hidden units; y = output units; ReLU = rectified linear unit.

4.5.2 Hyperparameter Optimization

The SGD hyperparameters, the sliding window size, and the number of hidden units in the dense layers were tuned via Bayesian hyperparameter optimization implemented using the KerasTuner library (O'Malley et al., 2019). The hyperparameter search space is shown in Table 4.1, where the selected values are bolded. The creation of this search space was guided by other research with similar tasks and instrumentation (Clouthier et al., 2020; Mavor et al., 2023). The objective was to minimize loss on the validation set, comprised of all data from three, randomly selected participants: 4, 11, and 18. For each set of hyperparameters, one trial was executed and was stopped early based on non-improving training accuracy with a patience of 2. Data were loaded with a batch size of 64 and hyperparameter optimization was limited to 24 trials per dataset. This hyperparameter optimization procedure was performed once per dataset, which led to different sets of hyperparameters for RAW and FUSION.

4.5.3 *k*-fold Cross-Validation

To train and evaluate the RAW and FUSION models, *k*-fold cross-validation was implemented. Based on the dataset size and the requirement to keep all data from a participant together (to prevent exposure to both training and evaluation), the value was set to $k = 8$, where each fold was comprised of all data from three participants. The folds were randomly selected, then kept constant across RAW and FUSION datasets.

Prior to training in each fold, a class weight dictionary was computed for the training set using the "balanced" setting from the SciKit-Learn library, which implemented a heuristic method that utilizes logistic regression to deal with rare events (Pedregosa et al., 2011). These class weights were implemented during training to mitigate the effects of class imbalance that varied in magnitude, depending on the fold.

Each cross-validation procedure was trained using a batch size of 64, maximum epochs of 500, and data were randomly shuffled prior to each epoch to avoid order effects. Training was stopped early based on non-improving training accuracy with a patience set to 3. Then, the model weights from the epoch with the lowest validation loss were saved and used to make predictions on the test fold. Performance on each fold was evaluated using loss, accuracy, weighted average F1-score, training time, and confusion matrices. Performance for each activity was evaluated using precision, recall, and F1-score.

Table 4.1. The hyperparameter search spaces for the models trained using RAW and FUSION datasets.

Dataset	Window size	Hidden units	SGD learning rate	SGD momentum
RAW	60, **120**	**64**, 128, 192, 256, 320	0.0001, **0.001**, 0.01, 0.1	0.7, 0.9, **0.95**, 0.98
FUSION	60, **120**	**64**, 128, 192, 256, 320	0.0001, 0.001, **0.01**, 0.1	0.7, **0.9**, 0.95, 0.98

The bolded values represent the selected hyperparameters; SGD = stochastic gradient descent.

4.6 Statistical Analysis

To test the hypothesis that performance differs between models, paired Student's t-tests were performed between RAW and FUSION for each performance metric, across all folds. For all tests, significance was set to α = 0.05. Mean and standard deviation of each performance metric were also computed for RAW and FUSION models to obtain a robust estimate of performance. All statistical analyses were performed using R version 4.3.2 (R Core Team, Vienna, Austria).

4.7 Results

4.7.1 Participants

Participant demographics are presented in Table 4.2 and the tally of reasons for stopping are presented in Table 4.3.

Table 4.2. Participant demographics.

	Age (yr)	Height (cm)	Mass (kg)	Fatiguing sets (no.)
M (*n* = 13)	27.5 (7.2)	178.6 (6.5)	79.8 (11.0)	19.9 (12.0)
F (*n* = 11)	25.6 (2.9)	168.9 (8.9)	68.2 (11.6)	16.5 (7.2)
All (*n* = 24)	26.7 (5.7)	174.2 (9.0)	74.5 (12.5)	18.4 (10.0)

Values are presented as mean (standard deviation). M = male; F = female.

Table 4.3. Tally of reasons for stopping the manual handling protocol.

Reason	Occurrences (No.)
MLS only	11
VAS only	6
MLS and VAS	3
Failure to maintain pace	0
Voluntary	4

MLS = maximal lift strength; VAS = fatigue visual analog scale

4.7.2 Dataset

In the dataset, a total of 441 manual handling sets were performed, totaling 36.75 hours of data where 8,820 rounds (i.e., one sequence of manual handling activities) or approximately 52,920 manual handling activities were performed. This amounted to 15,852,125 frames of data; after removing the frames labelled as “other”, 15,849,914 frames remained (0.0139% of data were removed). The number of frames labelled as each activity are shown in Fig. 4.4.

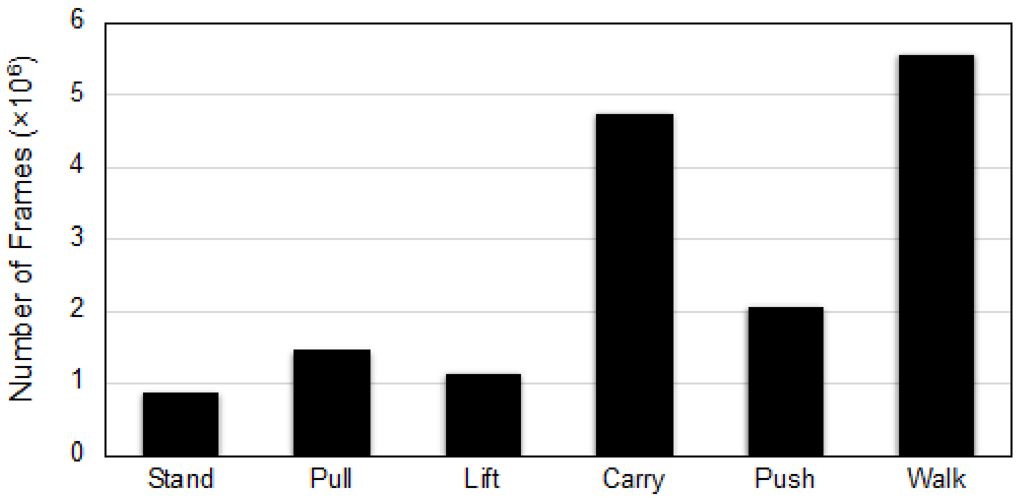


Fig. 4.4. The number of frames for each activity across all participants and manual handling sets. Movements that did not fall under these categories were labelled as "other", then removed (0.0139% of all data).

4.7.3 Classification Performance

Sample learning curve graphs for accuracy and loss across training epochs are shown in Fig. 4.5. The loss, accuracy, weighted average F1-score, and training time of the ANN models are shown in Fig. 4.6 and Table 4.4. The activity-specific performance metrics are shown in Table 4.5. Confusion matrices of RAW and FUSION models summed across all eight folds for *k*-fold cross-validation are presented in Fig. 4.7.

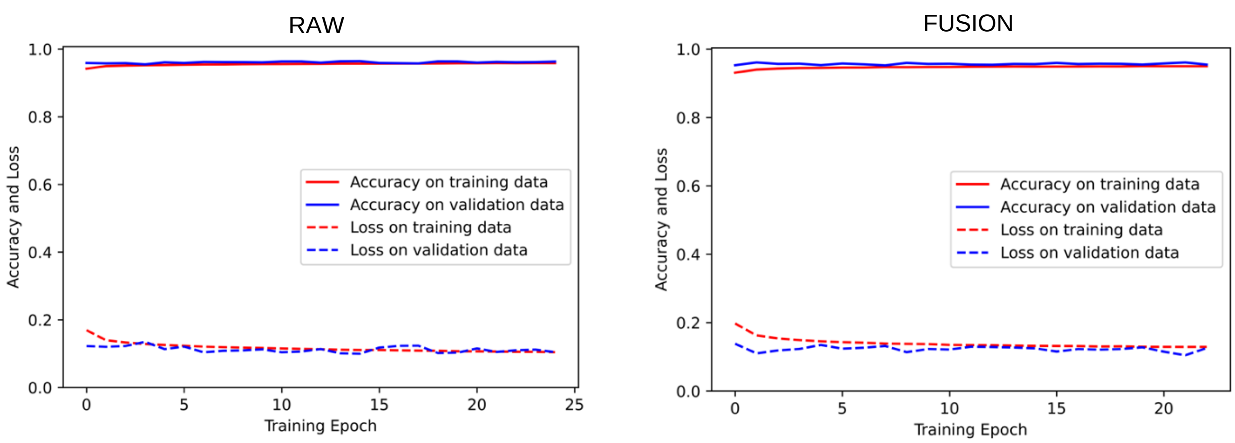


Fig. 4.5. Sample RAW and FUSION model learning curves for fold 1.

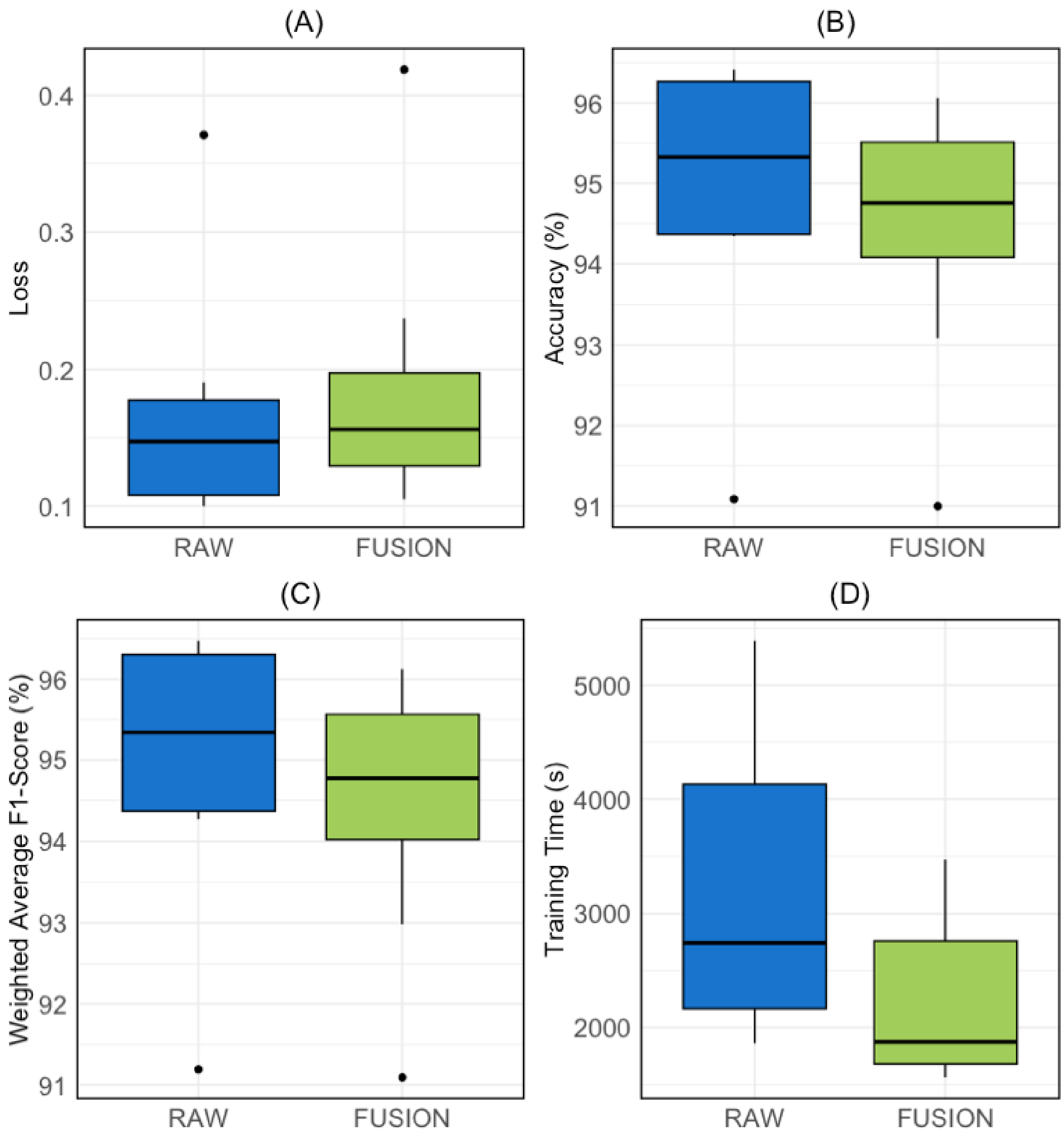


Fig. 4.6. Boxplots of (A) loss, (B) accuracy, (C) weighted average F1-score, and (D) training time for RAW and FUSION across all eight cross-validation folds. The boxes show the interquartile range, centre lines show the medians, whiskers show minimums and maximums, and dots show the outliers. Significant differences were observed between RAW and FUSION for all comparisons except training time.

Table 4.4. The performance and training time of the RAW and FUSION models averaged over cross-validation folds where k = 8.

Dataset	Loss	Accuracy (%)	Weighted average F1-Score (%)	Training time (s)
RAW	0.1676 (0.0888)	94.93 (1.77)	94.96 (1.75)	3174.7 (1283.2)
FUSION	0.1893 (0.1013)	94.39 (1.64)	94.41 (1.65)	2246.5 (756.3)

Values are presented as mean (standard deviation). Greyed values indicate a significant difference between RAW and FUSION.

Table 4.5. The precision, recall, and F1-score (%) of each activity for RAW and FUSION averaged over cross-validation folds where k = 8.

Metric	Dataset	Stand	Pull	Lift	Carry	Push	Walk
Precision	RAW	85.11 (7.93)	94.69 (2.19)	90.49 (3.80)	97.95 (1.13)	89.68 (5.76)	96.69 (2.47)
	FUSION	84.11 (6.24)	94.12 (1.23)	89.39 (5.09)	97.40 (1.30)	89.57 (5.01)	96.23 (2.71)
Recall	RAW	96.13 (1.85)	94.85 (1.95)	94.99 (2.33)	96.01 (2.57)	90.22 (6.69)	95.77 (1.61)
	FUSION	94.85 (2.39)	94.97 (2.72)	94.71 (2.17)	95.64 (2.46)	88.24 (7.13)	95.42 (1.16)
F1-score	RAW	90.06 (4.60)	94.76 (1.79)	92.64 (2.32)	96.95 (1.38)	89.77 (4.59)	96.21 (1.74)
	FUSION	89.03 (3.64)	94.53 (1.82)	91.88 (2.70)	96.49 (1.32)	88.71 (4.41)	95.81 (1.60)

Values are presented as mean percentage (standard deviation). Greyed values indicate a significant difference between RAW and FUSION.

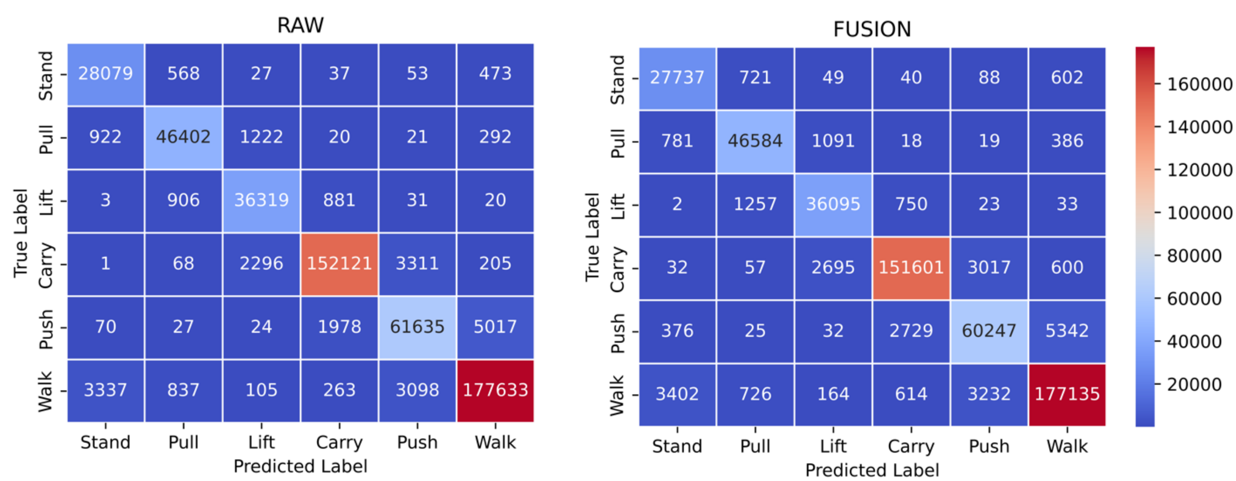


Fig. 4.7. Confusion matrices of the RAW and FUSION models summed across all eight cross-validation folds.

4.7.4 RAW Versus FUSION Comparison

The RAW model yielded significantly lower loss ($t(7) = -3.57$, $p < 0.01$), greater accuracy ($t(7) = 3.57$, $p < 0.01$), and greater weighted averaged F1-score ($t(7) = 3.74$, $p < 0.01$) compared to the FUSION model. There was no difference in training time ($p = 0.13$).

4.8 Discussion

The aim of the current work was to compare the performance and training time of an HAR model to classify fatiguing, manual handling activities measured using IMUs without (RAW) and with (FUSION) sensor fusion. The findings showed equivalent or better performance for RAW when compared to FUSION. Specifically, training on the RAW dataset led to marginally improved loss, accuracy, and weighted average F1-score, with no difference in training time when evaluated with *k*-fold cross-validation. Thus, the findings partially supported the hypothesis: performance differed between raw versus fused IMU data, but no difference was observed between approaches for training time.

Both approaches achieved accuracies and F1-scores ≥ 94.39%, which was marginally better (Bances et al., 2022; Bangaru et al., 2021) and worse (Hosseinian et al., 2019; K. Kim & Cho, 2020) than other works that used IMUs to capture similar activities. It should be noted that few HAR studies evaluate their performance using *k*-fold cross-validation, which in comparison to hold-out testing, may provide a more robust estimate since it does not depend on one train-test split. However, the performances in the current study are still strong when considering the size of the dataset, the inter-individual variability from the 24 participants, and the intra-individual variability from the lack of movement constraints and the development of excessive fatigue. Furthermore, no activity was poorly classified in this study in either RAW or FUSION approaches despite the class imbalance present in the dataset: the minimum and maximum average F1-scores were 89.03% for "stand" in FUSION and 96.95% for "carry" in RAW, respectively. This was likely attributable to large number of training samples and the use of class weighting.

To delve into why marginal yet statistically significant differences existed between RAW and FUSION approaches, activity-specific differences were examined. The recall and F1-scores for "push" were significantly greater in RAW compared to FUSION. This difference was speculated to be due to the nature of the activity – a pushing motion may be characterized by a large acceleration of the upper extremities directed anteriorly. Linear acceleration was only directly available in the RAW dataset, whereas it is a double derivative of the position signal in the FUSION dataset. Given that the window size was the same (i.e., 120 frames or 1 s) in both HAR approaches, RAW may have had more frames of relevant information pertaining to the "push" activity per window. RAW also elicited minor improvements in precision for "walk" and F1-scores for "lift" and "carry", likely due to the importance of lower extremity angular velocity in these activities. In the same line of reasoning for the "push" activity, angular velocity was directly available in RAW, whereas it is only indirectly observable as a derivative of the orientation signal in FUSION. Overall, the performance differences between RAW and FUSION were minute and are unlikely to lead to noticeable detriments upon implementation.

4.8.1 Limitations

This study was presented with two main limitations. First, only one architecture was employed for machine learning. The ANN was selected because of its suitability for the type and size of this dataset. Although only one architecture was used, no interaction effects were observed in other related research where various model architectures were trained using the same dataset (Mavor et al., 2023; Wan et al., 2020). Furthermore, the ANN presented in this work is commonly-used and is widely applicable for a variety of use cases (Victor C.H. Chan et al., 2022). Thus, the results of the current RAW and FUSION comparison may still hold regardless of the model architecture. Despite this, future work should explore this comparison with other classification methods, including shallow learning algorithms (e.g., support vector machines) or hybrid architectures (e.g., neural networks with inception-attention (Mim et al., 2023) for further validation. The second limitation is that feature selection was not performed to reduce the feature sets. This would have increased training speed by alleviating memory requirements and reduced the chances of overfitting without major detriments to performance (K. Kim & Cho, 2020). However, feature selection was not performed for either dataset while the nature of ANNs allows it to be helpful for end-to-end workflows (i.e., emphasis is placed on the salient features) (Murad & Pyun, 2017).

4.9 Conclusion

This study found marginally greater performance when using raw signals from IMUs for machine learning-based manual handling activity recognition compared to using fused sensor orientation and position signals, with no difference in training time. Because these differences were minor, fused signals should be used if: 1) sensor orientation and position are required later in the workflow; 2) if a method to reduce dimensionality is required; and 3) the additional processing time does not pose a constraint / fused data are output automatically from the device/system. The results from these findings should inform future creation of HAR models, which are an important intermediary step for injury and accident mitigation systems in occupational contexts.

Chapter 5. Study 3: Sensor data required for the automatic recognition of fatiguing manual handling activities

Submitted for publication as:

Chan, V. C., Graham, R. *"Sensor data required for automated activity recognition of manual handling tasks"* Submitted to Human Factors.

Abstract

Objective: Perform analysis on the quantity and placement of inertial measurement units (IMUs) needed to recognize fatiguing manual handling activities. **Background:** Wearable sensors, such as IMUs, can be used to unobtrusively perform human activity recognition (HAR). This is an intermediate step in novel interventions to mitigate work-related musculoskeletal disorder and accident risk. However, the quantity and placement of IMUs impact the HAR performance and implementation cost; thus, it would be beneficial to know how quantity and placement of IMUs affect HAR performance. **Methods:** 24 participants performed rounds of manual handling tasks with a crate until exhaustion while their kinematics were measured using an inertial motion capture suit and smartwatch. These data were divided into 18 subsets (each corresponding a sensor combination), then used to train fully connected and DeepConvLSTM neural networks to perform HAR. **Results:** All permutations of sensor combinations and model classified manual handling activity with ≥ 89.98% accuracy and 90.11% weighted average F1-score. One IMU with a deep neural network could classify manual handling activity with 93.17% accuracy when evaluated using *k*-fold cross-validation, up to a maximum of 94.26% accuracy when six IMUs were used. **Conclusion:** While all sensor combinations studied led to strong classification performances, more IMUs tended to improve performance, while placement on the upper body or extremities are recommended for recognizing manual handling activities. **Application:** How the quantity and placement of IMUs affects performance of manual handling activity recognition was presented, and this can inform future projects on what IMUs to include.

Keywords: manual materials handling, kinematics, artificial intelligence, job risk assessment, physical ergonomics

5.1 Background

Workers who perform manual handling are at risk for sustaining musculoskeletal disorders (MSDs) and accidents, which can lead to large compensation costs while impacting workers' quality of life and productivity (U.S. Bureau of Labor Statistics, 2023). As such, the primary prevention of work-related MSDs and accidents is of utmost importance. With technological advancements, collecting large amounts of data is possible and machine learning has made it feasible to leverage these data to develop effective interventions. Novel data-driven strategies have been put forth to mitigate work-related injury risk, including automatic ergonomic assessments (Giannini et al., 2020), risk prediction and management (Nath et al., 2018), and worker fatigue modelling (Maman et al., 2020). These methods can improve primary MSD and accident prevention efforts by informing timely breaks (Hong et al., 2023), job rotation or organization (Maman et al., 2020), or feedback to improve posture and kinematics (Owlia et al., 2020), and often rely on human activity recognition (HAR) to function.

The increased availability of wearable sensors, namely inertial measurement units (IMUs), has reduced the barrier to implement automated HAR. IMUs can be worn unobtrusively on or underneath equipment and are not limited to pre-defined capture volumes like cameras. However, there are considerations for implementing wearable sensors in the workplace: they must be perceived as acceptable by workers (Bootsman et al., 2019; Jacobs et al., 2019) and should be cost-effective for employers (Schall Jr et al., 2018). Thus, it is desirable to use fewer IMUs when developing MSD and accident prevention systems to reduce hindrance to users and costs for employers. Using fewer IMUs can also save time and resources for researchers and developers due to less equipment, setup, and data processing. However, it is unclear how many sensors, what placement, and which combinations should be used for automatic recognition of manual handling activities whilst achieving satisfactory classification performance.

There has been variation in both the quantity and placement of sensors for automatic recognition of occupational activities. Studies have used a single IMU, such as a smartphone on the upper arm (Akhavian & Behzadan, 2016) or an activity tracker on the wrist (Ryu et al., 2019) for construction activities, and an IMU on the back for military activities (Mavor et al., 2023). Three IMUs have been used on the waist and both ankles for load carrying activities (H. Lee et al., 2020), four IMUs attached to both wrists, sternum, and right ankle for masonry activities (Hong et al., 2023), and eight IMUs worn on the thoracic spine, lumbar spine, and bilaterally on the upper arms, forearms, and thighs for manual handling activities (Yan et al., 2021). Other works have employed inertial motion capture systems comprised of seventeen or more IMUs to classify activities during manual handling (Bastani et al., 2016; S. Kim & Nussbaum, 2014) and manufacturing (Roth et al., 2020). The variation in sensor quantity and placement is partially attributable to differences in activity: waist or back IMU placement can provide insight into centre of gravity, whereas placements on the arm can provide insight into hand movements (Ryu et al., 2019). Placement variation can also be attributed to the minimum classification performance needed for a given application. Thus, understanding and quantifying the trade-offs between performance and number/combination of IMUs can help inform selection for future work requiring automatic recognition of manual handling activities.

Few efforts have been directed towards studying sensor quantity and placement for manual handling activity recognition. One group evaluated 32 IMU combinations for their ability to recognize manual handling activities (K. Kim & Cho, 2020). On one participant, their random forest classifier achieved similarly strong classification performances between twenty-one (whole body) and two (hip and neck) IMUs; then, a long short-term memory neural network (LSTM) was trained using the two-IMU subset to classify activities from three additional participants with a hold-out test accuracy of 94.73% (K. Kim & Cho, 2020). While many sensor combinations were studied, more training samples and participants are warranted to ensure the findings are robust to inter- and intra-individual variability in movement patterns. Another group studying IMU placements for manual handling activity recognition involved ten participants and nine sensor combinations (Porta et al., 2021). This work incorporated greater inter-individual variability while achieving strong classification performance, but the number of trials may have limited intra-individual kinematic variability (i.e., three groups of four cycles) and combinations with more than three IMUs were not explored, except for all sensors. Thus, additional investigation is needed to establish optimal sensor quantity and placement for automatic recognition of manual handling tasks that builds off existing work.

The objective of the current study was to examine the ability of sensor combinations to classify manual handling activities during a fatiguing protocol using neural network classifiers. To this end, kinematics of individuals performing sets of general manual handling activities were measured: standing, walking, and pulling, lifting, carrying, and pushing a crate. Twenty-four participants were recruited to perform these activities with minimal technique constraints to allow for kinematic variability. Furthermore, as fatigue can induce individual-specific changes in movement (V C H Chan et al., 2020), participants were asked to perform the manual handling activities until exhaustion. Eighteen sensor combinations involving consumer and research-grade devices were explored, involving one to six IMUs per combination. Fully connected and deep convolutional LSTM (DeepConvLSTM) neural networks were trained and evaluated using *k*-fold cross-validation to provide robust estimates of classification performance. The overarching aim of this work was to inform the sensor selection process in the future development of systems/frameworks requiring HAR for manual handling.

5.2 Method

5.2.1 Overview

This research complied with the American Psychological Association Code of Ethics and was approved by the Research Ethics Board at the University of Ottawa. Informed consent was obtained from each participant. This study consisted of data collection of simulated manual handling performed until exhaustion, data processing, hyperparameter optimization, classifier training, activity recognition, and performance evaluation. An overview is depicted in Fig. 5.1.

5.2.2 Participants

Twenty-four participants aged 16 years of age and older from the University of Ottawa area were recruited via convenience sampling. Participants were free of pain, injury, and MSDs in the previous 12 months; additionally, they were screened using the physical activity readiness questionnaire (PAR-Q+; Warburton et al., 2021). Upon arrival to the biomechanics laboratory, participants read the consent form (Appendix E), received a verbal explanation of the experimental protocol, then completed the PAR-Q+ and consent form.

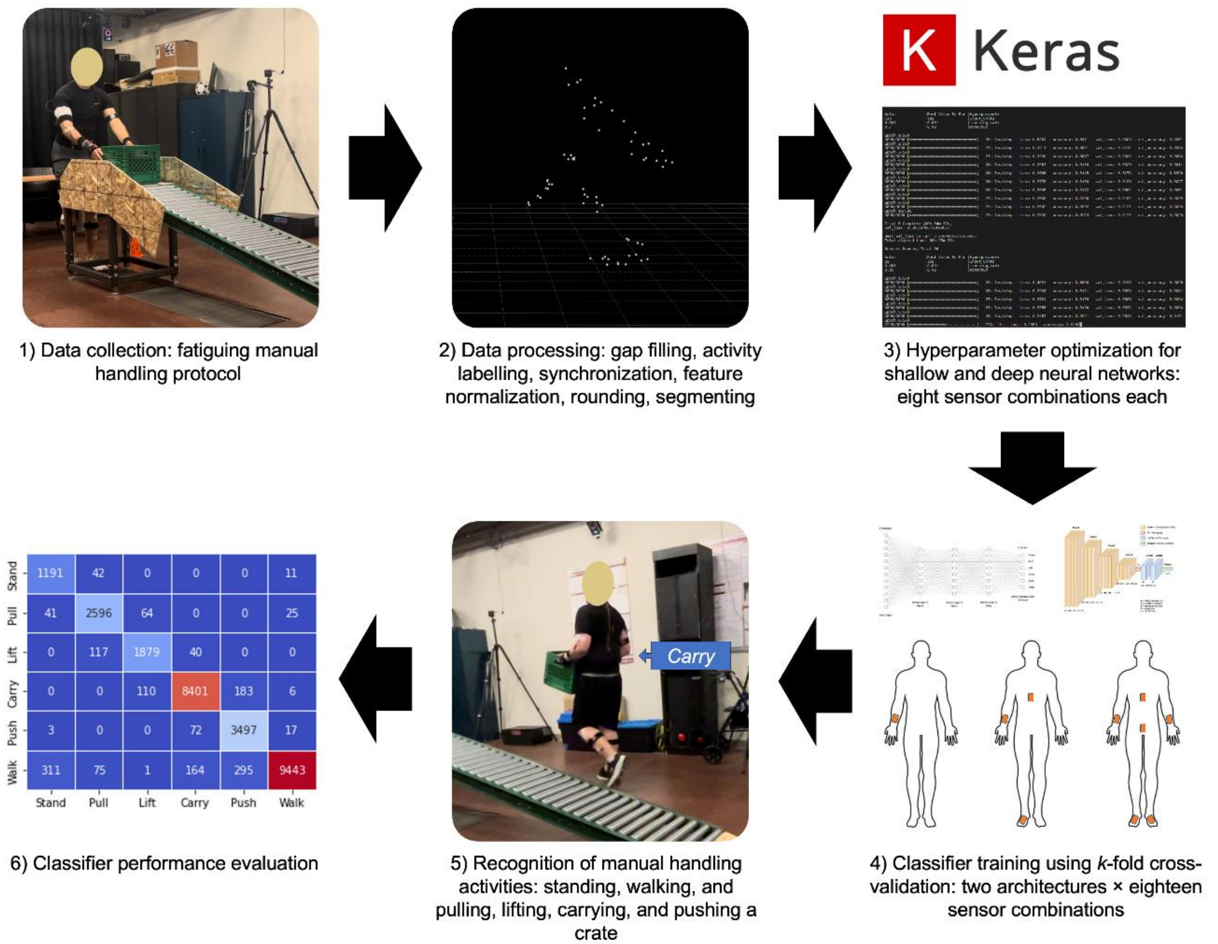


Fig. 5.1. Study overview.

5.2.3 Instrumentation

The data collection protocol has been described in subchapter 4.3; only the elements relevant to this study have been presented here. Participants changed into active clothing and footwear then donned the Xsens MVN Link inertial motion capture system (Xsens, Enschede, Netherlands) to capture whole body kinematics. This system included 17 IMUs secured/worn using neoprene bands/tight-fitting Xsens shirt: one on the head, sternum, pelvis, and bilaterally on the shoulders, upper arms, forearms, hands, thighs, shanks, and feet. To calibrate the system, participants stood in an upright posture, walked forward for 5 s, then turned around and returned to the starting position. Xsens data were collected using Xsens MVN Analyze 2022.02 software (Xsens, Enschede, Netherlands), sampled at 120 Hz. Participants also wore an Apple Watch Series 5 smartwatch (Apple Inc., Cupertino, USA) on their right wrist, where kinematic data from the IMU were sampled at 40 Hz using SensorLog 5.3.1 software (Bernd Thomas, Stuttgart, Germany). The IMUs from the Xsens and smartwatch contain a tri-axial accelerometer (linear acceleration) and gyroscope (angular velocity), while the Xsens IMUs also contain a magnetometer (magnetic field).

5.2.4 Manual Handling Protocol & Fatigue Assessments

After receiving a verbal explanation, participants received a two-minute opportunity to practice the protocol. The manual handling protocol was a repetitive circuit where participants were asked to pull, lift, carry, and push a crate (8.1k kg; 30 cm tall × 33 × 33 cm; handles for good coupling) down a roller conveyor, walk to the crate, then repeat (Fig. 5.2). One round was defined as one sequence of pull, lift, carry, push, and walk. Participants were instructed to handle the crate with both arms and to follow a pace of 15 s per round, enforced via metronome. Depending on their preferred movement speed, participants sometimes performed a "stand" activity at A) or C) to wait for the next round. Aside from these constraints, participants were encouraged to perform the protocol with self-selected technique. One set of the manual handling protocol was comprised of 20 rounds (5 min), with a fatigue assessment (approximately 2 min) before the first and between each set. To start each set, participants stood upright at position A) facing C), then concurrently performed a stomp and slapped their right hand on their right thigh to indicate the start of each trial for the Xsens and smartwatch IMUs.

The fatigue assessments consisted of a subjective fatigue rating using a 10 cm visual analog scale (VAS; Chiarotto et al., 2019) and a maximal lift strength assessment (Zemková et al., 2019). Participants were asked to mark a 10 cm VAS line on paper to indicate their perceived, holistic fatigue; the ends were anchored by the statements "No fatigue at all" and "The most fatigue I've ever experienced in my life, and I cannot continue". Participants never saw their previous fatigue VAS responses. Force generation capacity was assessed using a maximal lift strength assessment where three isometric deadlifts were attempted.

Participants were instructed to pull upwards on a static handle set to knee-height as hard as possible for 3 s, for three attempts with 5 s break between attempts. These data were measured using an s-type load cell sampled at 100 Hz using a custom program in LabVIEW software (National Instruments, Austin, TX, USA). Participants were instructed to maintain the same technique throughout the study to the best of their ability and were blinded to their results. After these two assessments, they immediately began another set of the manual handling protocol.

Participants performed as many sets as possible of the manual handling protocol until any of the following conditions were met: VAS became 9 cm or greater; maximal lift strength declined below 70% of their baseline value; pace was not maintained for three consecutive manual handling rounds; or they wished to stop. For this study, the fatigue assessments were only used to determine cessation.

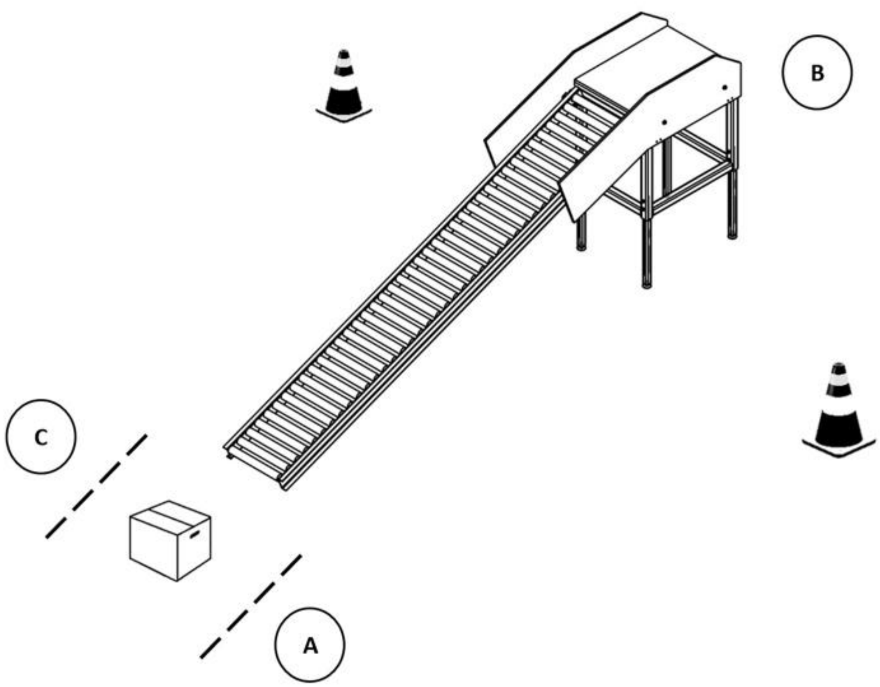


Fig. 5.2. The manual handling protocol. Participants stood behind the line at A) (approx. 0.5 m away from the crate), descended, and pulled the crate towards themselves. Then, they lifted the crate, turned, and carried it to B) by walking around the pylon (approx. 3 m). At B), they released the crate onto the low-friction table (0.75 m tall) and pushed it across and down the roller conveyor where it stopped between A) and C). Then, participants walked around the pylon to C), where they performed the same sequence on the other side of the roller conveyor. All activities from A) to C), or C) to A), was considered one round (i.e., one sequence of pulling, lifting, carrying, pushing, and walking). Participants sometimes stood at A) or C) to wait for the next round, depending on their movement speed.

5.2.5 Data Processing

Xsens IMU data were collected live on a Dell desktop and force recordings from the maximal lift strength assessments were collected on a Dell laptop, both running Windows 10 (Microsoft, Redmond, USA). Smartwatch IMU data were collected on-board. All data were processed using an Acer desktop (Acer, New Taipei City, Taiwan) running Windows 11 and an Apple M2 Mac Mini desktop running Sonoma 14.

The raw signals were exported from all Xsens IMUs (9 signals per IMU) and Apple Watch (6 signals). Raw IMU signals were used as previous work has shown no performance improvements when using sensor orientation and position information for HAR (Chapter 4; Chan & Graham, unpublished). Missing data from the smartwatch IMU were gap filled using data from the previous frame, implemented using the “ffill” method from the Pandas library (McKinney & others, 2010). Then, the primary researcher labelled all movements as being one of six manual handling activities (stand, pull, lift, carry, push, and walk) by visually observing simulated markers computed from Xsens IMU data in Mokka software (Biomechanical ToolKit, Capitale-Nationale, Canada). Major movements that were not manual handling activities were labelled as “other”, such as hitting a pylon and bending over to fix it. For the Xsens and smartwatch IMU data, the start of each set was defined as the moment the stomp and thigh slap were completed, respectively. Then, all activity labels from the Xsens IMUs were downsampled to 40 Hz and combined with the smartwatch IMU data. All data prior to the completion of the stomp and thigh slap were discarded.

One matrix each was created for Xsens and smartwatch IMU data by concatenating all sets from all participants together. The following processing steps were the same for both matrices. First, frames labelled “other” were discarded. Next, feature normalization was conducted separately for each feature by subtracting the mean and dividing by the standard deviation (SD) of all frames across within a subset (i.e., within training and validation folds). Then, all data were rounded to four decimal places. A sliding window approach was used to segment the data into “windows” containing an equal number of frames (selected using hyperparameter optimization; HPO), while the step size was kept constant at ¼ of the window size (Clouthier et al., 2020; Mavor et al., 2023). Each window was assigned a label based on the activity performed for the majority of that window. The resulting length and width of the data used for input into the machine learning algorithms was the window size and number of features, respectively.

5.2.6 Neural Networks

All machine learning was conducted using Python 3.10.12 (Van Rossum & Drake, 2009) and TensorFlow 2.12 (Abadi et al., 2016). Three Nvidia Titan RTX GPUs (Nvidia, Santa Clara, USA) with 24 GB of GDDR6 memory each were employed concurrently using a distributed mirrored strategy. These GPUs were installed on a Lambda Labs server (Lambda Labs, San Jose, USA) running Ubuntu 18.04 (Canonical, London, United Kingdom) with two Intel Xeon Gold 6248 CPUs @ 2.50 GHz (Intel, Santa Clara, USA) and 384 GB of DDR4 RAM.

Two existing neural network architectures were employed for manual handling activity recognition: fully connected neural network (FC; (Mavor et al., 2023), and deep convolutional long short-term memory neural network (DeepConvLSTM; Clouthier et al., 2020; Mavor et al., 2023). FC had an input layer, three fully connected dense layers with rectified linear unit (ReLU) activation functions, and a dense output layer with a softmax activation function (Fig. 5.3). For regularization, a drop-out operator with a probability of $p = 0.05$ was implemented on the output layer. FC model parameters were optimized during training using stochastic gradient descent (SGD) with the goal of minimizing loss, computed using the categorical cross-entropy function. DeepConvLSTM was comprised of an input layer, four convolutional layers with ReLU activation functions, two long short-term memory recurrent layers with hyperbolic tangent activation functions, and a dense output layer with a softmax activation function (Fig. 5.4). For regularization, drop-out operators were implemented on the recurrent layers with a probability of $p = 0.05$. DeepConvLSTM model parameters were optimized during training using batch gradient descent and the root mean squared propagation (RMSProp) update rule, where the objective was to minimize cross-entropy loss. For RMSProp, the decay factor was set to $\rho = 0.9$, and the momentum was set to 0.0. HPO was performed for select sensor combinations (seven procedures per model), while one FC and DeepConvLSTM model was developed for each sensor combination (36 models total).

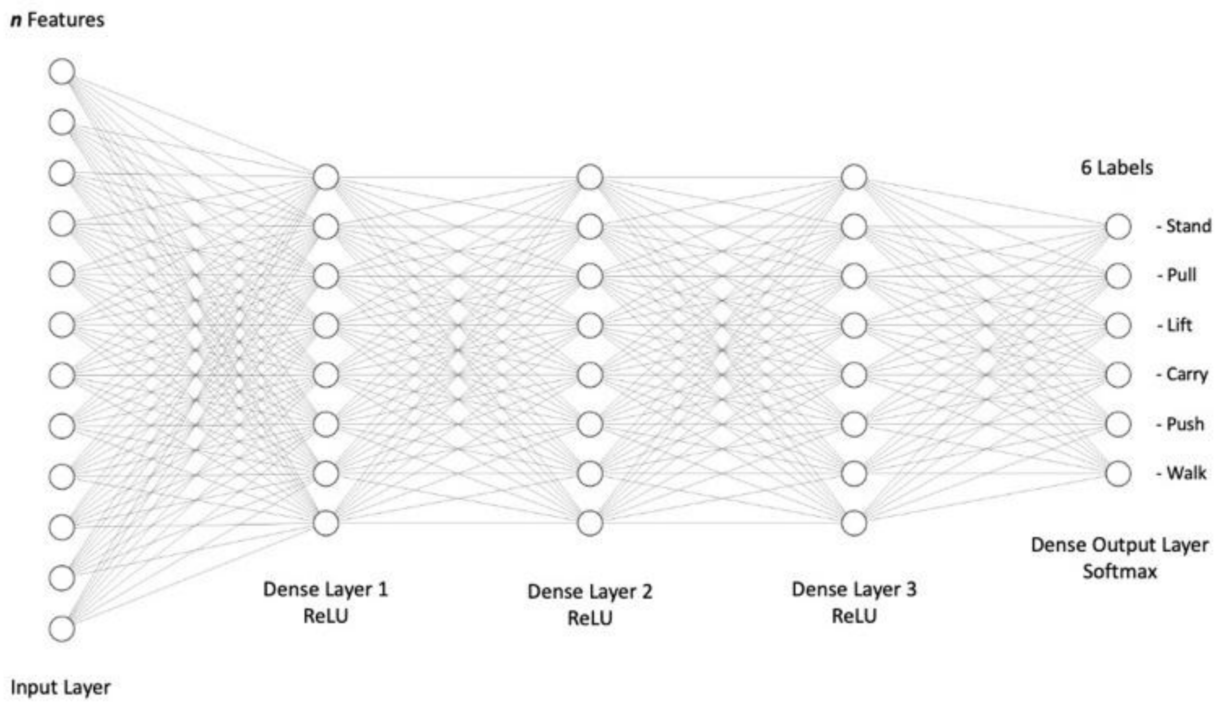


Fig. 5.3. The architecture of the fully connected neural network. The number of input features (*n*) was the number of sensors in the combination × nine signals (or six for the watch). ReLU = rectified linear unit.

5.2.7 Sensor Combinations

Data from eight Xsens IMUs and the Apple smartwatch IMU were included, configured to make up eighteen sensor combinations (Table 5.1). The selection of these combinations was informed by literature while considering practical implementation. Each sensor combination was a subset of the original matrix. HPO was employed for one combination per number of sensors, except for one-sensor combinations, where HPO was conducted separately for the Xsens T8 vertebrae and the smartwatch IMU. The number of features in each sensor combination subset was number of sensors × nine features, except Watch, comprised of six features. These were used as the input features for the FC and DeepConvLSTM neural networks.

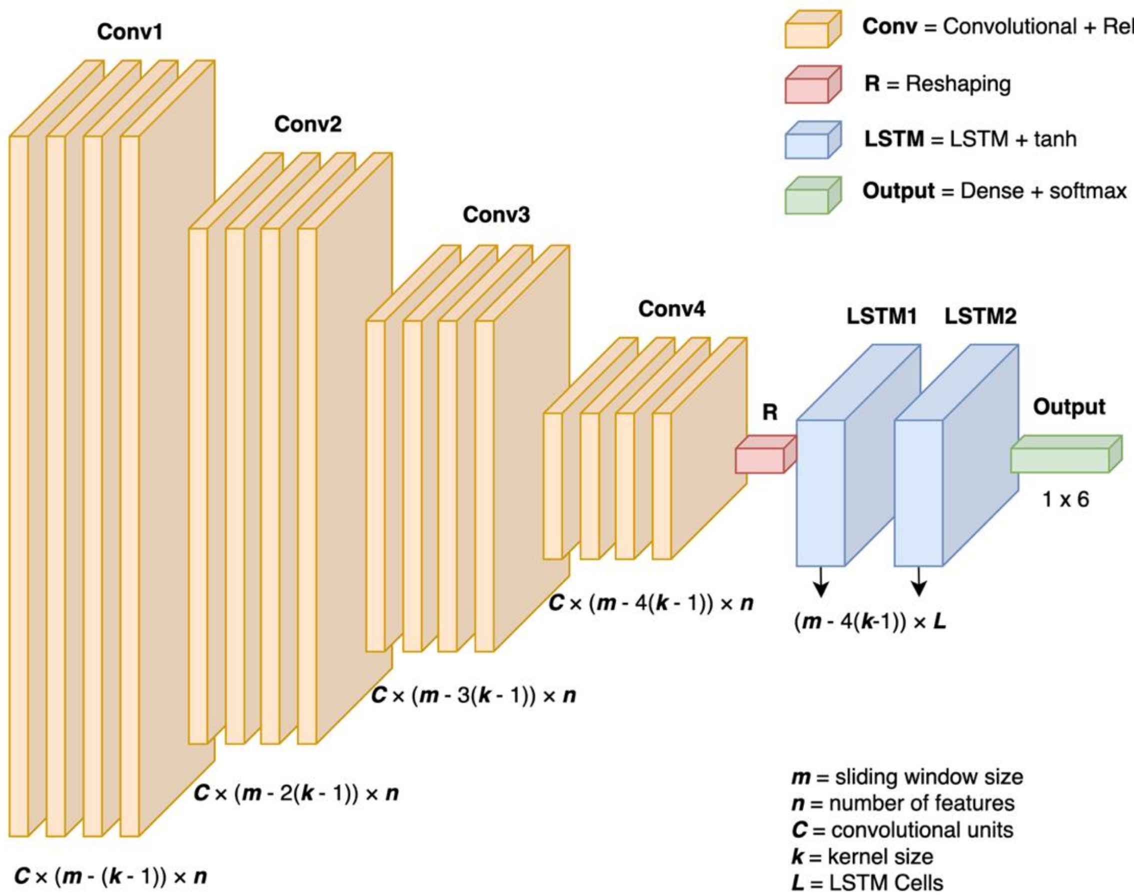


Fig. 5.4. Deep convolutional long short-term memory (LSTM) neural network architecture. The number of input features (*n*) was the number of sensors in the combination × nine signals (or six for the watch). ReLU = rectified linear unit.

5.2.8 Hyperparameter Optimization (HPO)

Bayesian HPO was implemented using the KerasTuner library (O'Malley et al., 2019) to search the hyperparameter search spaces for FC and DeepConvLSTM, shown in TablesTable **5.2** and Table **5.3**, respectively. These search spaces were guided by previous work employing similar tasks and instrumentation (Mavor et al., 2023). The sole difference in the search spaces between the Xsens and smartwatch IMU data was window size; due to different sampling frequencies, the values were selected to correspond to 0.5, 1, and 2 s of data from both devices. The objective was to minimize loss on the validation set, comprised of all data from three, randomly selected participants: 5, 12, and 15. For each set of hyperparameters, one trial was executed and was stopped early based on non-improving validation loss with a patience of three. Data were loaded with a batch size of 64 and HPO was limited to 24 trials per sensor combination. HPO was performed once per architecture, per sensor combination (16 total procedures). The selected hyperparameters were used for all combinations with that number of sensors (except Watch).

Table 5.1. The eighteen sensor combinations examined to perform activity recognition. All sensor presented are Xsens inertial measurement units (IMUs), except for Watch.

# of Sensors	Combination Code	Sensor(s) in Combination
1	**Watch**	Apple Watch on right wrist
	T8	T8 vertebrae
	Pel	Pelvis
	RFt	Right foot
	Head	Head
	RUA	Right upper arm
2	**2A**	T8 vertebrae, right foot
	2B	Right & left feet
	2C	Right foot & forearm
	2D	T8 vertebrae, right forearm
3	**3A**	T8 vertebrae, right & left feet
	3B	Right & left feet, right forearm
	3C	T8 vertebrae, right foot & forearm
	3D	T8 vertebrae, right & left forearms
4	**4A**	Right & left feet & forearms
	4B	T8 vertebrae, pelvis, right foot & forearm
5	**5**	T8 vertebrae, right & left feet & forearms
6	**6**	T8 vertebrae, pelvis, right & left feet & forearms

Bolded code indicates hyperparameter optimization was performed using this combination.

Table 5.2. The hyperparameter search space and for the fully connected artificial neural networks, and the hyperparameters selected for each sensor combination.

Sensor Combination	Window size	Input units	SGD learning rate	SGD momentum
Search space	60, 120, 240 (Watch: 20, 40, 80)	32, 64, 128, 192	0.0001, 0.001, 0.01, 0.1	0.7, 0.9, 0.95, 0.98
Watch	80	128	0.001	0.95
T8	240	32	0.01	0.9
2A	120	192	0.001	0.7
3A	120	192	0.0001	0.95
4A	120	32	0.01	0.9
5	120	32	0.001	0.7
6	120	32	0.001	0.9

SGD = stochastic gradient descent

Table 5.3. The hyperparameter search space and for deep convolutional long short-term memory neural network, and hyperparameters selected for each sensor combination.

Sensor Combination	Window size	Conv. Units	Kernel	LSTM cells	RMSProp learning rate
Search space	60, 120, 240 (Watch: 20, 40, 80)	16, 32, 48, 64	5, 7, 9, 11	16, 32, 48, 64	0.0001, 0.001, 0.01, 0.1
Watch	80	16	9	48	0.001
T8	120	16	5	48	0.0001
2A	240	48	11	64	0.0001
3A	240	64	9	32	0.0001
4A	240	16	11	64	0.001
5	240	64	11	48	0.001
6	120	32	5	16	0.0001

Conv. = convolutional; LSTM = long short-term memory; RMSProp = root mean squared propagation

5.2.9 *k*-fold Cross-Validation

k-fold cross-validation was employed to train and evaluate models for all sensor combinations. Due to the necessity to keep all data from a participant in the same fold (to avoid exposing participants to training and evaluation), the value of was *k* was set to eight, where each fold consisted of all data from three participants. Each fold was randomly selected (Table 5.4) then kept constant across all models and sensor combinations. A class weight dictionary was computed using the training set using the "balanced" setting in the SciKit-Learn Library (Pedregosa et al., 2011), then was employed during training. This was done separately for each fold to accommodate the varying degree of class imbalance.

Table 5.4. The participants in each fold for *k*-fold cross validation. These were randomly selected and kept consistent across all sensor combinations and neural networks.

Fold	Participant
1	1, 3, 13
2	6, 15, 21
3	4, 14, 20
4	18, 19, 24
5	9, 22, 23
6	2, 5, 7
7	11, 12, 16
8	8, 10, 17

Each *k*-fold cross-validation procedure was trained using a batch size of 64, maximum epochs of 500, and data were randomly shuffled prior to each epoch to prevent models from learning the activity order. Early stopping was implemented when validation loss ceased to improve with a patience set to three. Then, model weights from the epoch with the lowest validation loss were saved and used to make predictions on the test fold. Performance on each fold was evaluated using loss (categorical cross-entropy), accuracy (proportion of correct predictions to the total number of predictions), weighted average F1-score (harmonic mean of precision and recall, weighted based on instances), training time, and confusion matrices. Activity-specific precision (positive predictive value), recall (true positive rate), and F1-score were also computed.

5.3 Results

5.3.1 Participants

Participant demographics and the tally of reasons for stopping the manual handling protocol are presented in Tables Table **5.5** and Table **5.6**, respectively.

Table 5.5. Participant demographics. Values are presented as mean (standard deviation).

	Age (yr)	Height (cm)	Mass (kg)	Completed sets (no.)
Male (*n* = 13)	27.5 (7.2)	178.6 (6.5)	79.8 (11.0)	19.9 (12.0)
Female (*n* = 11)	25.6 (2.9)	168.9 (8.9)	68.2 (11.6)	16.5 (7.2)
Combined (*n* = 24)	26.7 (5.7)	174.2 (9.0)	74.5 (12.5)	18.4 (10.0)

Table 5.6. Tally of reasons for stopping the manual handling protocol.

Reason	Occurrences (no.)
MLS only	11
VAS only	6
MLS and VAS concurrently	3
Failure to maintain pace during manual handling	0
Voluntary	4

MLS = maximal lift strength; VAS = fatigue visual analog scale

5.3.2 Dataset

Across all participants, 441 sets of the manual handling protocol were completed, totaling 36.75 hours of data. This amounted to 15,852,125 frames of Xsens IMU data and 5,343,988 frames of smartwatch IMU data; once the frames labelled as "other" were removed, 15,849,914 and 5,343,245 frames remained in the Xsens and smartwatch datasets (0.0139% of data removed). In both datasets, 8,820 rounds (i.e., sequences of manual handling activities), or approximately 52,920 separate manual handling activities, were performed. The number of Xsens IMU frames labelled as each activity are shown in Fig. 5.5.

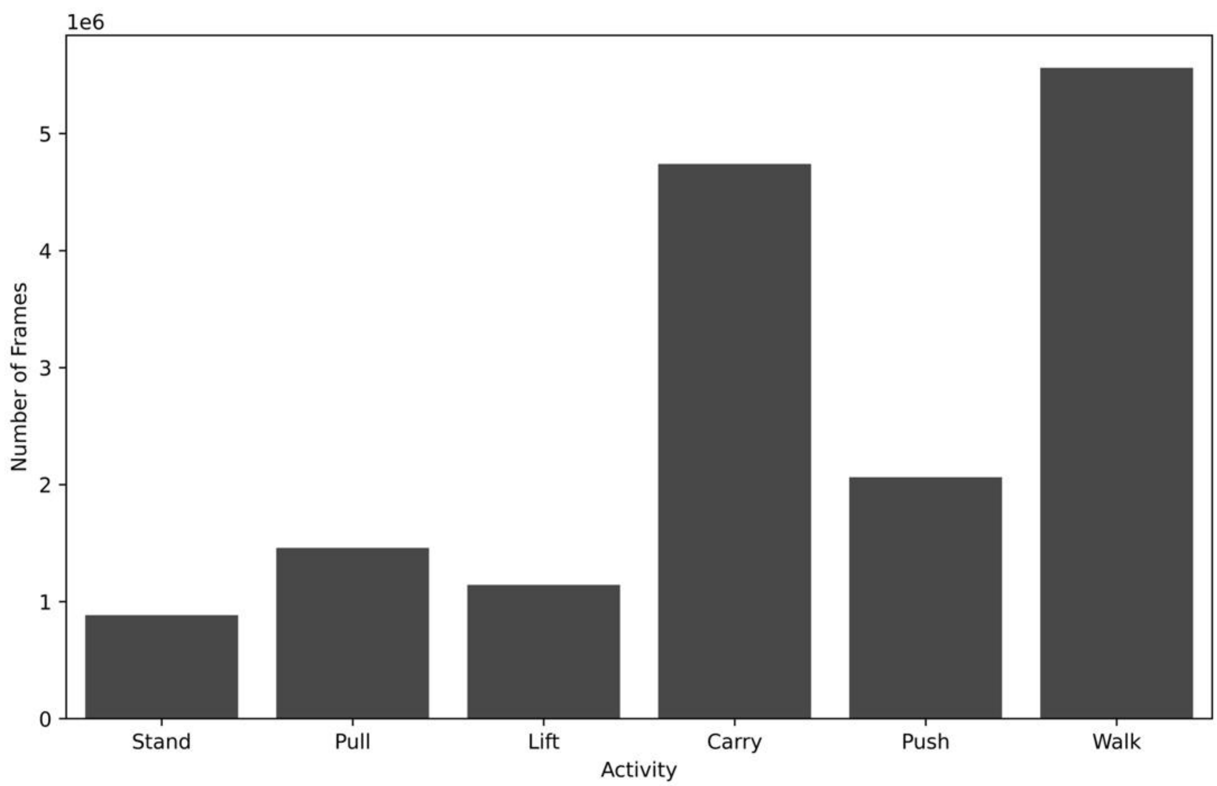


Fig. 5.5. The number of frames labelled as each activity across all participants and sets in the Xsens inertial measurement unit dataset, after removing activities labelled as "other" (0.0139% of all data).

5.3.3 Classification Performance

The loss, accuracy, weighted-average F1-score, and training time of the FC and DeepConvLSTM models are presented in Table 5.7, while the loss, accuracy, and weighted-average F1-scores of both models are depicted in Fig. 5.6. Sample confusion matrices are presented in Fig. 5.7. Activity-specific precision, recall, and F1-score averaged across all sensor combinations for each neural network are presented in Table 5.8. Full activity-specific results are included in the Supplementary Materials.

Table 5.7. The classification performance of the fully connected (FC) and deep convolutional long short-term memory (DeepConvLSTM) neural networks. Values are presented as mean (standard deviation). The sensor combinations were defined in Table 6.

Neural network	Sensor combination	Loss	Accuracy (%)	Weighted average F1-score (%)	Training time (s)
FC	Watch	0.2654 (0.0726)	91.61 (1.79)	91.75 (1.82)	389.1 (168.0)
	T8	0.2509 (0.9880)	92.48 (2.24)	92.55 (2.19)	159.6 (57.2)
	Pel	0.3092 (0.1710)	90.75 (5.28)	90.82 (5.33)	135.1 (59.6)
	RFt	0.2965 (0.0690)	90.25 (1.71)	90.41 (1.71)	128.8 (42.4)
	Head	0.2830 (0.0747)	91.59 (1.57)	91.73 (1.53)	174.4 (42.0)
	RUA	0.2677 (0.1157)	92.42 (2.06)	92.52 (2.01)	135.3 (50.8)
	2A	0.2339 (0.0987)	93.39 (1.86)	93.43 (1.85)	1551.5 (650.8)
	2B	0.2891 (0.1164)	91.12 (1.65)	91.25 (1.61)	1776.6 (380.3)
	2C	0.2266 (0.1053)	93.32 (1.95)	93.39 (1.90)	1552.5 (328.3)
	2D	0.2409 (0.1236)	93.20 (2.20)	93.26 (2.13)	1411.9 (730.4)
	3A	0.2166 (0.1142)	94.03 (1.66)	94.05 (1.66)	1984.9 (907.8)
	3B	0.2262 (0.1344)	93.42 (2.46)	93.49 (2.40)	2082.2 (1085)
	3C	0.2139 (0.1030)	93.77 (1.87)	93.82 (1.83)	1427 (762.7)
	3D	0.2231 (0.1089)	93.55 (1.82)	93.62 (1.77)	1520.7 (869.2)
	4A	0.1919 (0.1054)	94.18 (1.96)	94.22 (1.92)	625.6 (347.9)
	4B	0.1946 (0.0932)	94.26 (1.78)	94.30 (1.77)	654.5 (228.9)
	5	0.1976 (0.1086)	94.15 (1.93)	94.20 (1.90)	1314.2 (860.9)
	6	0.1927 (0.1089)	94.26 (1.99)	94.30 (1.97)	1783.8 (805.7)
DeepConvLSTM	Watch	0.2641 (0.0821)	91.53 (2.21)	91.64 (2.21)	620.0 (185.3)
	T8	0.2255 (0.0872)	93.17 (1.94)	93.23 (1.92)	2384.0 (624.9)
	Pel	0.2807 (0.1675)	91.39 (5.10)	91.41 (5.18)	2235.6 (828.7)
	RFt	0.3081 (0.0730)	89.98 (1.73)	90.11 (1.70)	2829.1 (1015.7)
	Head	0.2468 (0.0567)	92.24 (1.2)	92.26 (1.23)	2491.6 (609.7)
	RUA	0.2410 (0.0867)	92.43 (2.18)	92.52 (2.12)	2408.5 (630.6)
	2A	0.2108 (0.0794)	93.64 (1.78)	93.69 (1.74)	8429.1 (2876.6)
	2B	0.2702 (0.0788)	91.26 (1.72)	91.34 (1.75)	8749.1 (1730.3)
	2C	0.2138 (0.0961)	93.57 (2.23)	93.61 (2.24)	10257.6 (4632.3)
	2D	0.2099 (0.0775)	93.47 (1.92)	93.52 (1.92)	8832.0 (3893.9)
	3A	0.2027 (0.0700)	93.87 (1.84)	93.89 (1.83)	15291.5 (7774.5)
	3B	0.2162 (0.0825)	93.47 (1.90)	93.49 (1.88)	17030.9 (3133)
	3C	0.2058 (0.0785)	93.76 (1.95)	93.79 (1.96)	16950.9 (2244)
	3D	0.2271 (0.0935)	93.09 (2.08)	93.16 (2.06)	14950.8 (4225)
	4A	0.2348 (0.0877)	92.52 (2.01)	92.60 (1.98)	3417.9 (1156.1)
	4B	0.2544 (0.1414)	92.00 (4.39)	92.09 (4.27)	3242.1 (693)
	5	0.2370 (0.0993)	92.91 (2.35)	92.96 (2.30)	13462.8 (2331)
	6	0.2009 (0.0794)	94.09 (1.60)	94.11 (1.62)	12587.5 (3635.3)

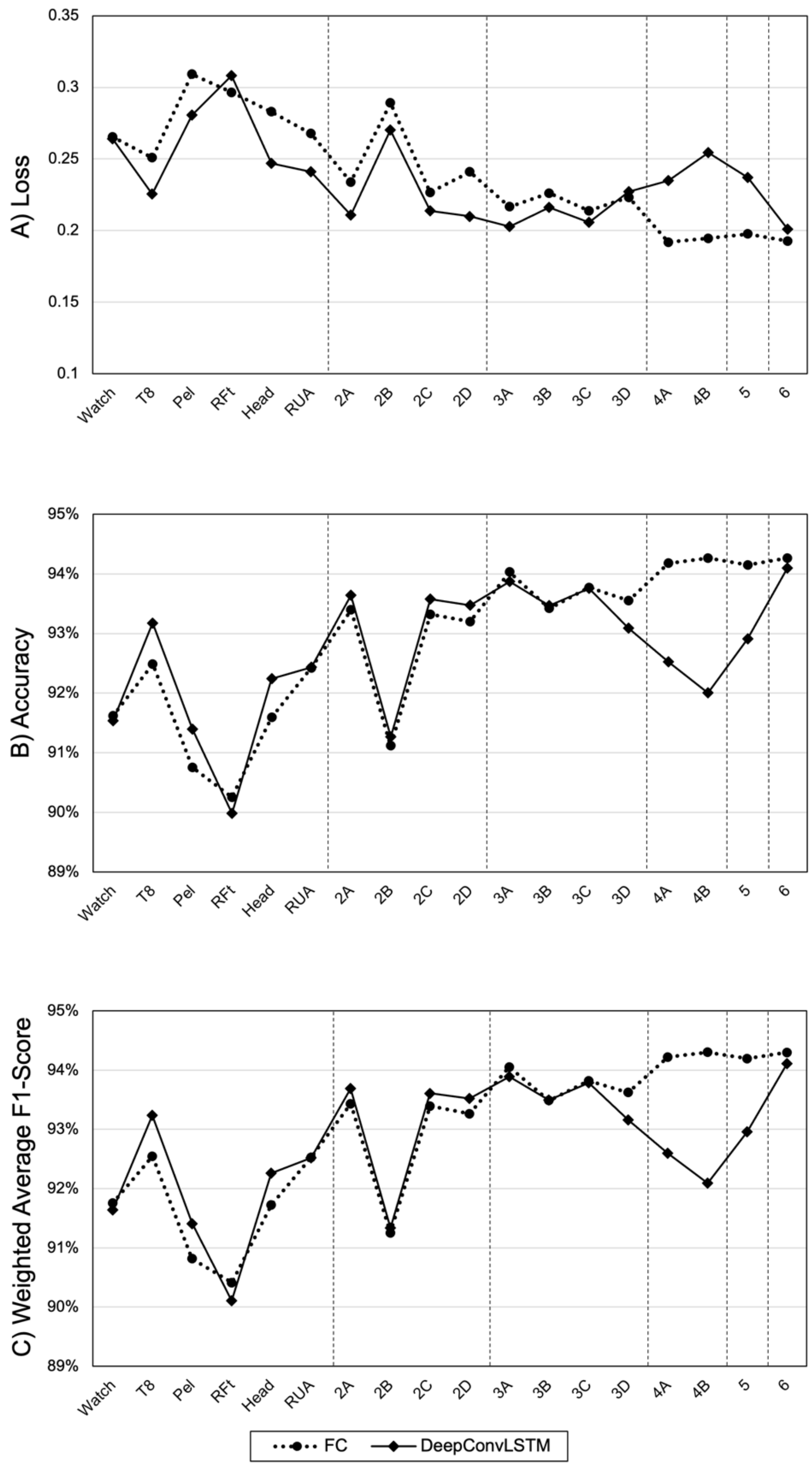


Fig. 5.6. The loss (A), accuracy (B), and weighted average F1-score (C) of the fully connected (FC) and deep convolutional long short-term memory (DeepConvLSTM) neural networks. The sensor combinations (x-axes) were defined in Table 5.1. The vertical dashed lines separate the number of sensors per combination from one (left) to six (right).

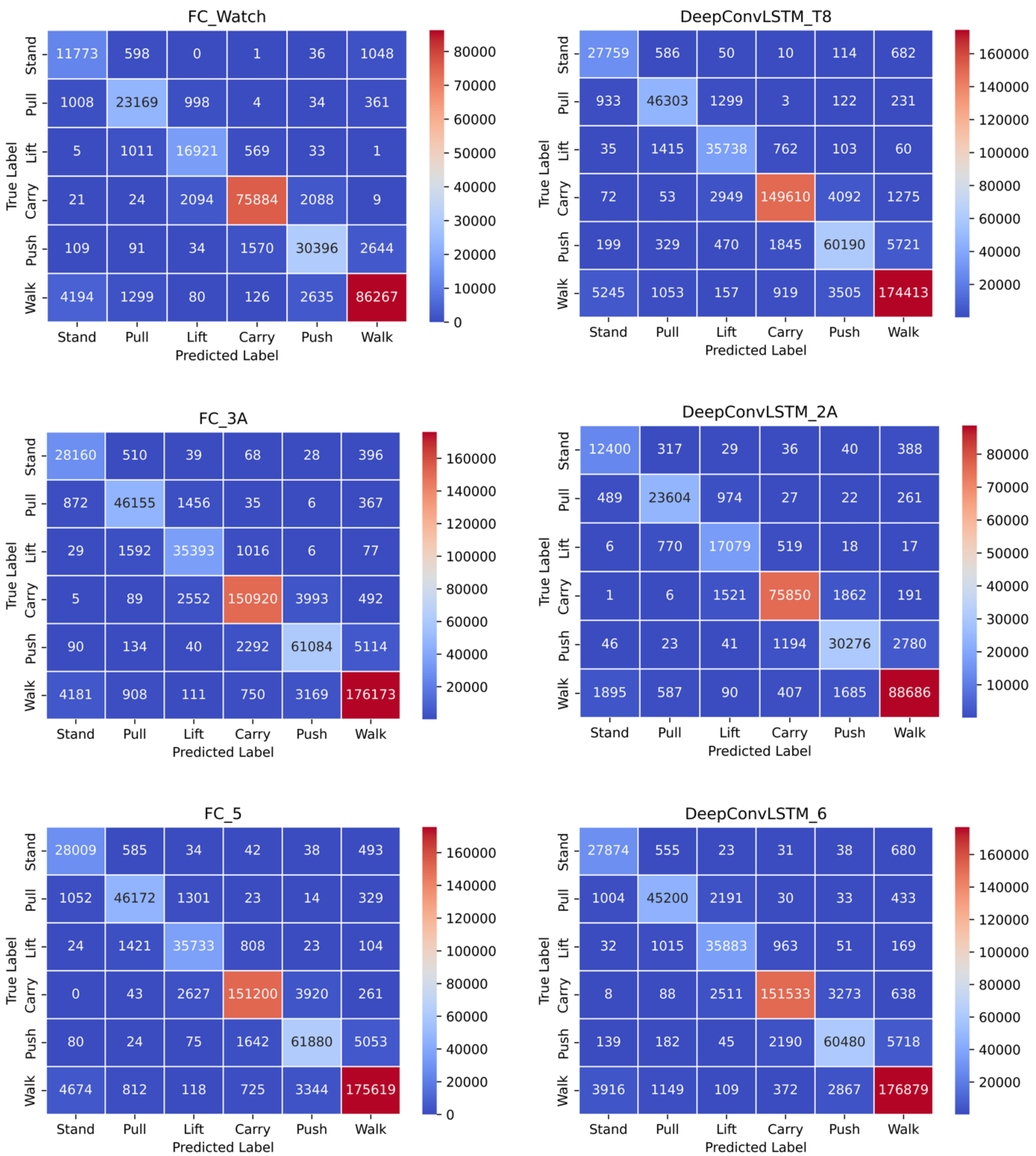


Fig. 5.7. Sample confusion matrices from fully connected (FC; left column) and deep convolution long short-term memory (DeepConvLSTM; right column) neural networks. Values were summed across all eight folds from cross-validation.

Table 5.8. Precision, recall, and F1-score for each activity when averaged across all sensor combinations for each model.

Neural Network	Metric	Stand	Pull	Lift	Carry	Push	Walk
FC	Precision (%)	79.44	90.67	86.53	97.12	86.93	96.26
	Recall (%)	92.81	91.44	92.44	94.34	89.89	93.44
	F1-Score (%)	85.36	90.97	89.30	95.68	88.02	94.79
DeepConvLSTM	Precision (%)	79.82	90.46	84.76	97.17	87.98	95.46
	Recall (%)	91.59	90.18	92.04	94.23	88.54	94.08
	F1-Score (%)	85.02	90.20	88.11	95.64	87.95	94.74

5.4 Discussion

The risk for developing MSDs or sustaining an accident while performing fatiguing manual handling can be mitigated using novel interventions that often require activity recognition. The purpose of the current study was to analyze the quantity and placement of IMUs for recognizing manual handling activities. A general trend was observed where classification performance improved as the number of sensors increased. However, all classification models showed strong performance when trained and evaluated using k-fold cross-validation (e.g., weighted average F1-scores ≥ 90.11%), regardless of the number of sensors in the combination, neural network architecture used, or IMU quality (i.e., commercial or research-grade). Observed trends and recommendations for use are discussed.

Despite the kinematic variability introduced from large sample size and the presence of fatigue, strong classification performances were observed in the current study across all permutations of sensor combinations and models. This was likely due (in part) to the large amount of data available, exceeding what has been collected in similar work (K. Kim & Cho, 2020; S. Kim & Nussbaum, 2014; Porta et al., 2021). Furthermore, the kinematic variability present in this study may strengthened the generalizability of the presented classifiers (Barshan & Yurtman, 2016). Another potential reason for the observed performance was the number of activities (six), which was less than eight in Porta et al. (2021) and thirteen in K. Kim & Cho (2020), while more than two to five activities per category in Akhavian & Behzadan (2016). While fewer activities can improve classification performance, it is noteworthy that some activities in the present study have been divided and/or differentiated into two or more activities in other work. For example, lifting is differentiated by the origin of the crate (i.e., ground or knuckle-height) in Porta et al. (2021), while carrying is differentiated by method (i.e., one-shoulder or two-handed) in Lee et al. (2020). The decision to differentiate between types of activities should depend on the information/resolution needed for the final objective. Overall, the classification performances presented are suitable for implementation, especially as the lack of movement constraints in this study ensures that these models can be generalizable to a variety of manual handling techniques.

While no activities were classified poorly, certain activities were classified better than others. Using F1-score averaged across all sensor combinations, the activities classified from best to worst were carry, walk, pull, lift, push, stand. This finding was true for both models and did not rank according to the number of samples per activity. This suggests that other variables caused some activities to be easier to recognize than others. It is speculated that stand was the poorest classified activity due to the fewest samples, being characterized by stationary lower extremities which are present in other activities (e.g., push), whilst upper extremity movements were inconsistent (i.e., participants may or may not have checked their watch while standing). If an under-represented or difficult activity is particularly important for an application, methods at the data (e.g., re-sampling) and algorithmic (e.g., class weighting) level can be used to improve performance (K. Chen et al., 2021).

Several sensor combinations were of note in this study, namely: Watch, T8, 3A, and 6. Using only raw accelerometer and gyroscope data from the smartwatch sampled at 40 Hz led to classification accuracy $\geq$ 91.53%, which was within 3% of combination 6, the best in the study. When using one Xsens IMU, the placement was best at the level of the T8 vertebrae with accuracy $\geq$ 92.48%. These locations were likely effective because they provided relevant information on either hand movement or centre of gravity movement, both relevant to manual handling (Ryu et al., 2019). Three IMUs may be the optimal number when at least one sensor is placed on the upper and lower body (e.g., 3A), achieving similar performance to six IMUs in this study or eight IMUs in related work (Yan et al., 2021). Overall, these findings demonstrate that activity recognition can be accessible and implemented with or without research-grade devices.

This study showed a general trend that more IMUs lead to stronger classification performance, a finding that has been observed in other HAR research (Clouthier et al., 2020). This is because more IMUs provide more data for training and describe the activity of interest with greater detail. However, this trend did not hold up for certain combinations; namely, RFt and 2B did not show better performance than other one-sensor combinations. Combinations RFt and 2B only consisted of foot data, suggesting that these configurations may not provide sufficient information for effective manual handling activities recognition on their own. While solely using IMU-instrumented shoe insoles have been suitable for gait activities (D'arco et al., 2022), data including the upper extremities and/or centre of gravity are warranted for manual handling.

Another deviation from the trend of increasing sensors equating to improved performance was the deep neural network (DeepConvLSTM) for combinations 3D, 4A, 4B and 5. Benefits of using deeper, complex neural networks with more layers and parameters is the ability to model highly non-linear relationships (Ramamurthy & Roy, 2018). However, for this certain set of combinations, using a complex model may have led to overtraining; high variance (i.e., noise) in the training data may have been learned, leading to poorer generalization on unseen data. Another possibility is that the predefined hyperparameter search space was suitable for select sensor combinations but not others, a problem that may not have existed when training the fully connected neural network as fewer hyperparameters and layers were optimized. An additional consideration for implementing deeper neural networks is the longer training time involved; unless computational resources are not a limitation, shallower neural networks may be more accessible with merely minor detriments in performance when using one or two IMUs.

This work was presented with several delimitations. First, the HPO procedure was not exhaustive. The search spaces were constrained, and the search strategy was limited to one attempt per permutation and 24 permutations per sensor combination and model pairing. Thus, it is possible that other hyperparameters may have led to improved performance. Furthermore, HPO was only conducted for seven of eighteen sensor combinations (per model), likely inflating the performance metrics on combinations that HPO was performed on. The decisions to limit the thoroughness of the HPO procedures ensured that the study was feasible to conduct, whilst ensuring the hyperparameters were still systematically selected using heuristics. Furthermore, this study included six activities and other manual handling activities were not included. However, the activities included were performed with high variability between and within individuals while encompassing movements that are common across many sectors of work (Dick et al., 2016); thus, the results can still be informative.

5.5 Conclusion

Workers performing fatiguing manual handling work are at risk for sustaining MSDs and accidents. Wearable sensors used to recognize manual handling activities can be part of preventative interventions by enabling automatic ergonomic assessments, risk estimation, or fatigue modelling. As the quantity and placement of IMUs can influence the quality of activity recognition and the cost of implementation, this study examined two models' performance when classifying six manual handling activities using eighteen different sensor combinations. Using over 36 hours of fatiguing manual handling data to train a deep neural network, one-sensor combinations could be used to identify activity with up to 93.17% accuracy. It was found that incorporating more IMUs tended to improve performance, and commercial-grade IMUs only led to a small detriment in performance compared to research-grade IMUs.

5.6 Key Points

- The quantity and placement of inertial measurement units (IMUs) can impact the cost and performance of manual handling activity recognition.
- While wearing IMUs, 24 participants performed a simulated manual handling protocol until exhaustion.
- Artificial neural networks were trained on 18 combinations of IMUs to recognize 1 of 6 activities being performed.
- More IMUs tended to improve performance, while a smartwatch or an IMU on the T8 vertebrae led to strong classification accuracies ≥ 91.53%.

Chapter 6. Study 4: Continuous Fatigue Estimation During Manual Handling Tasks

This work has yet to be submitted for publication.

Abstract

Fatigue at work increases WMSD and occupational accident risk, but this risk could be mitigated if fatigue could be monitored and prevented. Because of the multi-faceted nature of fatigue progression, an unobtrusive method to monitor holistic fatigue at work is warranted. An approach was presented in this work to adapt to various manual handling tasks using previously-developed human activity recognition models, then estimate holistic fatigue status using kinematics from an inertial measurement unit (IMU) with or without supplementary heart rate (HR) data. A simulated manual handing protocol comprised of six tasks was performed by 24 participants until excessive fatigue was detected or volitional failure occurred. Participants performed the manual handling protocol for an average of 92 min, during which six fatigue metrics were measured then combined in a fatigue composite index (FCI). Artificial neural networks were trained to estimate the FCI for each manual handling task, using input data from one of four sensor combinations. A single IMU on the T8 with HR data could be used to estimate the FCI with mean absolute error (MAE) of 0.05 and Pearson's correlation of r = 0.59 across all tasks and participants, with peak performances of MAE = 0.023 and r = 0.94 on select combinations. The results of this work represent a component of a system that can estimate holistic fatigue using accessible, unobtrusive devices during the performance of various manual handling tasks, which can be employed to mitigate injury and accident risk.

6.1 Introduction

Fatigue is extremely pervasive at work (Williamson & Friswell, 2013). Not only can perceived fatigability influence reaction time, awareness, and decision making (Jeklin et al., 2021), but performance fatigability can alter kinematics, motor control, and force generation capacity in a manner that affects workers' affordances in movement solutions (Côté et al., 2005; Enoka & Duchateau, 2016; Gribble & Hertel, 2004; Place et al., 2010). Together, their interaction can increase the risk of developing WMSDs and occupational accidents, which have resulted in and continue to result in significant lost work time and costs worldwide

(M. F. Antwi-Afari et al., 2017; Kang et al., 2014; Namian et al., 2021; Raenović, 2023). Thus, it is desirable for workers and employers alike to mitigate risks associated with fatigue at work.

A useful method to inform the provision of breaks, job rotation, and feedback is to monitor fatigue status. However, common, gold-standard methods are cumbersome and difficult to implement at work, such as assessments of maximal force generation capacity or endurance (V C H Chan et al., 2020; Mehta & Agnew, 2012), self-reports or surveys of fatigue status (Abdel-Malek et al., 2021; Saito, 1999), electroencephalography (EEG; Saito, 1999), or electromyography (EMG; Shair et al., 2017). However, because fatigue alters movement behaviours, kinematics have been used successfully to monitor changes in fatigue status, which can be measured unobtrusively using inertial measurement units (IMUs; Chan et al., 2020). Combined with heart rate (HR) data measured using electrocardiogram (EKG) or a photoplethysmography (PPG) sensor, an inexpensive setup can be implemented for fatigue modelling (Maman et al., 2017).

Modelling fatigue using kinematic data is relatively novel and has primarily relied on the use of machine learning models (ML) to classify fatigue states, often "fatigued" or "not fatigued' (Baghdadi et al., 2018; Escobar-Linero et al., 2022; Karvekar et al., 2021; Marotta et al., 2021). However, this simplification limits the preventative capabilities of a fatigue modelling system, and does not accurately represent the phenomenon, which is a continuous process instead of discrete states (Enoka & Duchateau, 2016). Thus, it is desirable to estimate fatigue as a continuous variable, done by only a few studies. Karg et al. (2014) estimated self-reported fatigue during repetitive squats, using marker-based motion capture data, while Maman et al. (2017) detected fatigue then estimated self-reported fatigue during three different occupational tasks using four IMUs and an HR monitor. Gholami et al. (2020) estimated self-reported fatigue during treadmill running using eight textile strain sensors in their garments to measure kinematics. Lastly, Lee et al. (2024) used an HR monitor to compute HR reserve and used a critical power model to estimate perceived whole-body fatigue in construction workers. While these pioneering studies have presented various novel approaches to model fatigue as a continuous variable, several limitations still exist.

First, kinematic-based fatigue models are highly task-specific, which is why many proposed methods focus on one, repetitive task at a time. While this focus may be necessary, methods have not been designed to automatically adapt the modelling approach to the changes in task. Second, when fatigue is estimated as a continuous variable, the variable is typically a self-reported rating of fatigue. While useful, they may be inaccurate without a calibration procedure and only represent one component of fatiguability (Abdel-Malek et al., 2021; Christodoulou, 2017; Enoka & Duchateau, 2016). Lastly, more training data may be needed to improve fatigue estimation performance and reduce the quantity of input data required. The aim of this work was progress towards the development of a method to improve on these shortcomings for manual handling fatigue estimation.

The overarching goal of this work was to develop a fatigue estimation system driven by kinematics (with or without HR) that is capable of adapting to the manual handling task being performed, then estimating fatigue as a continuous variable in near real-time. The first objective of the present study was to develop a fatigue composite index (FCI) to represent a holistic ground truth for fatigue. To achieve this, six variables previously associated with fatigue were quantified during a simulated manual handling protocol to comprise the FCI: self-reported fatigue (fatigue visual analog scale; VAS), instantaneous HR (HRinst), HR variability (low frequency high frequency ratio; LFHFratio), and movement jerk (pelvic jerk) which increase with fatigue (Tran et al., 2009; Ueda et al., 2006; L. Zhang et al., 2019), in addition to force generation capacity (MLS) and muscle activity power spectrum (EMG median frequency; MDF) which decrease with fatigue (Cifrek et al., 2009; Nocella et al., 2011). Then, task-specific machine learning (ML) models were trained and evaluated using four different sensor combinations to estimate the FCI continuously. Integrated in a cascading ensemble with previously-developed human activity recognition (HAR) models (Chapters 4 and 5), this system would be capable of estimating continuous, holistic fatigue across various manual handling tasks in near real-time using minimal instrumentation.

6.2 Methods

6.2.1 Participants

Twenty-four (13 male and 11 female) participants from the University of Ottawa downtown and Lees campuses were recruited for this study via convenience sampling. Participant characteristics are presented in Table 6.1. All participants were able to read and communicate in English or French, ensuring that they thoroughly understood the study information, forms and questionnaires, and instructions from the researchers. To screen their readiness for physical activity, all participants completed the physical activity readiness questionnaire (2021 PAR-Q+; Appendix D; Warburton et al., 2021). To be included in this study, they had to answer "NO" to all questions on page 1, indicating that they were medically-cleared for physical activity. If they answered "YES" to any question on page 1, they must have answered "NO" to all questions on pages 2-3 to participate; otherwise, participants were excluded. To reduce the risk for injury resulting from overexertion and/or accidents, participants must not have had any injuries or MSDs in the previous 12 months prior to the date of their session. To ensure that they were not over-estimating their current physical abilities, the researchers explained that repetitive manual tasks were required during the testing session, performed until they experienced very high levels of fatigue. Lastly, participants were required to be 16 years of age or older so that they could provide informed consent by themselves.

Table 6.1. Participant demographics. Values are presented as mean (standard deviation).

	Age (yr)	Height (cm)	Mass (kg)	Completed sets (no.)
Male (*n* = 13)	27.5 (7.2)	178.6 (6.5)	79.8 (11.0)	19.9 (12.0)
Female (*n* = 11)	25.6 (2.9)	168.9 (8.9)	68.2 (11.6)	16.5 (7.2)
Combined (*n* = 24)	26.7 (5.7)	174.2 (9.0)	74.5 (12.5)	18.4 (10.0)

6.2.2 Instrumentation

The participants began by changing into athletic clothing and footwear. Then, participants were prepared to don seven EMG electrodes on their body, which included shaving to improve the electrode-skin interface (when necessary) and sanitizing the attachment sites with alcohol wipes. The Delsys Trigno Avanti (Delsys, Natick, USA) EMG sensors were secured to right side of the body (unilaterally) on the following locations using medical-grade adhesive: vastus medialis obliquus (VMO), vastus lateralis (VL), lumbar erector spinae (LES), biceps brachii (BB), triceps brachii (TB), anterior deltoid (AD), and posterior deltoid (PD). A Delsys electrocardiogram (EKG) Biofeedback sensor (Delsys, Natick, USA) was worn on the participants chest on the left side. The sensor setup is depicted in Fig. 6.1. These sensors were further securing using medical tape and, in some instances, elastic strapping, to minimize the risk of sensors falling off during the manual handling protocol. Nexus software (Vicon Motion Systems Ltd., Oxford, UK) was used to collect all EMG and EKG data, sampled at 2000 Hz.

Then, an Xsens MVN Link inertial motion capture system (Xsens, Enschede, Netherlands) was worn to capture whole body kinematics. This system consisted of 17 IMUs secured/worn by the participants using sanitized neoprene bands and a tight-fitting Xsens shirt: one on the head, sternum, and pelvis; and bilaterally on the shoulders, upper arms, forearms, hands, thighs, shanks, and feet (Fig. 6.2). To calibrate the IMU system, participants were asked to stand in a neutral posture, walk for 5 s forward, turn around, and return to the starting location and position. The kinematic data was collected using Xsens MVN Analyze 2022.0.2 software (Xsens, Enschede, Netherlands), sampled at 120 Hz. Participants also wore an Apple Watch Series 5 (Apple Inc., Cupertino, USA) on their right wrist, where kinematic data were sampled at 40 Hz using SensorLog 5.3.1 software (Bernd Thomas, Stuttgart, Germany).

Fig. 6.1. The anterior (left) and posterior (right) perspectives of the electromyography and electrocardiogram (EKG) sensor setup. AD = anterior deltoid; BB = biceps brachii; VL = vastus lateralis; VMO = vastus medialis obliquus; PD = posterior deltoid; TB = triceps brachii; LES = lumbar erector spinae. Note that there was only one VL sensor despite being depicted in both perspectives.

6.2.3 Manual Handling Protocol & Fatigue Assessments

The manual handling protocol and fatigue assessments were previously described in Chapter 4.3.2; the only difference being that the fatigue assessments were used as ground truth fatigue measures (fatigue VAS and MLS) in addition to determining study cessation.

6.2.4 Data Processing

All Xsens IMU and Delsys EMG and EKG data were collected on a Dell desktop (Dell Inc., Round Rock, USA) running Windows 10 (Microsoft, Redmond, USA). Raw Apple watch IMU data were collected onboard then exported to an iPhone 13 Mini smartphone (Apple Inc., Cupertino, USA), prior to being transferred to the desktop. Maximal lift force was collected on a Dell laptop running Windows 10. Data processing was conducted using an Apple M2 Mac Mini desktop computer (Apple Inc., Cupertino, USA) running Sonoma 14.0. Python 3.11 (Van Rossum & Drake, 2009) was used for all data processing, except for the EMG data, which were processed using MATLAB R2020a (MathWorks, Natick, USA).

Fig. 6.2. The Xsens inertial motion capture suit with 17 inertial measurement units: one each on the head, sternum, and pelvis; and bilaterally on the shoulders, upper arms, forearms, hands, thighs, shanks, and feet.

6.2.4.1 Kinematics

The raw IMU signals and sensor orientation and position signals obtained using sensor fusion (Roetenberg et al., 2013; Schepers et al., 2018) were exported from all Xsens IMUs (3 linear acceleration, 3 angular velocity, 3 magnetometer, 4 orientation, and 3 position signals per IMU) and the Apple Watch (3 linear acceleration, 3 angular velocity, and 4 orientation signals). Missing data from the smartwatch IMU were gap filled using data from the previous frame, implemented using the “ffill” method from the Pandas library (McKinney & others, 2010). Then, the primary researcher labelled all movements as being one of six manual handling activities (stand, pull, lift, carry, push, and walk) by visually observing simulated markers computed from Xsens IMU data in Mokka software (Biomechanical ToolKit, Capitale-Nationale, Canada). Major movements that were not manual handling activities were labelled as “other”, such as hitting a pylon and bending over to fix it. For the Xsens and smartwatch IMU data, the start of each set was defined as the moment the stomp and thigh slap were completed, respectively. Then, all activity labels determined using the Xsens IMUs were downsampled to 40 Hz and combined with the smartwatch IMU data. This resulted in an Xsens dataset (120 Hz) and a smartwatch dataset (40 Hz) All data prior to the completion of the stomp and thigh slap were discarded. These kinematic data were used as inputs in the fatigue estimation model.

A finite differences method was used compute jerk of the pelvis Xsens IMU (Hamill et al., 2014). First, a Butterworth low-pass filter with a cutoff frequency of 5 Hz was applied to the raw linear acceleration data from the pelvis. The filtered acceleration data were then used to compute jerk by differentiating the acceleration signals in the X, Y, and Z axes at each timestep. Forward and backward differencing were performed if the frame was at the start or end of a set, respectively; all other frames were differentiated using a three-point central difference method (Hamill et al., 2014). Within each axis, the difference between consecutive acceleration values was divided by the time step(s) (1/120 seconds). Next, the magnitude of pelvic jerk was computed as the resultant of its X, Y, and Z components using Equation 1:

$$J_{Pel,Mag} = \sqrt{{J_{Pel,X}}^2 + {J_{Pel,Y}}^2 + {J_{Pel,Z}}^2} \tag{1}$$

where J_{Pel} is pelvic jerk and *Mag* is the magnitude (i.e., resultant).

To compute the mean and standard deviation of the pelvic jerk across all participants, the pelvic jerk signals across all sets for each participant were rubber banded to 0 – 100% to represent percentage of protocol completion. Then, the mean and standard deviation were computed at each percentage of completion. To produce a pelvic jerk signal that aligned with the Apple Watch IMU data, down-sampling was conducted to produce a signal with length equal to the number of frames for each participant in the Apple Watch data. The down-sampling was conducted using the Piecewise Cubic Hermite Interpolating

Polynomial (PCHIP) interpolator (Fritsch & Butland, 1984). The corresponding pelvic jerk signal was concatenated to the Xsens and Apple Watch matrices to be used as a ground truth for fatigue in the FCI.

6.2.4.2 Electromyography (EMG)

The seven EMG signals were analyzed during the MLS assessments only. The direct current bias of all EMG signals were removed by taking the average of the signal and subtracting it from the signal (Cudlip et al., 2015). Then, the EMG signals was smoothed using a fourth-order Butterworth band-pass filter with cut-off frequencies of 20 and 450 Hz. EMG MDF was computed for each MLS assessment using methods outlined by Karthick et al. (2018). First, the EMG bursts to analyze were identified as the largest moving average during a 1.5 s window of the linear enveloped EMG signal (Fig. 6.3). Next, these bursts were isolated, then a fast Fourier transform was applied. The resulting power spectrum was divided into equal halves to determine the MDF (Fig. 6.4).

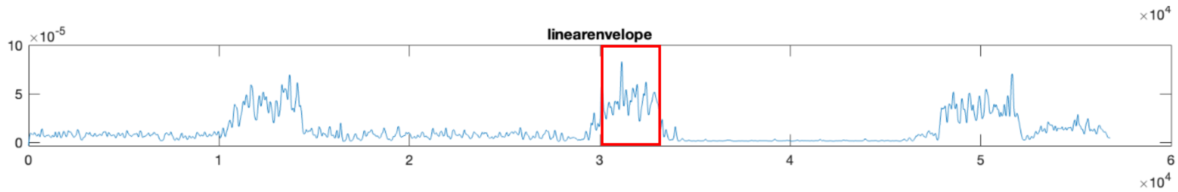


Fig. 6.3. A depiction of a 1.5 s window being identified as the largest moving average during a maximal lift strength assessment in the linear enveloped lumbar erector spinae electromyography (EMG) signal. The EMG amplitude is on the y-axis, the frame number is on the x-axis, and the red box shows the 1.5 s window.

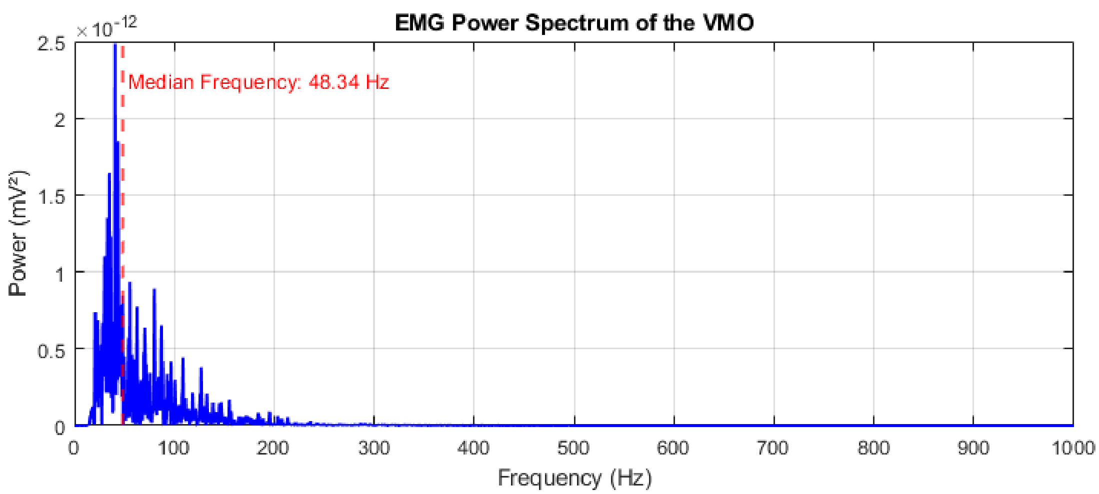


Fig. 6.4. Electromyography (EMG) power spectrum of the vastus medialis obliquus (VMO). The red dashed line at 48.34 Hz divides the power of the signal into two equal halves, which represents the median frequency for this sample.

To compute the mean and standard deviation of all EMG MDFs (seven EMG sensors) across all participants, MDFs across all assessments for each participant were rubber banded to 0 – 100% to represent percentage of protocol completion. Then, the mean and standard deviation were computed at each percentage of completion. To produce EMG MDF signals that aligned with the Xsens and Apple Watch IMU data, up-sampling was conducted to produce EMG MDF signals with length equal to the number of frames for each participant in the Apple Watch and Xsens IMU data. The rubber banding and up-sampling were conducted using the PCHIP interpolator (Fritsch & Butland, 1984). Finally, the seven EMG MDF signals were averaged to provide one signal to represent EMG MDF, weighted using Equation 2:

$$EMG\ MDF\ for\ FCI = \frac{(VMO + VL + BB + TB + AD + PD)}{8} + \frac{LES}{4} \tag{2}$$

where EMG MDF for FCI was the averaged signal for use in the FCI. This equation effectively averaged all signals together, while weighting the LES twice as much as all the other signals. This was done to have equal representation of four areas of the body: thigh (VMO and VL), upper arm (BB and TB), shoulder (AD and PD), and back (LES). The corresponding EMG MDF signals were concatenated to the Xsens and Apple Watch matrices to be used as ground truths for fatigue in the FCI.

6.2.4.3 Electrocardiogram (EKG)

The raw EKG signals from the fatiguing sets were processed using the biosppy.signals.ecg.ecg function (Bota et al., 2024). 600,000 frames (corresponding to 5 min of data; Catai et al., 2020; Tran et al., 2009) from each fatiguing set were used for both HRinst and LFHFratio calculations by discarding the final 1000 frames, then keeping the final 600,000 frames (i.e., a variable number of frames removed from the beginning of each set). The signal was filtered using a 600-order finite impulse response filter with bandpass cutoff frequencies set to 3 and 45 Hz. The indices of the R-peaks were located using the Hamilton QRS detection algorithm (Hamilton, 2002). This was followed by correction to the maximum within a tolerance set to *tol* = 0.05s, where the tolerance was defined as the time interval (*R – tol, R + tol*; Bota et al., 2024). A sample histogram of the normal-to-normal interval (NNI; time intervals between R-peaks; Gomes et al.,

2019) is shown in Fig. 6.5. Next, the instantaneous heart rate (HRinst) in beats per minute (bpm) was computed using Equation 3:

$$HRinst = \frac{60\ s}{NNI} \tag{3}$$

Heart rate variability (HRV) was quantified for each fatiguing set using the low-frequency high-frequency ratio (LFHFratio), using the pyHRV toolbox (Gomes et al., 2019). Using the previously-identified R-peaks and NNI, the NNI were interpolated using a 4 Hz resampling frequency, then the signal was detrended by subtracting the mean NNI (Gomes et al., 2019). Next, the signal was transformed to the frequency domain by using Welch's method to compute the power spectral density (Fig. 6.6); Welch's method divides data into overlapping segments, computes a periodogram using fast Fourier transform for each segment, then averages the periodograms (Solomon Jr, 1991). The frequency bands were defined as very low (0.00 – 0.04 Hz), low (0.04 – 0.15 Hz), and high (0.15 – 0.40 Hz), where LFHFratio was computed as the ratio of low to high frequency bands (Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, 1996).

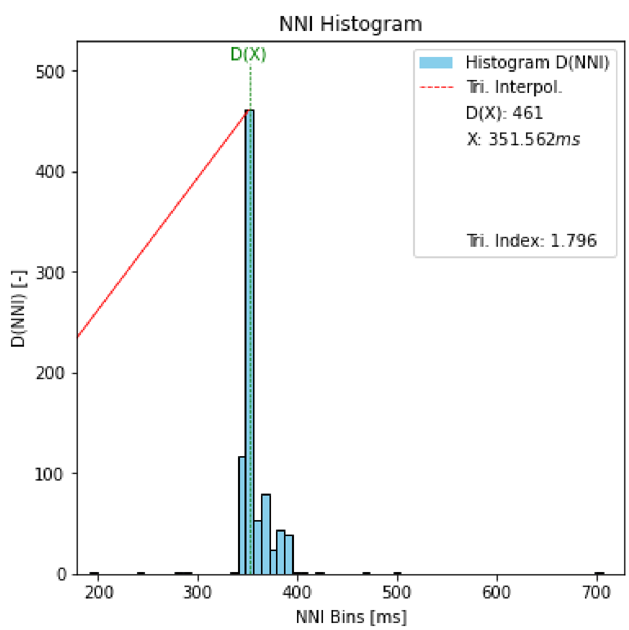


Fig. 6.5. A normal-to-normal interval (NNI) histogram for a sample 5 min manual handling set. The bins of NNI lengths are on the x-axis, while the frequency of their occurrence is on the y-axis. The mode NNI length was 351.562 ms (X), occurring 461 times (D(X)). The triangular index (Tri. Index: 1.796; red line) was computed as the total number of NNIs bins divided by the height of the histogram peak, providing insight into the risk of cardiovascular events (Hämmerle et al., 2020).

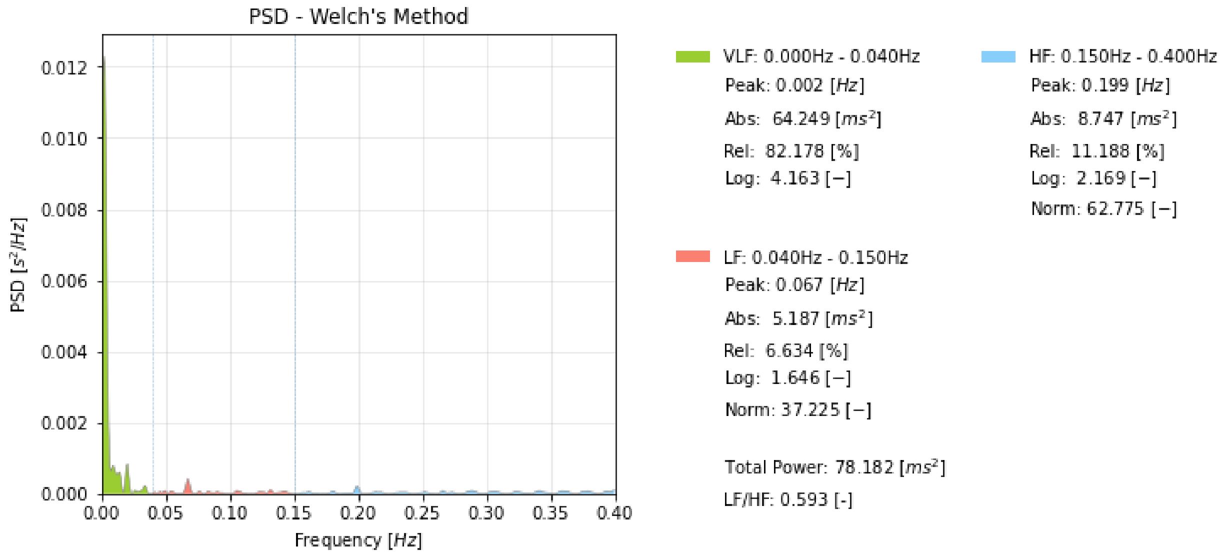


Fig. 6.6. The power spectral density (PSD) computed using Welch's method of a sample 5 min manual handling set. PSD is shown on the y-axis and frequency is shown on the x-axis. Very low frequency (VLF), low frequency (LF) and high frequency (HF) are depicted using green, red, and blue, respectively. Peak = peak frequencies; Abs = absolute powers; Rel = relative powers; Log = logarithmic powers; Norm = normalized powers.

To compute the mean and standard deviation HRinst and LFHFratio signals across all participants, the mean of the HRinst and LFHFratio signals across all sets for each participant were rubber banded to 0 – 100% to represent percentage of protocol completion. Then, the mean and standard deviation were computed at each percentage of completion. To produce HRinst and LFHFratio signals that were aligned with the Xsens and Apple Watch IMU data, up-sampling was conducted to produce HRinst and LFHFratio signals with lengths equal to the number of frames for each corresponding set in the Xsens and Apple Watch IMU data. The rubber banding and up-sampling were conducted using the PCHIP interpolator (Fritsch & Butland, 1984). The corresponding HRinst and LFHFratio signals were concatenated to the Xsens and Apple Watch matrices to be used as ground truths for fatigue in the FCI, while the HRinst was also used as an input in the fatigue estimation model.

6.2.4.4 Self-reported fatigue rating

The 10 cm fatigue VAS from each fatigue assessment was measured with a ruler to the nearest 0.01 cm, defined as the distance between the zero-line and the location where the mark made by participants intersected the VAS line. This resulted in one value per fatigue assessment, per participant. To compute the mean and standard deviation fatigue VAS signal across all participants, the fatigue VAS signal across all sets for each participant was rubber banded to 0 – 100% to represent percentage of protocol completion. Then, the mean and standard deviation were computed at each percentage of completion. To produce a fatigue VAS signal that aligned with the Xsens and Apple Watch IMU data, up-sampling was conducted to produce a fatigue VAS signal with length equal to the number of frames for each participant in the Apple Watch and Xsens IMU data. The rubber banding and up-sampling were conducted using the PCHIP interpolator (Fritsch & Butland, 1984). The corresponding fatigue VAS signals were concatenated to the Xsens and Apple Watch matrices to be used as a ground truth for fatigue in the FCI.

6.2.4.5 Force generation capacity

The average of the three peak values obtained during each MLS assessment was computed, resulting in one value per fatigue assessment, per participant. Then, this value was normalized to the participants' maximum value across all assessments by dividing by their maximum value, multiplied by 100%. To compute the mean and standard deviation MLS signal across all participants, the MLS signal across all sets for each participant was rubber banded to 0 – 100% to represent percentage of protocol completion. Then, the mean and standard deviation were computed at each percentage of completion. To produce an MLS signal that aligned with the Xsens and Apple Watch IMU data, up-sampling was conducted to produce an MLS signal with length equal to the number of frames for each participant in the Apple Watch and Xsens IMU data. The rubber banding and up-sampling were conducted using the PCHIP interpolator (Fritsch & Butland, 1984). The corresponding MLS signals were concatenated to the Xsens and Apple Watch matrices to be used as a ground truth for fatigue in the FCI.

6.2.5 Fatigue composite index (FCI)

The FCI was comprised of six, equally-weighted variables known to be associated with fatigue progression. These variables were fatigue VAS (self-reported fatigue rating), MLS (force generation capacity), EMG MDF of seven muscles (muscle contractile function), HRinst (blood flow and oxygenation), LFHFratio (sympathetic arousal), and pelvic jerk (motor control). Once all variables were processed (Chapters 6.2.4.1 – 6.2.4.5), they were globally normalized and averaged together to obtain the FCI.

To normalize fatigue VAS, HRinst, LFHFratio, and pelvic jerk, values were min-max normalized to 0 – 1 using Equation 4:

$$z_i = \frac{x_i - min_{(X)}}{max_{(X)} - min_{(X)}} \tag{4}$$

where x_i is the i^{th} original value, z_i is the i^{th} normalized value, and X is data from all participants. To normalize MLS and EMG MDF, values were min-max normalized to 0 – 1 using Equation 5:

$$z_i = 1 - \left(\frac{x_i - min_{(X)}}{max_{(X)} - min_{(X)}}\right) \tag{5}$$

where x_i is the i^{th} original value, z_i is the i^{th} normalized value, and X is data from all participants. Variables were treated differently during the normalization procedure based on whether they had a positive (VAS, HRinst, LFHFratio, pelvic jerk; Tran et al., 2009; Ueda et al., 2006; Zhang et al., 2019) or negative (MLS, EMG MDF; Cifrek et al., 2009; Nocella et al., 2011) association with fatigue. Equations 3 and 4 ensured that values closer to 0 represented little fatigue while values closer to 1 meant greater fatigue for all variables. The decision to normalize the fatigue variables to the minimum and maximum values across all participants was to allow for comparability across participants, ensure the scale is consistent across participants, and ensure estimation models can generalize to unseen individuals. Once normalized, the six variables were averaged together to produce the FCI. Lastly, the FCI was smoothed using a moving average with a window size set to 40 × sampling frequency (i.e., 120 Hz for Xsens, 40 Hz for smartwatch) shown in Fig. 6.7. These steps were performed once each for the Xsens and smartwatch data.

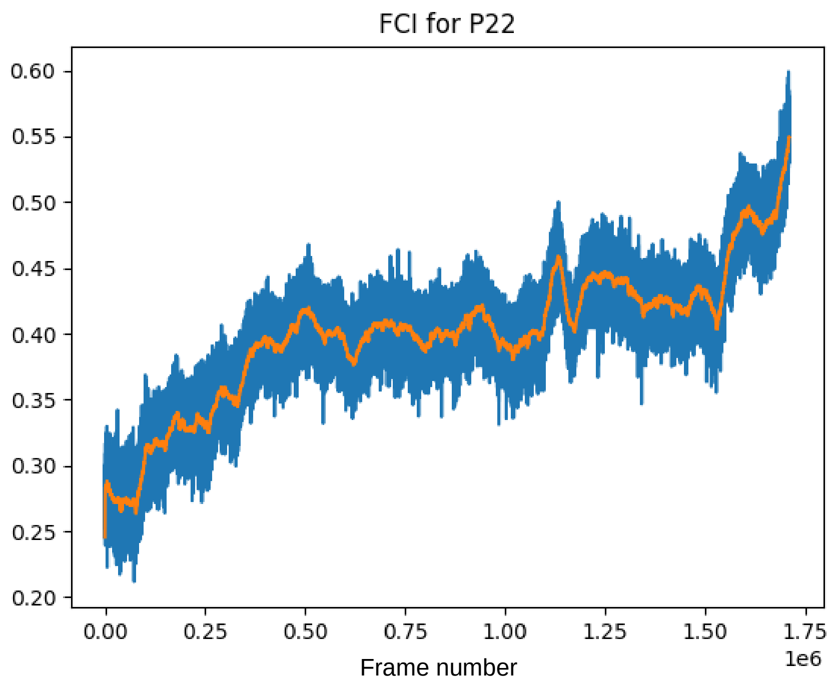


Fig. 6.7. The fatigue composite index (FCI) across all manual handling sets completed by participant 22, illustrating the effect of the moving average smoothing procedure with a window size of 40 × 120 Hz (the Xsens IMU sampling frequency). The blue line indicates the unsmoothed FCI, the orange line indicates the smoothed FCI, and the FCI is on the y-axis.

6.2.6 Data Preparation

One matrix for Xsens and smartwatch data each were created using data from all participants. The datasets were constructed by concatenating the data from all manual handling sets, across all participants. The task labels were previously included for each frame of data; all frames labelled as "other" were removed. Feature normalization was then conducted on a feature-by-feature basis using z-score normalization shown in Equation 6 (Clouthier et al., 2020; Friedman & Komogortsev, 2019; Mavor et al., 2023):

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{6}$$

where z_{ij} is the normalized value of feature j for a given participant i, x_{ij} is the value of feature j for a given participant i, and μ_j and σ_j are the mean and standard deviation of feature j in the subset (i.e., training or validation set), respectively. Lastly, all features were rounded to four decimal places.

The FCI was combined to the Xsens and Apple Watch matrices. Once prepared, the Xsens matrix was comprised of 277 features total: participant number, set number, time, 272 kinematic features (3 linear acceleration, 3 angular velocity, 3 magnetometer, 4 orientation, and 3 position for each of the 17 IMUs) the task label, and the FCI. The smartwatch matrix was comprised of 15 features total: participant number, set number, time, 10 kinematic features (3 linear acceleration, 3 angular velocity, and 4 orientation), the task label, and the FCI. Because the FCI estimation models were task-specific, the label feature was used to isolate subsets of data for ML (i.e., hyperparameter optimization and *k*-fold cross-validation), segregated by task. To reshape the data for ML, a sliding window approach was used to divide the data into segments containing an equal number of data frames (comprising either 0.5, 1, or 2 s of data), where the step size was always ¼ of the window size (Clouthier et al., 2020; Mavor et al., 2023). Each segment was assigned an FCI value according to the mean FCI for that segment. The dimension of the input data segments for the ANN was the sliding window size by the number of features.

6.2.7 Sensor combinations

Four sensor combinations were tested in this study, based on the results of sensitivity analysis of sensors for HAR (study 3; Chapter 5). The sensor combinations are presented in Table 6.2. Correlation-based feature selection was employed to determine which features to use from each sensor type (Hall, 2000); in short, this algorithm selected a subset of features that are correlated to the output variable, that are not correlated to each other. The selected features per sensor type determined using correlation-based feature selection are presented in Table 6.3. An additional feature included with all combinations was set number, reflecting the number of 5 min manual handling sets performed.

Table 6.2. Sensor combinations tested for fatigue estimation. IMU = inertial measurement unit; EKG = electrocardiogram; T8 = T8 vertebrae; RFt = right foot; LFt = left foot.

Combination Name	Sensors Included	No. of Features
Watch	Apple Watch IMU, EKG	12
T8	Xsens T8 IMU, EKG	12
3A	Xsens T8, RFt, LFt IMUs, EKG	32
T8_only	Xsens T8 IMU	11

Table 6.3. Features selected from each sensor type as input variables for the machine learning estimation of continuous fatigue. EKG = electrocardiogram; HRinst = instantaneous heart rate.

Sensor Type	Features
Xsens	Magnetometer X, Y, Z Orientation 0, 1, 2, 3 Position X, Y, Z
Apple Watch	Linear acceleration X, Y, Z Angular velocity X, Y, Z Orientation 0, 1, 2, 3
EKG	HRinst

6.2.8 Machine learning (ML)

All ML model development (i.e., hyperparameter optimization and *k*-fold cross validation) were synchronously distributed using a mirrored strategy across three Nvidia Titan RTX GPUs (Nvidia, Santa Clara, USA) with 24 GB of GDDR6 memory, each. This was performed on a server running Ubuntu 18.04 powered by two Intel Xeon Gold 6248 CPUs @ 2.50 GHz (Intel, Santa Clara, USA) and 384 GB of DDR4 RAM. All ML was conducted using Python 3.11 (Van Rossum & Drake, 2009) using TensorFlow 2.4.0 (Abadi et al., 2016).

6.2.8.1 Fully connected artificial neural network (ANN)

A fully connected artificial neural network (ANN) was selected to estimate continuous fatigue levels because it can be used end-to-end to automatically extract features and effectively model complex, non-linear relationships between features and the output, whilst avoiding excessive model complexity. The model architecture was comprised of an input layer, three fully-connected dense hidden layers with ReLU activation functions, and a dense output layer with a linear activation function (Fig. 6.8). For regularization, a drop-out operator with a probability set to $p = 0.25$ was implemented on the output of the dense layers. The model inputs were the reshaped data segments, and the ground truth FCI was used to compute loss (i.e., error). The dense hidden layers consisted of hidden units (i.e., neurons or nodes), which computed an element of the layer's output. The input into each hidden layer was the output of the previous layer. A set of randomly initialized parameters were used to control the transformations for each hidden unit and layer. Feed-forward propagation through the ANN was used to obtain a continuous output representing the predicted FCI. Prediction loss was computed using the mean squared error (MSE) function. During training, the gradients of the cost function with respect to the parameters were computed using backpropagation, where SGD was used to update the parameters to minimize loss. This was performed iteratively until a stopping criterion was reached.

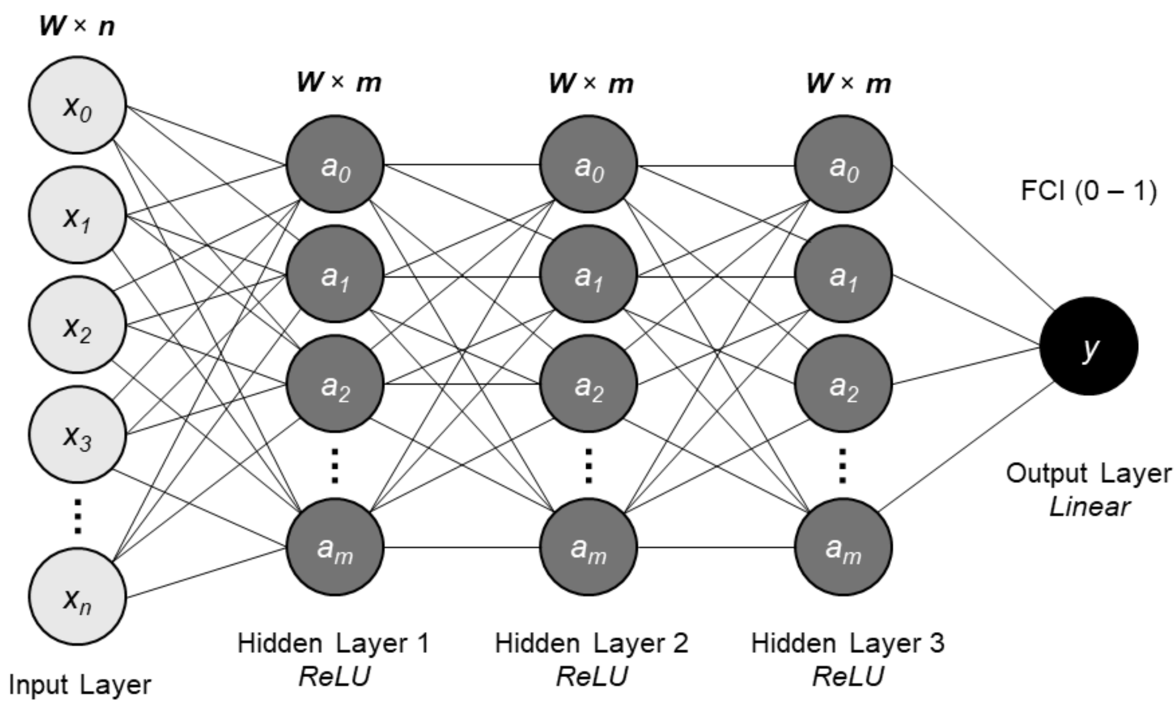


Fig. 6.8. Architecture of the fully connected artificial neural network. The number of input features (*n*) depended on the sensor configuration. *W* = sliding window size; *m* = number of hidden units; *x* = input units; *a* = hidden units; *y* = output unit; ReLU = rectified linear unit.

6.2.8.2 Hyperparameter optimization (HPO)

The sliding window size for each sensor combination was determined using the results from the sensitivity analysis of sensors for HAR (study 3; Chapter 5), presented in Table 6.4. The number of hidden units in the dense layers, SGD learning rate and momentum were optimized using Bayesian hyperparameter optimization implemented in KerasTuner library (O'Malley et al., 2019). The search space was guided by previous research with similar tasks and instrumentation (studies 2 and 3 in Chapters 4 and 5, respectively; Mavor et al., 2023). The objective was to minimize loss on the validation set, comprised of all data from three participants who were randomly selected for each trial. For each permutation of hyperparameters, a trial was executed and stopped early based on non-improving validation loss (MSE) with a patience set to 25. Data were loaded with a batch size of 64 and the number of trials were limited to 24. For feasibility, HPO was only conducted for the T8 sensor combination and the carry task only, then used for training and evaluation of all sensor combinations and task-specific models.

Table 6.4. Hyperparameter search space. The hyperparameter optimization procedure was performed only for the T8 combination during the carry task, then the results were used for all sensor combinations and tasks. Bolded values indicate the selected hyperparameter. SGD = stochastic gradient descent.

Window size	Hidden units	SGD learning rate	SGD momentum
Watch = 80 T8 = 120 3A = 120	64, 128, 192, **256**, 320	**0.0001**, 0.001, 0.01, 0.1	0.7, 0.9, 0.95, **0.98**

6.2.8.3 k-fold Cross-Validation

To train and evaluate task-specific ANN models, *k*-fold cross-validation was employed. To ensure all data from a participant was kept together (avoiding exposure to training and evaluation), the *k* was set to 8, where each fold was comprised of all data from three participants. The folds were randomly selected then kept constant across all task-specific models and sensor combinations.

Each cross-validation procedure was trained using a batch size of 64, maximum epochs of 500, and data were not shuffled prior to each epoch to preserve temporal dependencies that are important to estimating and forecasting human dynamics (Chiu et al., 2019). Training was stopped early based on non-improving validation loss with a patience set to 25. Then, the model weights from the epoch with the lowest validation loss were saved and used to make predictions on the test fold. Performance on each fold was evaluated using RMSE, MAE, Pearson's correlation coefficient, training time, and prediction time. Pearson's correlations were classified as weak ($r < 0.3$), moderate ($0.3 \leq r \leq 0.7$), and strong ($r > 0.7$) (Rusakov, 2023). Then, the average, standard deviation, maximum, and minimum of all performance metrics were computed across folds within a task-specific model to provide a performance estimate.

6.3 Results

6.3.1 Participants and dataset

Participant and dataset results were presented in chapters 5.3.1 and 5.3.2, respectively. The mean (SD) of the FCI at the start and end of the protocols were 0.25 (0.068) and 0.46 (0.11), respectively, with a mean FCI difference of 0.21. The mean start to end difference in FCI for participants who: stopped voluntarily was 0.20; stopped due to one or more criteria (i.e., decline in MLS and/or increase in fatigue VAS) was 0.21; and stopped due to two stopping criteria (i.e., decline in MLS and increase in fatigue VAS) was 0.28.

6.3.2 Changes in fatigue variables

The mean and standard deviation for all fatigue variables and the FCI across all participants from 0 – 100% of protocol completion are presented in this subchapter. When averaged across all participants, changes were visually observed on for fatigue VAS (Fig. 6.9), MLS (Fig. 6.10), and HRinst (Fig. 6.11), whereas no (consistent) changes were visually observed for LFHFratio (Fig. 6.12) and pelvic jerk (Fig. 6.14). Initial decreases in EMG MDF were observed in approximately the first 10% of protocol completion (Fig. 6.13), but did not persist throughout the protocol. A visible increase in the FCI can be observed (Fig. 6.15).

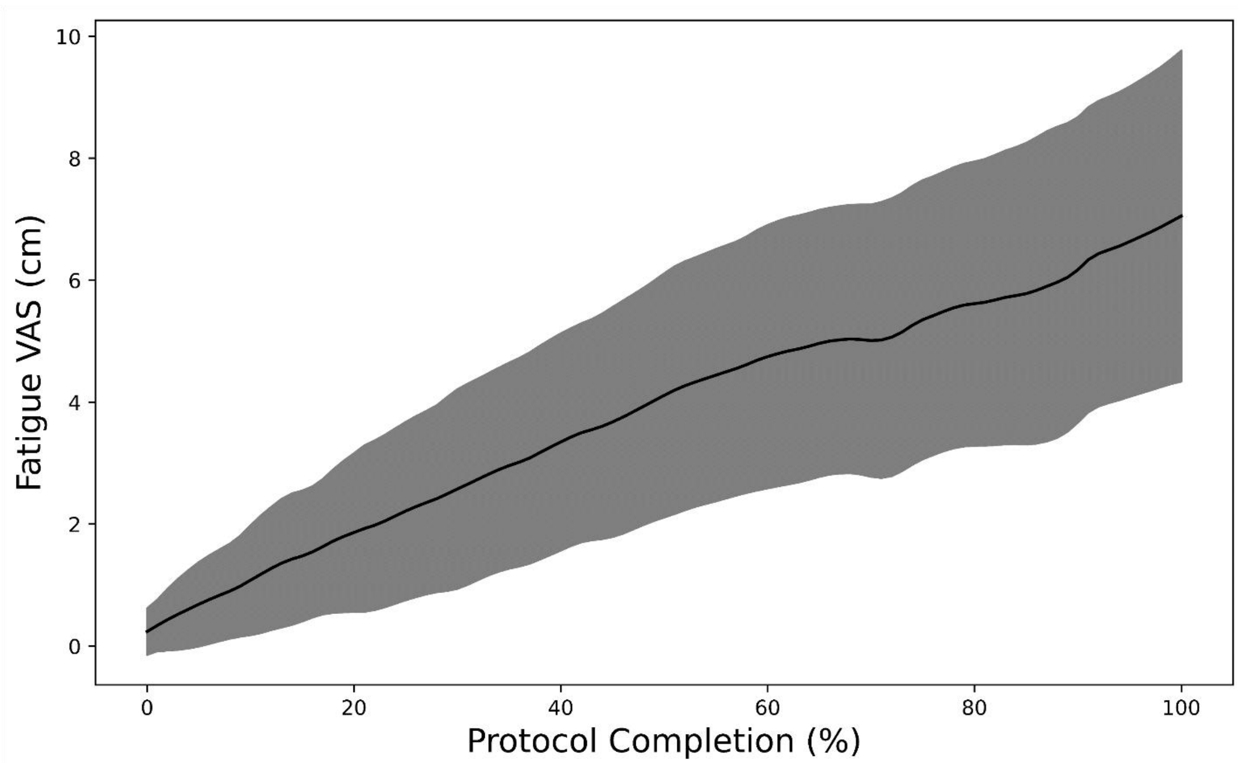


Fig. 6.9. Fatigue visual analog scale (VAS) across participants for 0 – 100% of manual handling protocol completion. Mean and standard deviation are indicated by the black line and grey shading, respectively.

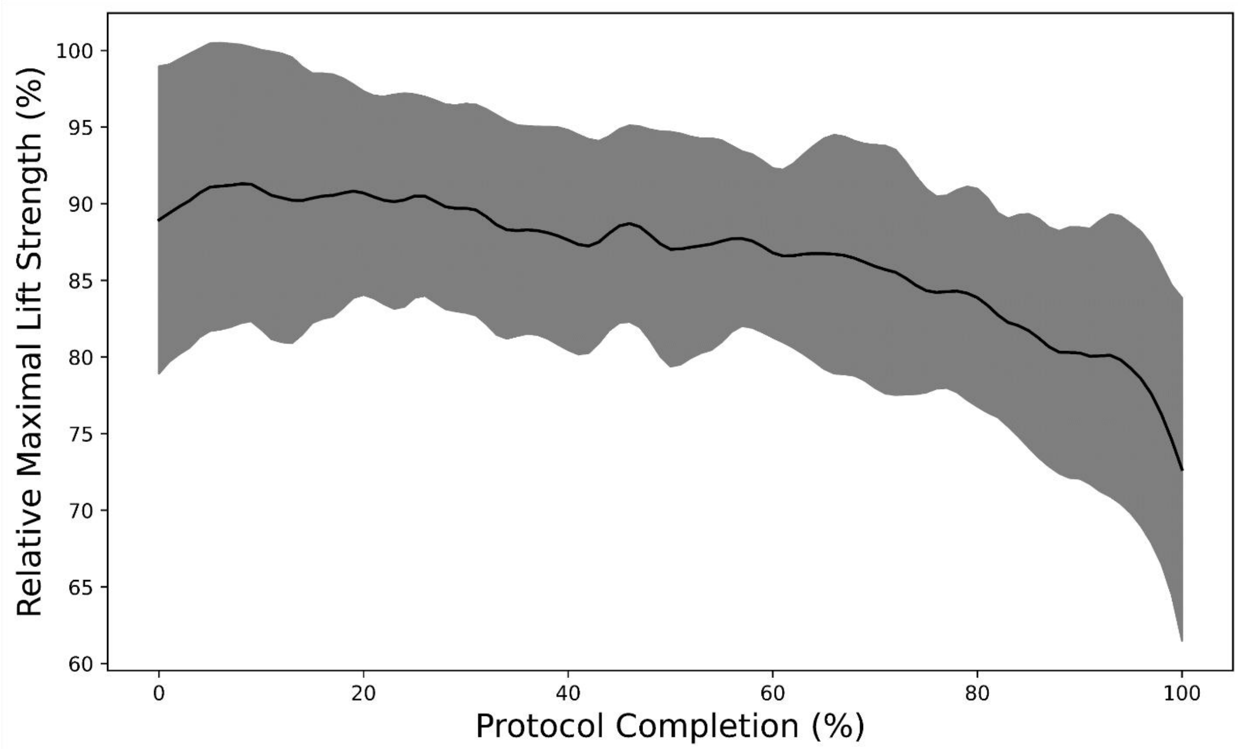


Fig. 6.10. Relative maximal lift strength across participants for 0 – 100% of manual handling protocol completion. Mean and standard deviation are indicated by the black line and grey shading, respectively.

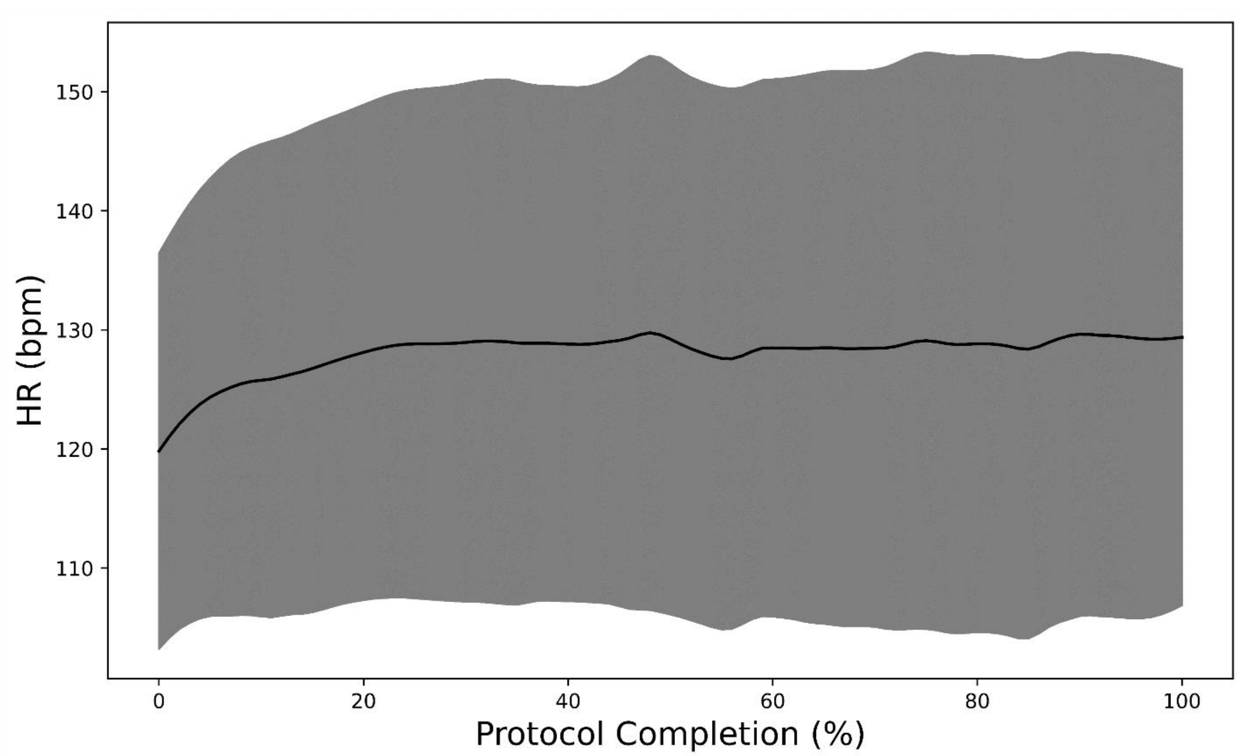


Fig. 6.11. Heart rate (HR) across participants for 0 – 100% of manual handling protocol completion. Mean and standard deviation are indicated by the black line and grey shading, respectively.

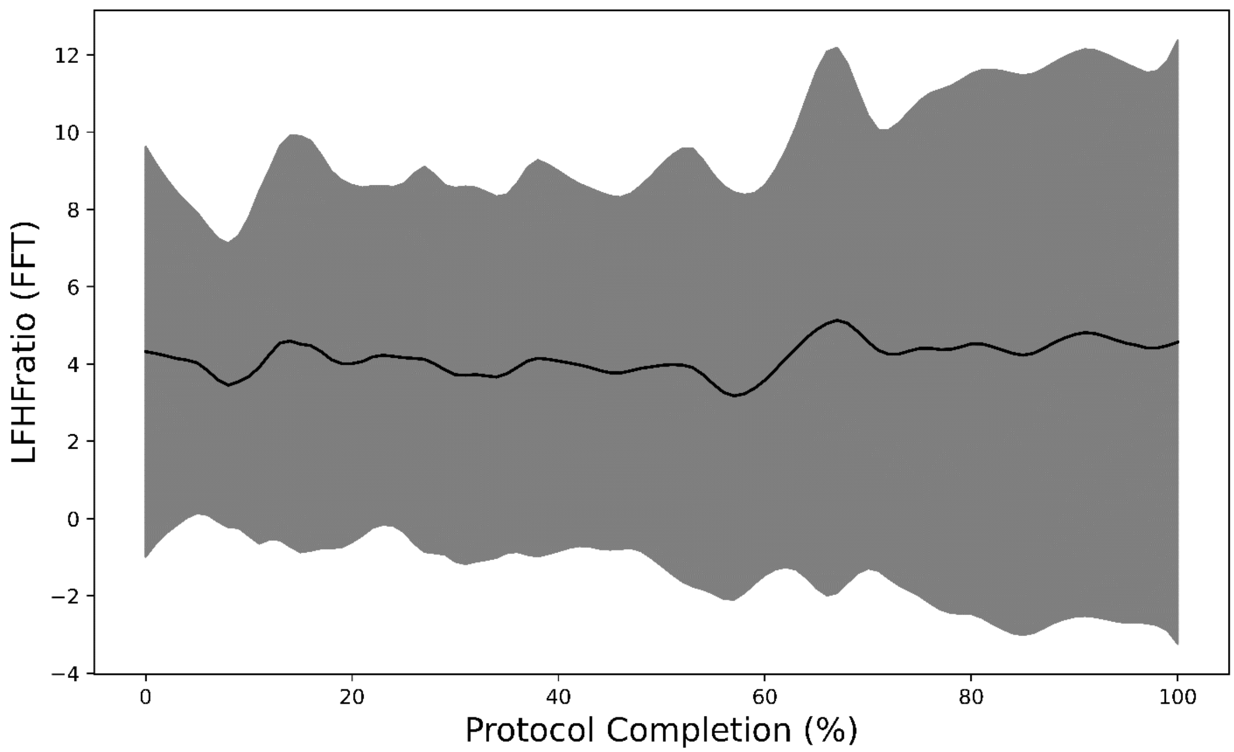


Fig. 6.12. Heart rate variability quantified using the low frequency high frequency ratio (LFHFratio) across participants for 0 – 100% of manual handling protocol completion. LFHFratio was computed using Welch's method, which employs fast Fourier transform (FFT) to generate a periodogram. Mean and standard deviation are indicated by the black line and grey shading, respectively.

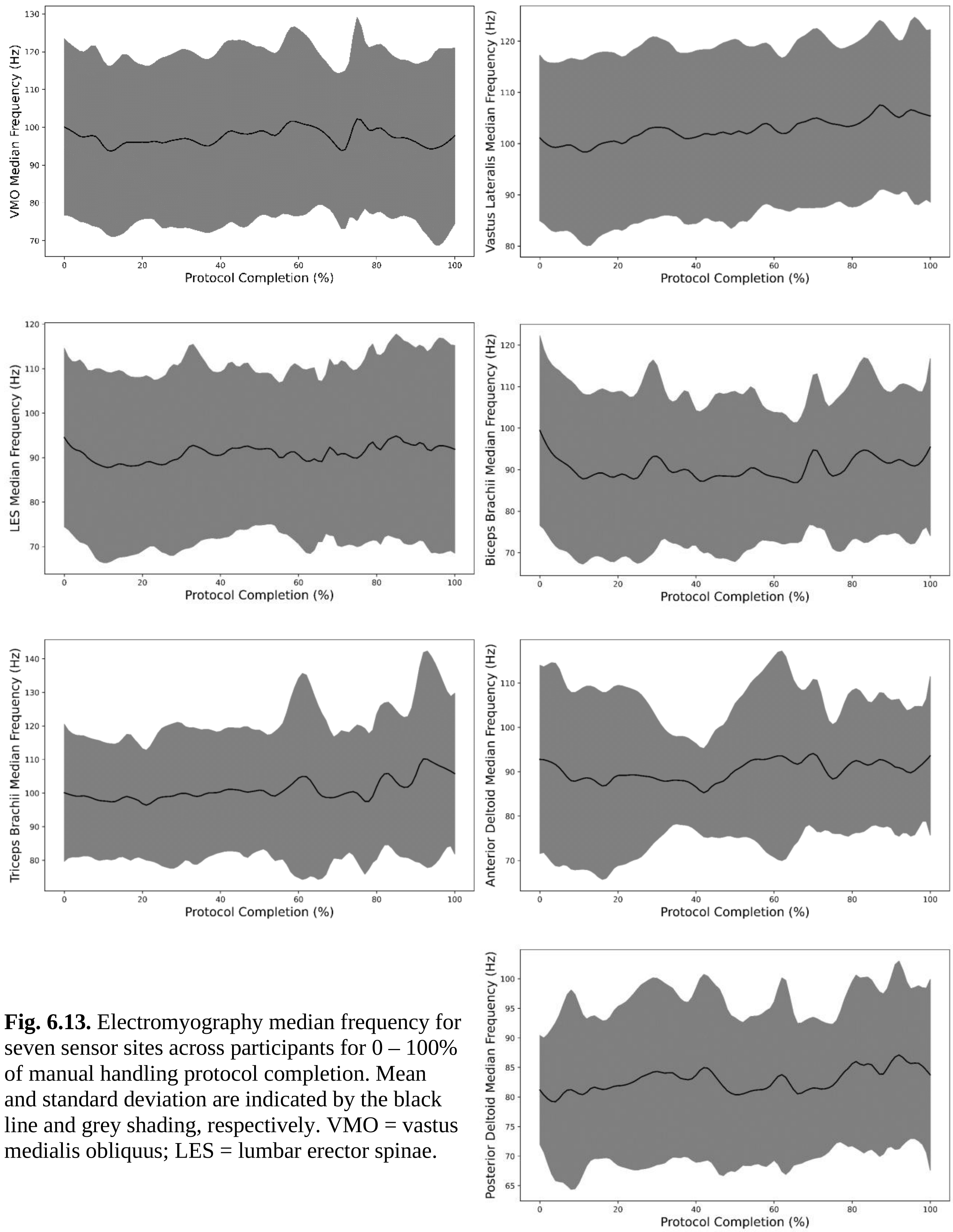


Fig. 6.13. Electromyography median frequency for seven sensor sites across participants for 0 – 100% of manual handling protocol completion. Mean and standard deviation are indicated by the black line and grey shading, respectively. VMO = vastus medialis obliquus; LES = lumbar erector spinae.

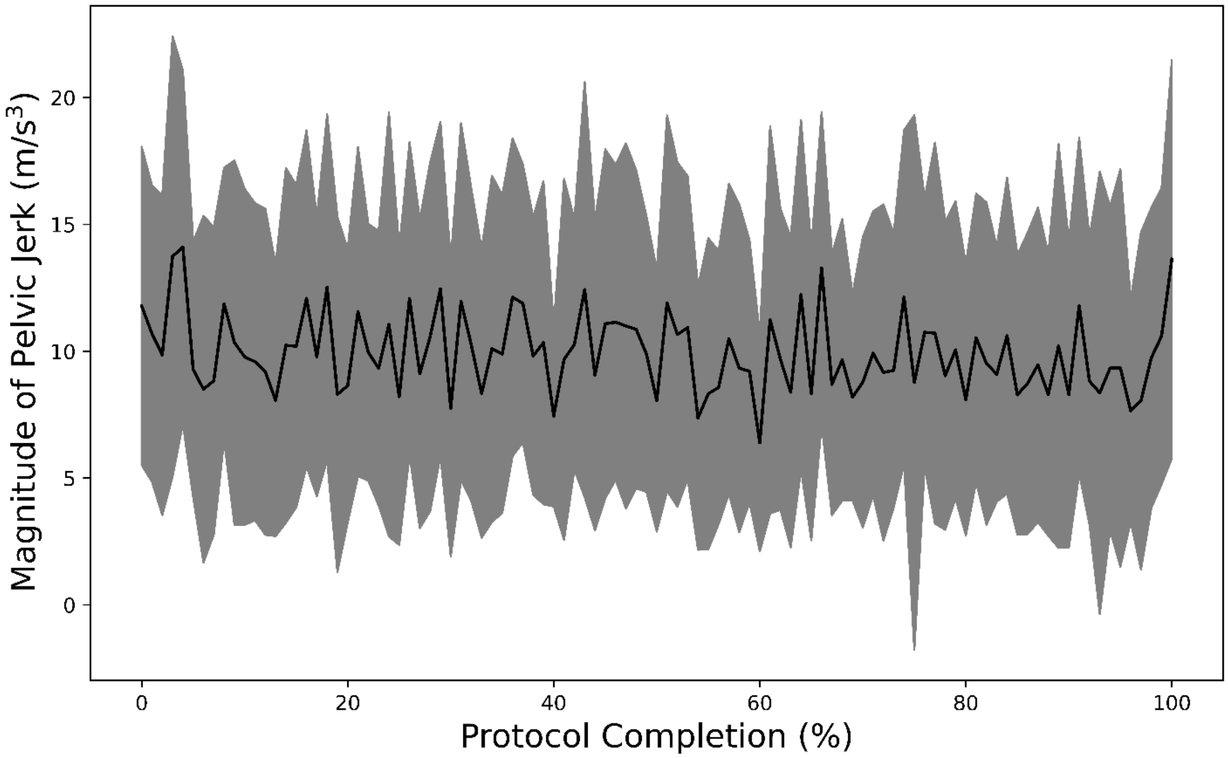


Fig. 6.14. Resultant of pelvic linear jerk across participants for 0 – 100% of manual handling protocol completion. Mean and standard deviation are indicated by the black line and grey shading, respectively.

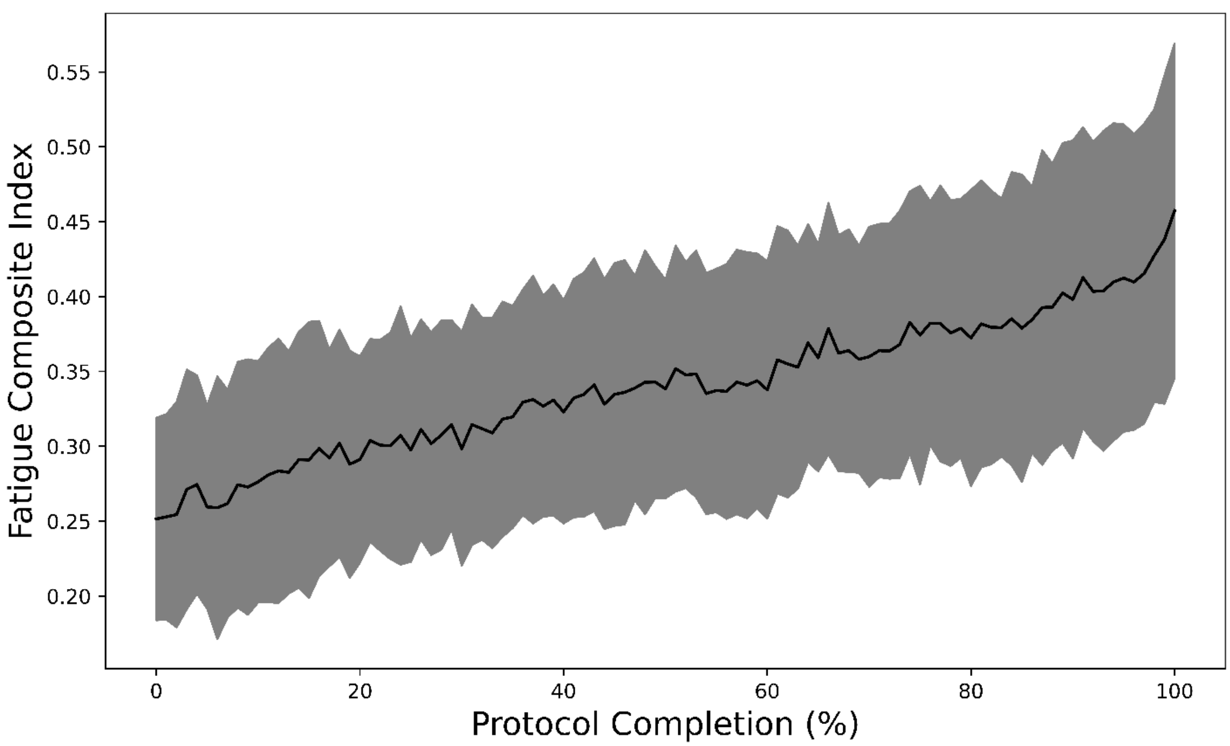


Fig. 6.15. The fatigue composite index across participants for 0 – 100% of manual handling protocol completion. Mean and standard deviation are indicated by the black line and grey shading, respectively.

6.3.3 Fatigue estimation performance

6.3.3.1 Overall performance of sensor combinations

The overall performances of the ANNs for each sensor combination are presented in Table 6.5. The average time required by the ANNs to make an estimation on each of second of input data are presented in Table 6.6. These summary results were obtained by aggregating results across all folds and tasks, for each sensor combination. On average, the T8 combination had the smallest RMSE (0.064 FCI) and MAE (0.051 FCI) and largest Pearson's correlation (r = 0.590). Only predictions from the T8 combination are shown below.

Table 6.5. Overall performance of the fully connected artificial neural network for each sensor combination. Performance metrics were aggregated across all folds and tasks for each sensor combination. SD = standard deviation; RMSE = root mean squared error; MAE = mean absolute error; FCI = fatigue composite index.

Summary metric	Sensor combination	RMSE (FCI)	MAE (FCI)	Pearson's correlation (r)	Training time (s)	Prediction time (s)
Mean (SD)	T8	0.064 (0.015)	0.051 (0.013)	0.590 (0.203)	689.81 (877.61)	1.78 (1.31)
	Watch	0.065 (0.012)	0.052 (0.010)	0.565 (0.170)	430.62 (536.38)	0.93 (0.65)
	3A	0.069 (0.017)	0.054 (0.013)	0.496 (0.240)	987.44 (1872.27)	1.98 (1.51)
	T8_only	0.070 (0.017)	0.056 (0.014)	0.461 (0.233)	561.09 (693.03)	1.76 (1.28)
Minimum - Maximum	T8	0.036 - 0.114	0.028 - 0.105	0.052 - 0.895	92.33 - 4618.74	0.42 - 5.51
	Watch	0.042 - 0.098	0.032 - 0.076	0.043 - 0.843	51.33 - 2534.24	0.24 - 2.73
	3A	0.038 - 0.114	0.029 - 0.087	0.043 - 0.941	95.42 - 10504.62	0.44 - 6.31
	T8_only	0.035 - 0.103	0.026 - 0.084	0.043 - 0.805	88.92 – 4267.97	0.40 – 5.32

Table 6.6. The average time required per second of input data for the artificial neural networks to make an estimation on the fatigue composite index for each sensor combination.

Sensor combination	Estimation time per second of input data (ms)
T8	0.108
Watch	0.0557
3A	0.120
T8_only	0.107

6.3.3.2 Correlations

The number of folds where the FCI was estimated with weak, moderate, and strong correlations is shown in Table 6.7. Across all folds, tasks, and sensor combinations, the ANNs were able to predict the FCI with weak, moderate, and strong correlations 27, 119, and 46 times (14.1, 62.0, and 24.0% of the time), respectively.

Table 6.7. A tally of the number of folds per sensor combination and task where the fatigue composite index was estimated with weak, moderate, and strong Pearson's correlation to the actual values.

Sensor Combination	Correlation strength	Stand	Pull	Lift	Carry	Push	Walk	Total
T8	Weak	2	0	0	0	2	0	4
	Moderate	4	3	5	7	3	7	29
	Strong	2	5	3	1	3	1	15
Watch	Weak	1	1	0	1	0	0	3
	Moderate	5	7	7	6	5	4	34
	Strong	2	0	1	1	3	4	11
3A	Weak	2	1	0	2	3	1	9
	Moderate	4	6	5	5	3	5	28
	Strong	2	1	3	1	2	2	11
T8_only	Weak	1	0	1	4	2	3	11
	Moderate	5	6	5	4	4	4	28
	Strong	2	2	2	0	2	1	9

6.3.3.3 Stand fatigue estimation performance

The errors for the ANNs to estimate the FCI during the stand task for each sensor combination are presented in Fig. 6.16, while the Pearson's correlations between the estimated and actual FCI are presented in Fig. 6.17. The training and prediction times are presented in Table 6.8. To provide a visual sample of the prediction, the best and worst folds from T8 are presented in Fig. 6.18.

6.3.3.4 Pull fatigue estimation performance

The errors for the ANNs to estimate the FCI during the pull task for each sensor combination are presented in Fig. 6.19, while the Pearson's correlations between the estimated and actual FCI are presented in Fig. 6.20. The training and prediction times are presented in Table 6.9. To provide a visual sample of the prediction, the best and worst folds from T8 are presented in Fig. 6.21.

6.3.3.5 Lift fatigue estimation performance

The errors for the ANNs to estimate the FCI during the lift task for each sensor combination are presented in Fig. 6.22, while the Pearson's correlations between the estimated and actual FCI are presented in Fig. 6.23. The training and prediction times are presented in Table 6.10. To provide a visual sample of the prediction, the best and worst folds from T8 are presented in Fig. 6.24.

6.3.3.6 Carry fatigue estimation performance

The errors for the ANNs to estimate the FCI during the carry task for each sensor combination are presented in Fig. 6.25, while the Pearson's correlations between the estimated and actual FCI are presented in Fig. 6.26. The training and prediction times are presented in Table 6.11. To provide a visual sample of the prediction, the best and worst folds from T8 are presented in Fig. 6.27.

6.3.3.7 Push fatigue estimation performance

The errors for the ANNs to estimate the FCI during the push task for each sensor combination are presented in Fig. 6.28, while the Pearson's correlations between the estimated and actual FCI are presented in Fig. 6.29. The training and prediction times are presented in Table 6.12. To provide a visual sample of the prediction, the best and worst folds from the T8 are presented in Fig. 6.30.

6.3.3.8 Walk fatigue estimation performance

The errors for the ANNs to estimate the FCI during the walk task for each sensor combination are presented in Fig. 6.31, while the Pearson's correlations between the estimated and actual FCI are presented in Fig. 6.32. The training and prediction times are presented in Table 6.13. To provide a visual sample of the prediction, the best and worst folds from T8 are presented in Fig. 6.33.

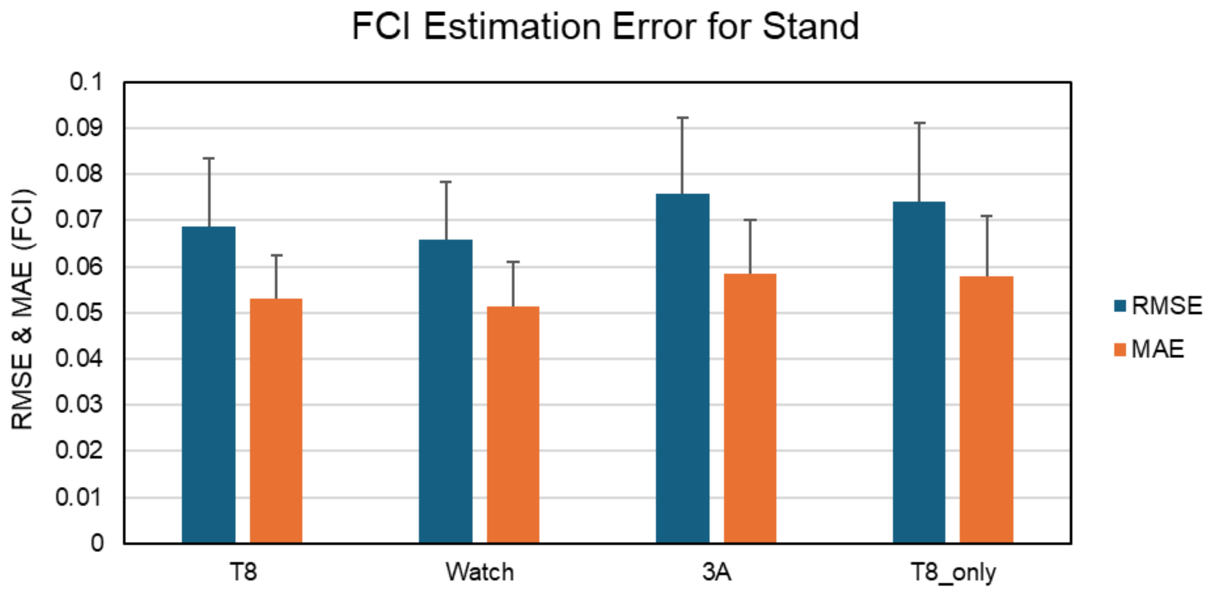


Fig. 6.16. The root mean squared error (RMSE) and mean absolute error (MAE) for the estimations of the fatigue composite index (FCI) during the stand task when averaged across all folds. The error bars indicate the standard deviation.

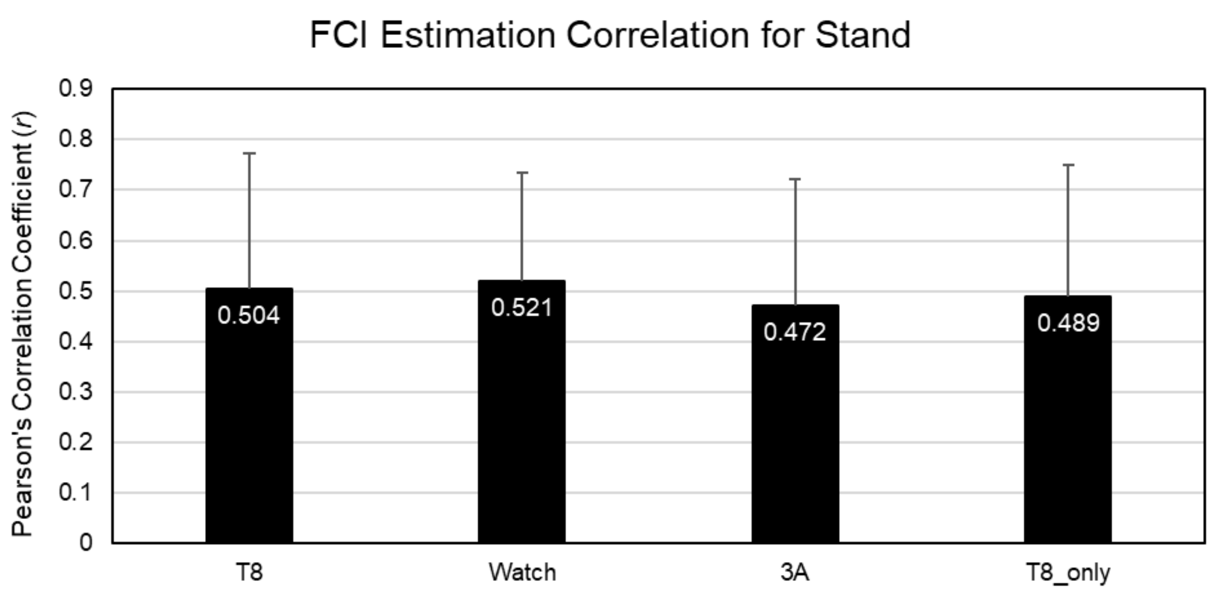


Fig. 6.17. Pearson's correlation coefficient (r) between the estimated and actual fatigue composite index (FCI) during the stand task when averaged across all folds. The error bars indicate the standard deviation.

Table 6.8. Training and prediction times of the artificial neural network for the stand task when averaged across all folds, for all sensor combinations. Values are presented as mean (standard deviation).

Sensor combination	Training time (s)	Prediction time (s)
T8	318.95 (284.45)	0.64 (0.22)
Watch	167.82 (143.39)	0.36 (0.11)
3A	412.31 (415.78)	0.70 (0.23)
T8_only	432.00 (443.39)	0.63 (0.22)

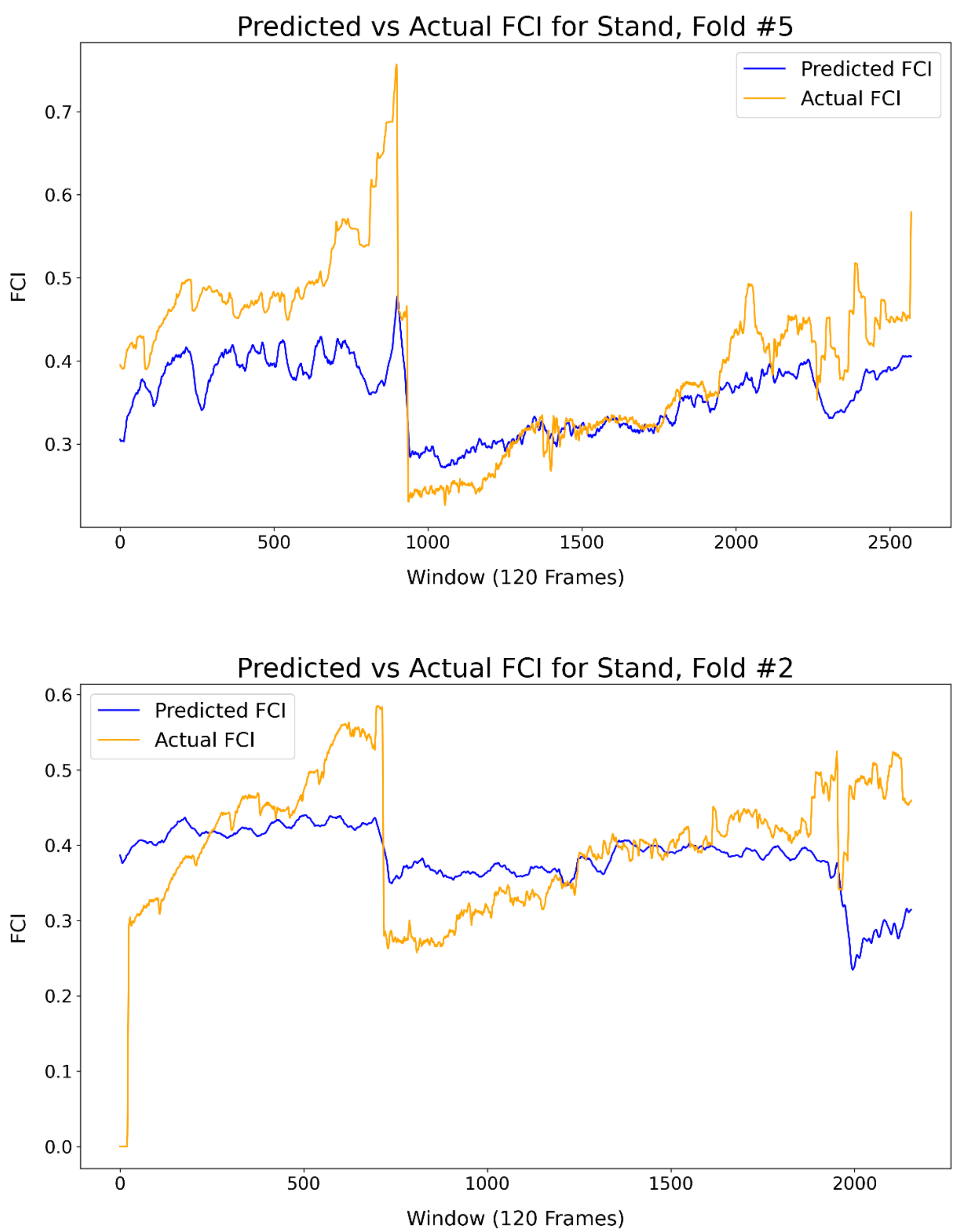


Fig. 6.18. The predicted versus actual fatigue composite index (FCI) for the T8 sensor combination during the stand task. The best (fold 5, MAE = 0.066, r = 0.85) and worst (fold 2, MAE = 0.068, r = 0.12) folds are pictured at the top and bottom, respectively.

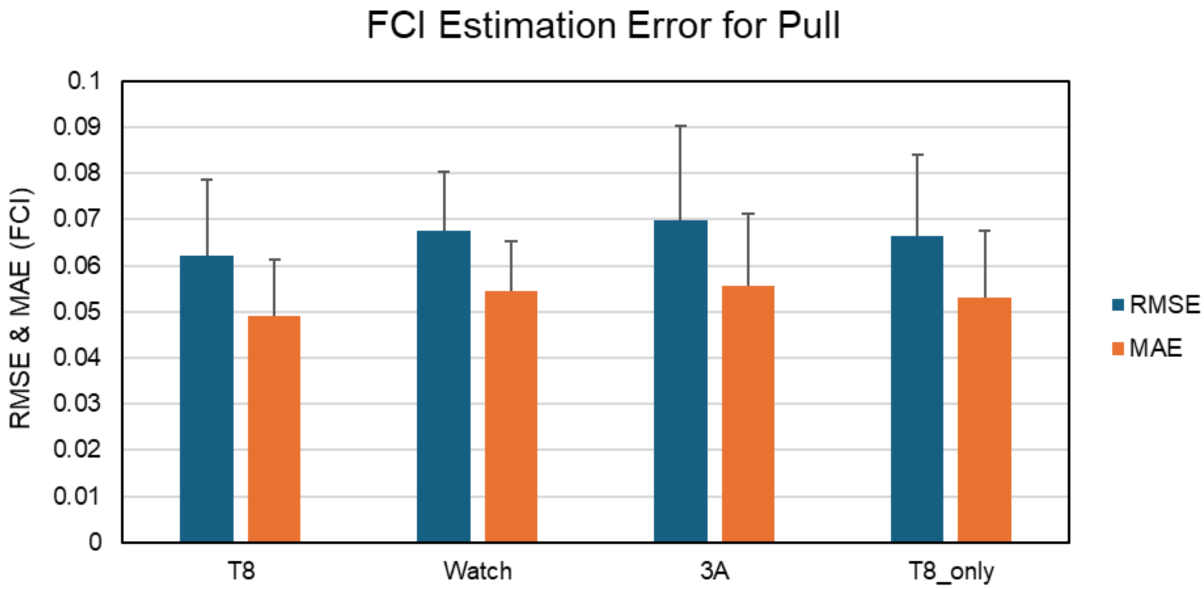


Fig. 6.19. The root mean squared error (RMSE) and mean absolute error (MAE) for the estimations of the fatigue composite index (FCI) during the pull task when averaged across all folds. The error bars indicate the standard deviation.

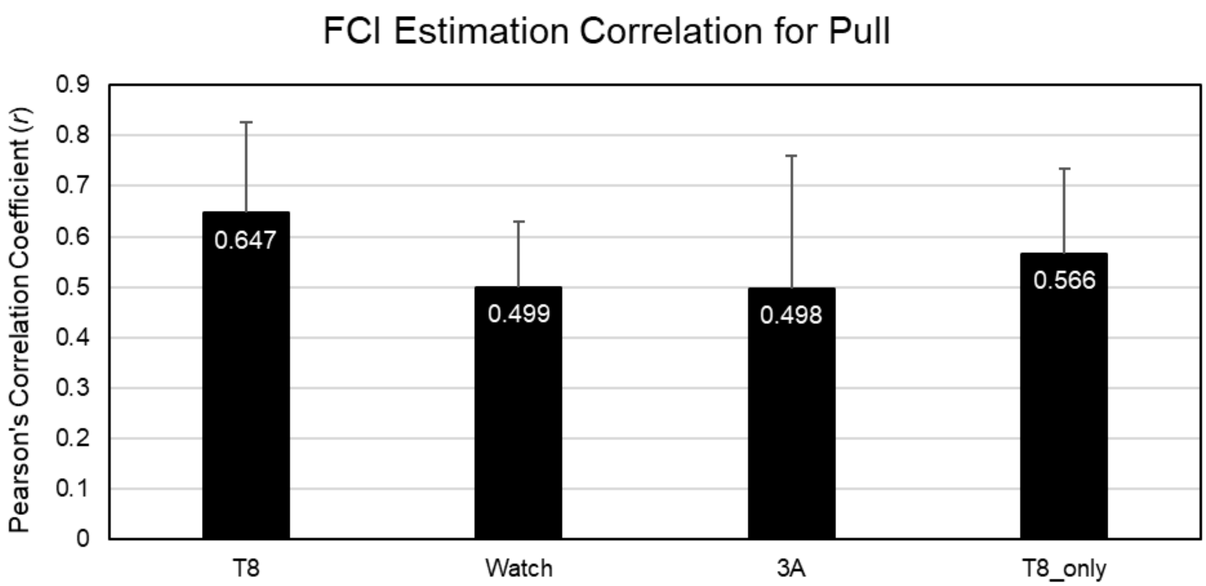


Fig. 6.20. Pearson's correlation coefficient (r) between the estimated and actual fatigue composite index (FCI) during the pull task when averaged across all folds. The error bars indicate the standard deviation.

Table 6.9. Training and prediction times of the artificial neural network for the pull task when averaged across all folds, for all sensor combinations. Values are presented as mean (standard deviation).

Sensor combination	Training time (s)	Prediction time (s)
T8	249.42 (219.63)	1.03 (0.34)
Watch	212.44 (159.17)	0.55 (0.15)
3A	297.85 (160.33)	1.11 (0.31)
T8_only	210.34 (125.12)	1.02 (0.28)

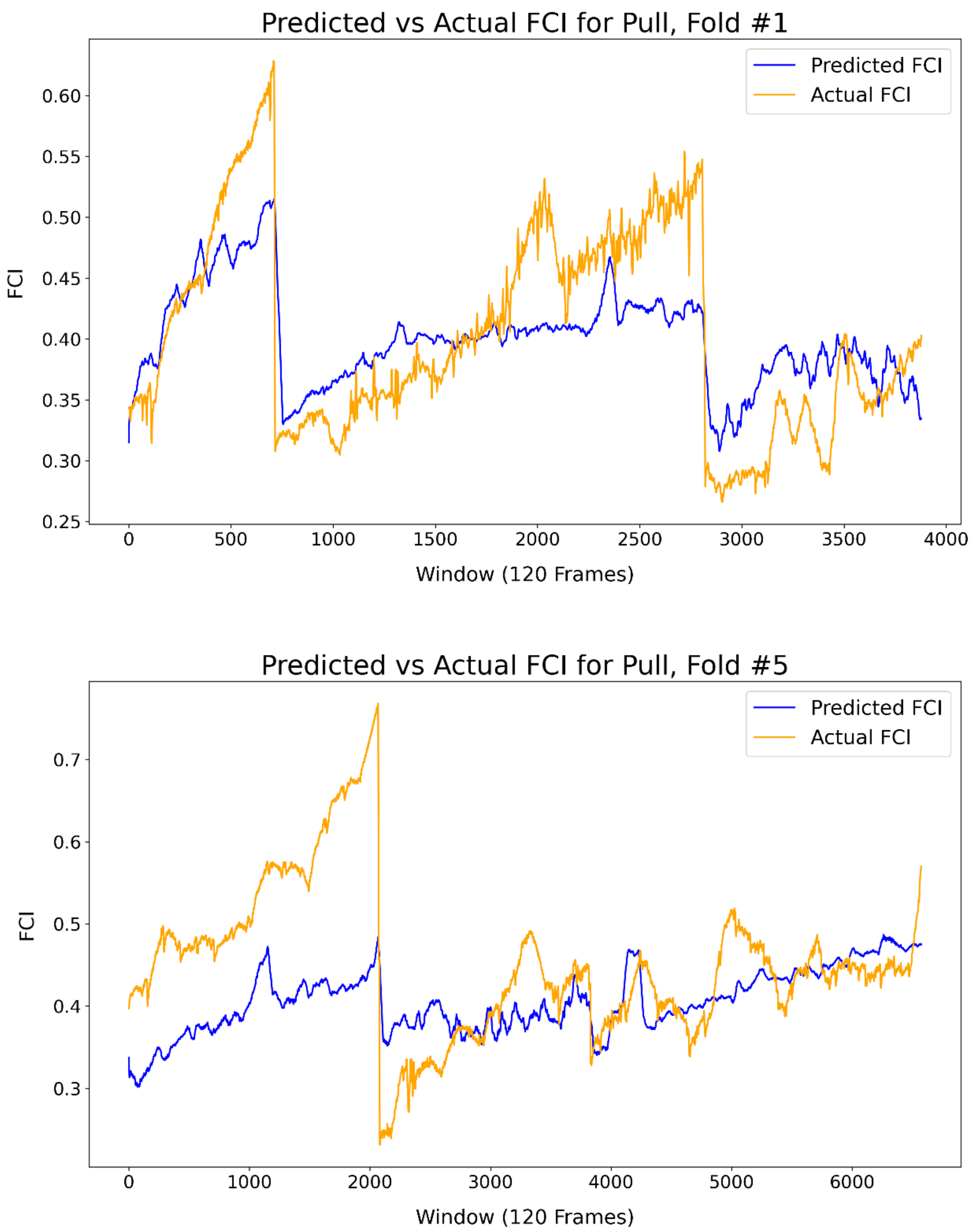


Fig. 6.21. The predicted versus actual fatigue composite index (FCI) for the T8 sensor combination during the pull task. The best (fold 1, MAE = 0.046, r = 0.84) and worst (fold 5, MAE = 0.074, r = 0.34) folds are pictured at the top and bottom, respectively.

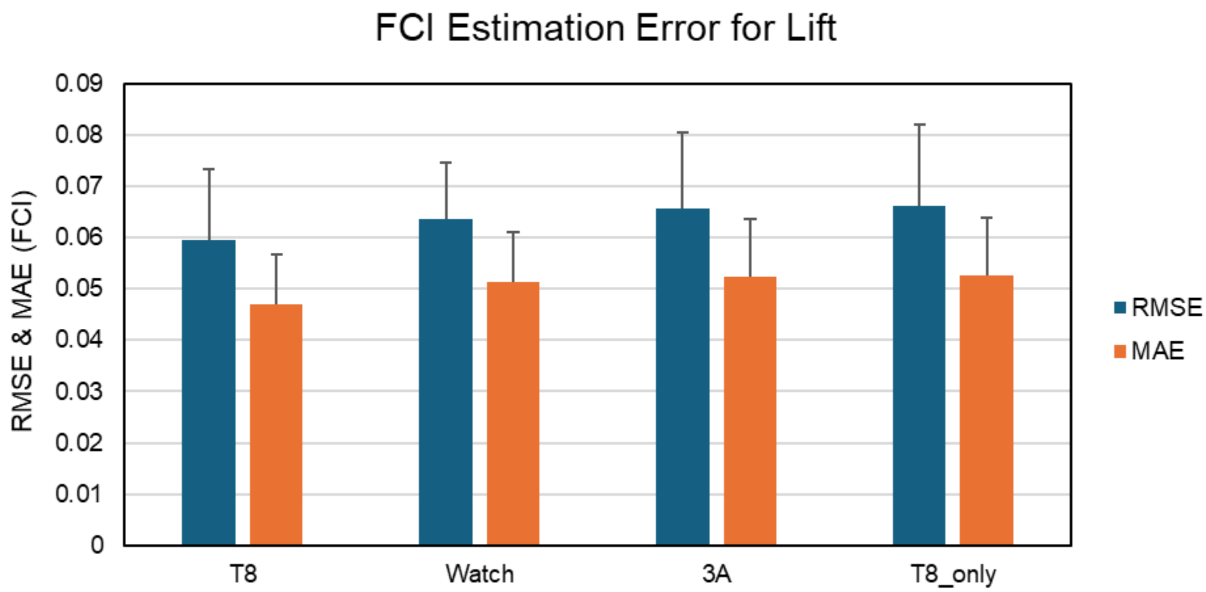


Fig. 6.22. The root mean squared error (RMSE) and mean absolute error (MAE) for the estimations of the fatigue composite index (FCI) during the lift task when averaged across all folds. The error bars indicate the standard deviation.

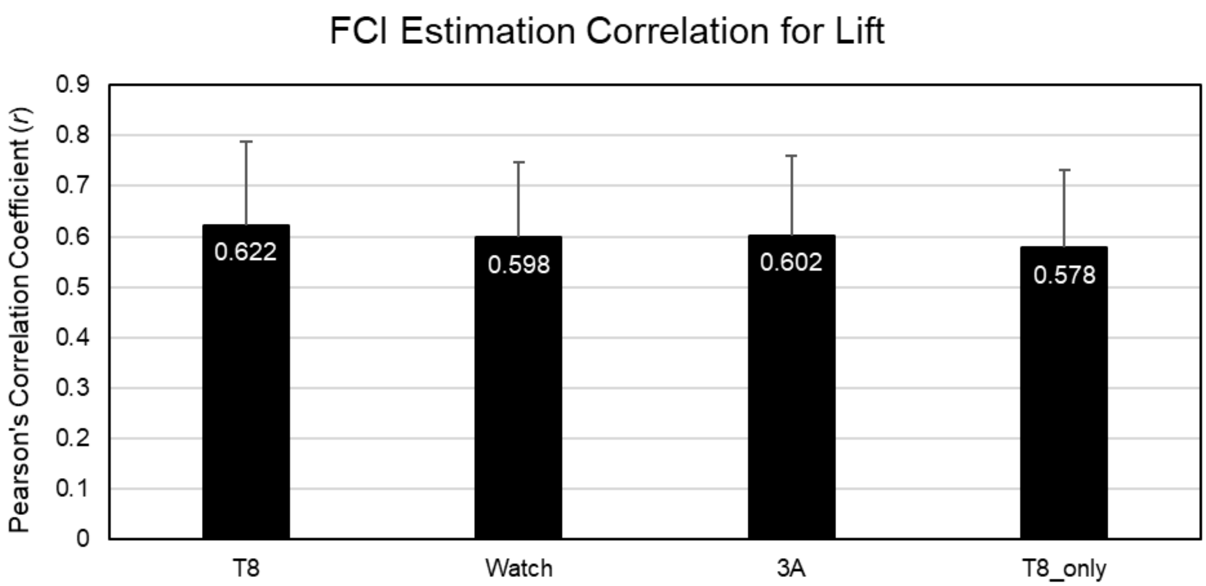


Fig. 6.23. Pearson's correlation coefficient (r) between the estimated and actual fatigue composite index (FCI) during the lift task when averaged across all folds. The error bars indicate the standard deviation.

Table 6.10. Training and prediction times of the artificial neural network for the lift task when averaged across all folds, for all sensor combinations. Values are presented as mean (standard deviation).

Sensor combination	Training time (s)	Prediction time (s)
T8	382.29 (397.91)	0.82 (0.19)
Watch	203.96 (156.57)	0.46 (0.10)
3A	230.34 (112.00)	0.88 (0.20)
T8_only	184.72 (111.07)	0.84 (0.20)

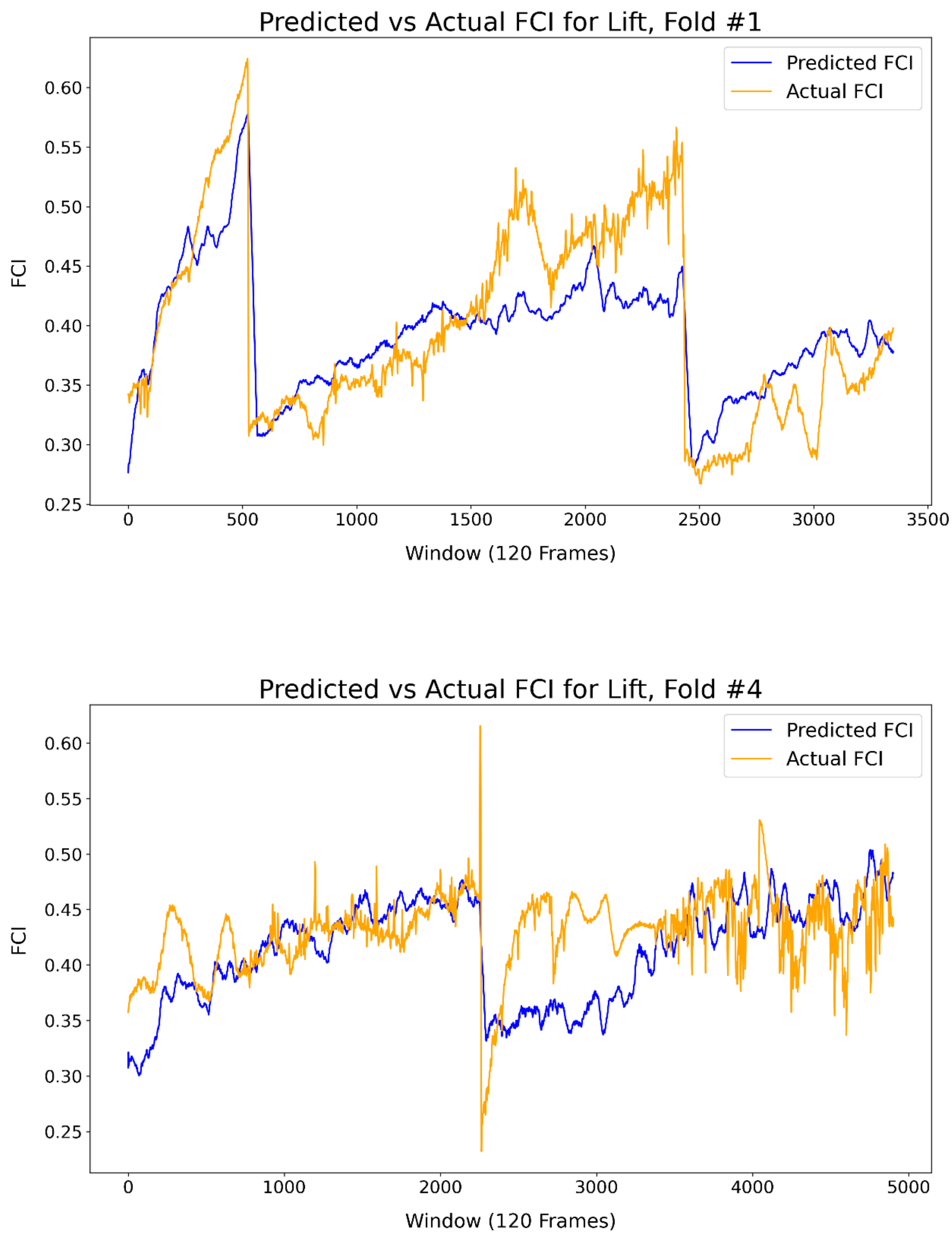


Fig. 6.24. The predicted versus actual fatigue composite index (FCI) for the T8 sensor combination during the lift task. The best (fold 1, MAE = 0.038, r = 0.84) and worst (fold 4, MAE = 0.043, r = 0.40) folds are pictured at the top and bottom, respectively.

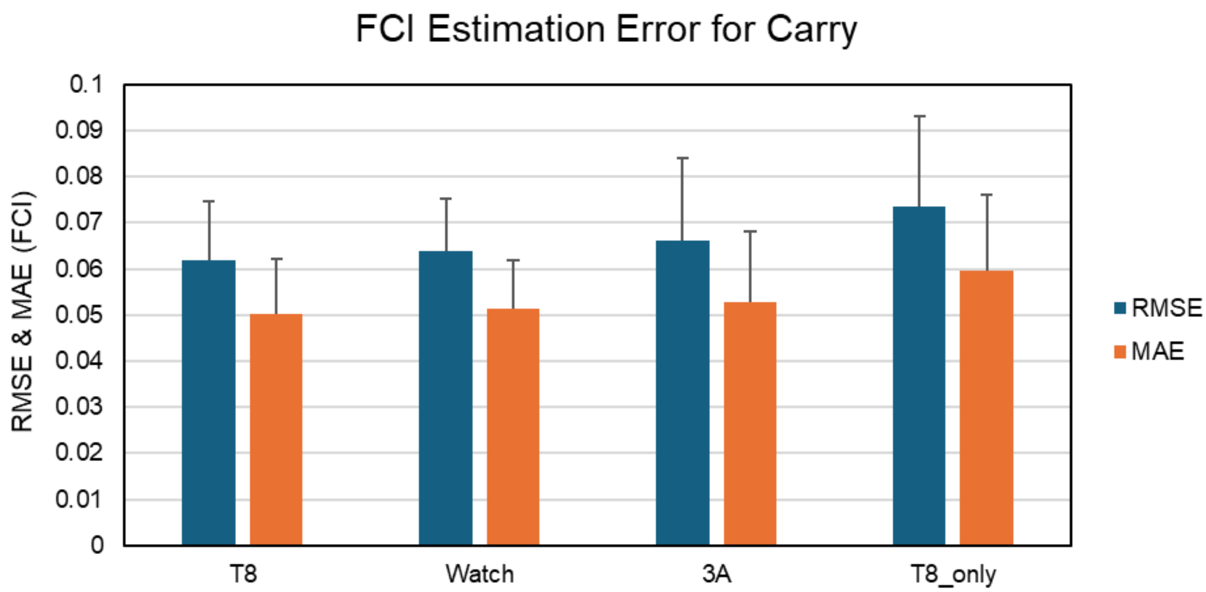


Fig. 6.25. The root mean squared error (RMSE) and mean absolute error (MAE) for the estimations of the fatigue composite index (FCI) during the carry task when averaged across all folds. The error bars indicate the standard deviation.

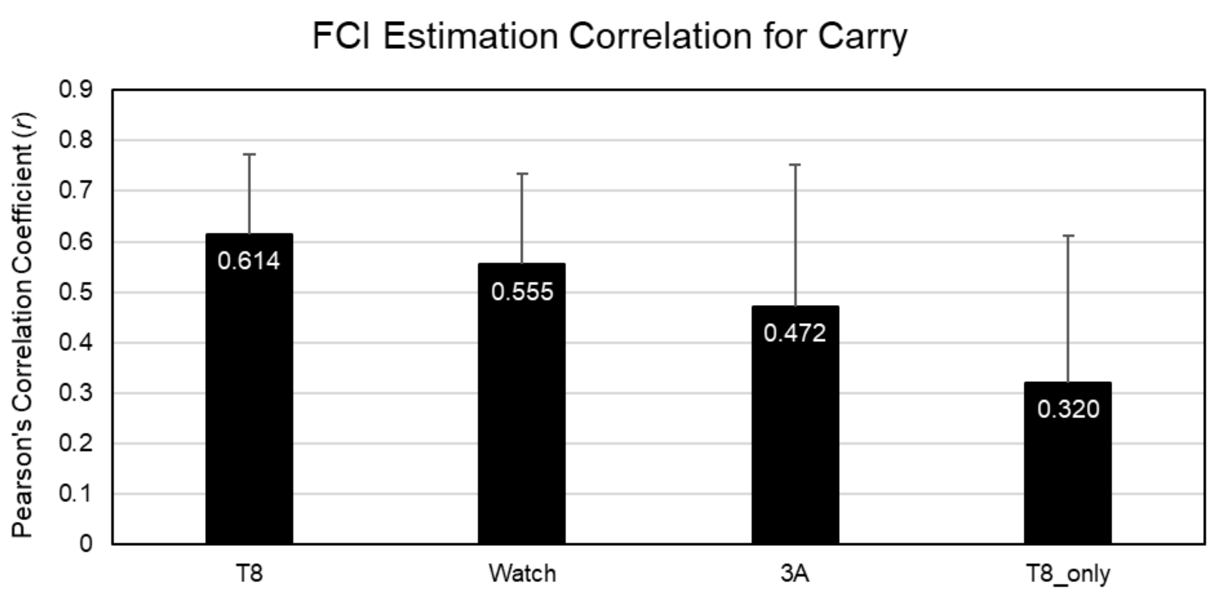


Fig. 6.26. Pearson's correlation coefficient (*r*) between the estimated and actual fatigue composite index (FCI) during the carry task when averaged across all folds. The error bars indicate the standard deviation.

Table 6.11. Training and prediction times of the artificial neural network for the carry task when averaged across all folds, for all sensor combinations. Values are presented as mean (standard deviation).

Sensor combination	Training time (s)	Prediction time (s)
T8	1271.65 (1431.14)	3.11 (0.85)
Watch	605.27 (549.42)	1.60 (0.38)
3A	1126.20 (983.17)	3.52 (0.95)
T8_only	805.74 (480.03)	3.07 (0.82)

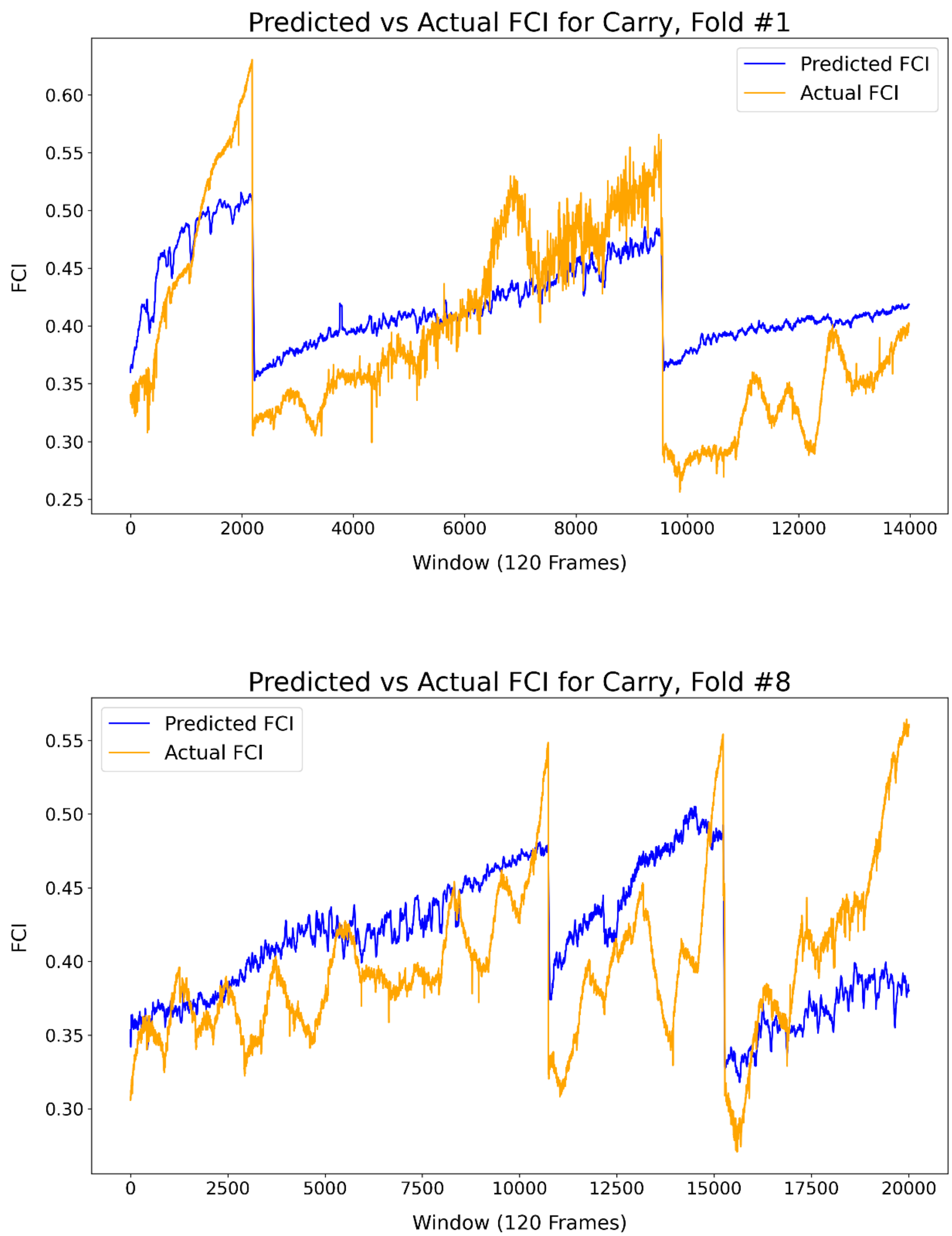


Fig. 6.27. The predicted versus actual fatigue composite index (FCI) for the T8 sensor combination during the carry task. The best (fold 1, MAE = 0.049, r = 0.89) and worst (fold 8, MAE = 0.045, r = 0.40) folds are pictured at the top and bottom, respectively.

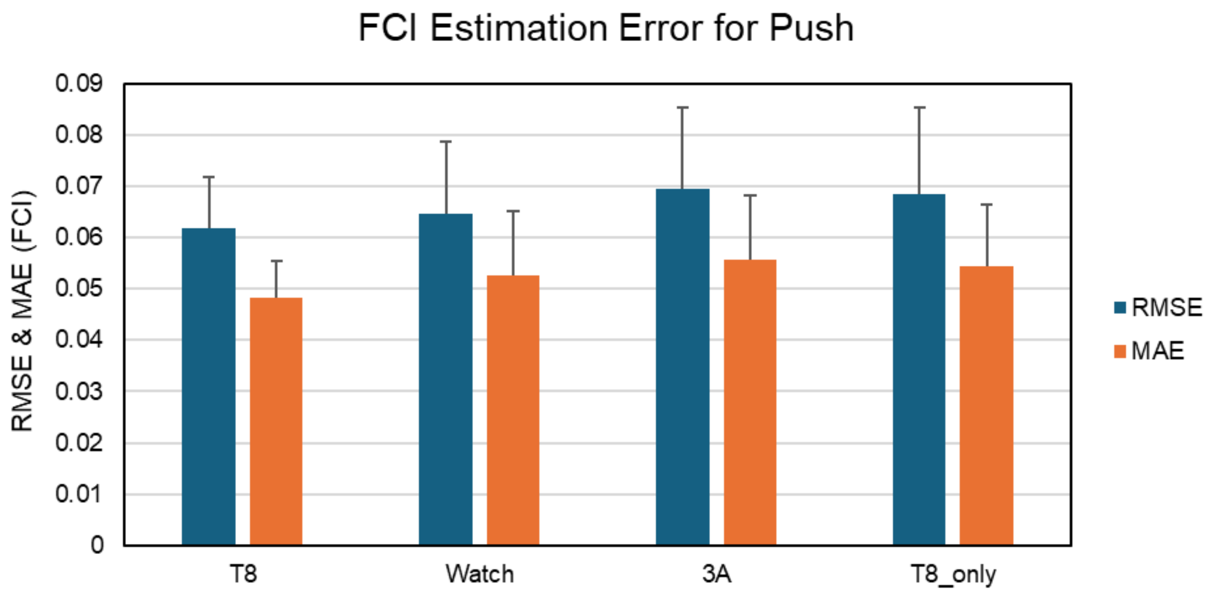


Fig. 6.28. The root mean squared error (RMSE) and mean absolute error (MAE) for the estimations of the fatigue composite index (FCI) during the push task when averaged across all folds. The error bars indicate the standard deviation.

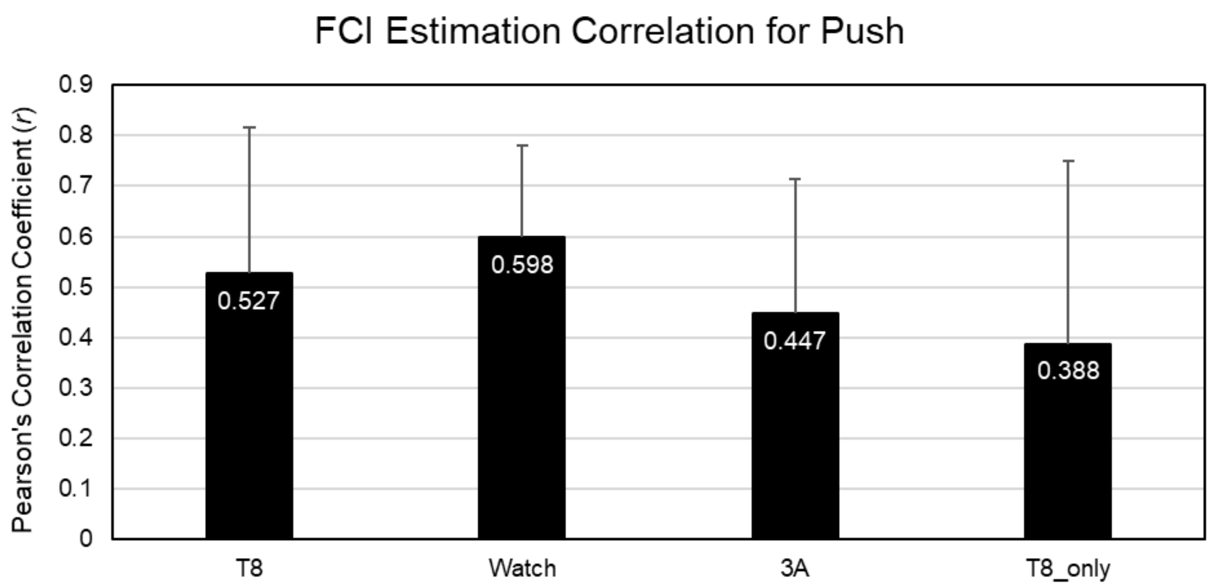


Fig. 6.29. Pearson's correlation coefficient (*r*) between the estimated and actual fatigue composite index (FCI) during the push task when averaged across all folds. The error bars indicate the standard deviation.

Table 6.12. Training and prediction times of the artificial neural network for the push task when averaged across all folds, for all sensor combinations. Values are presented as mean (standard deviation).

Sensor combination	Training time (s)	Prediction time (s)
T8	537.80 (451.36)	1.41 (0.43)
Watch	364.19 (217.52)	0.73 (0.20)
3A	532.10 (664.60)	1.50 (0.51)
T8_only	328.20 (150.93)	1.40 (0.41)

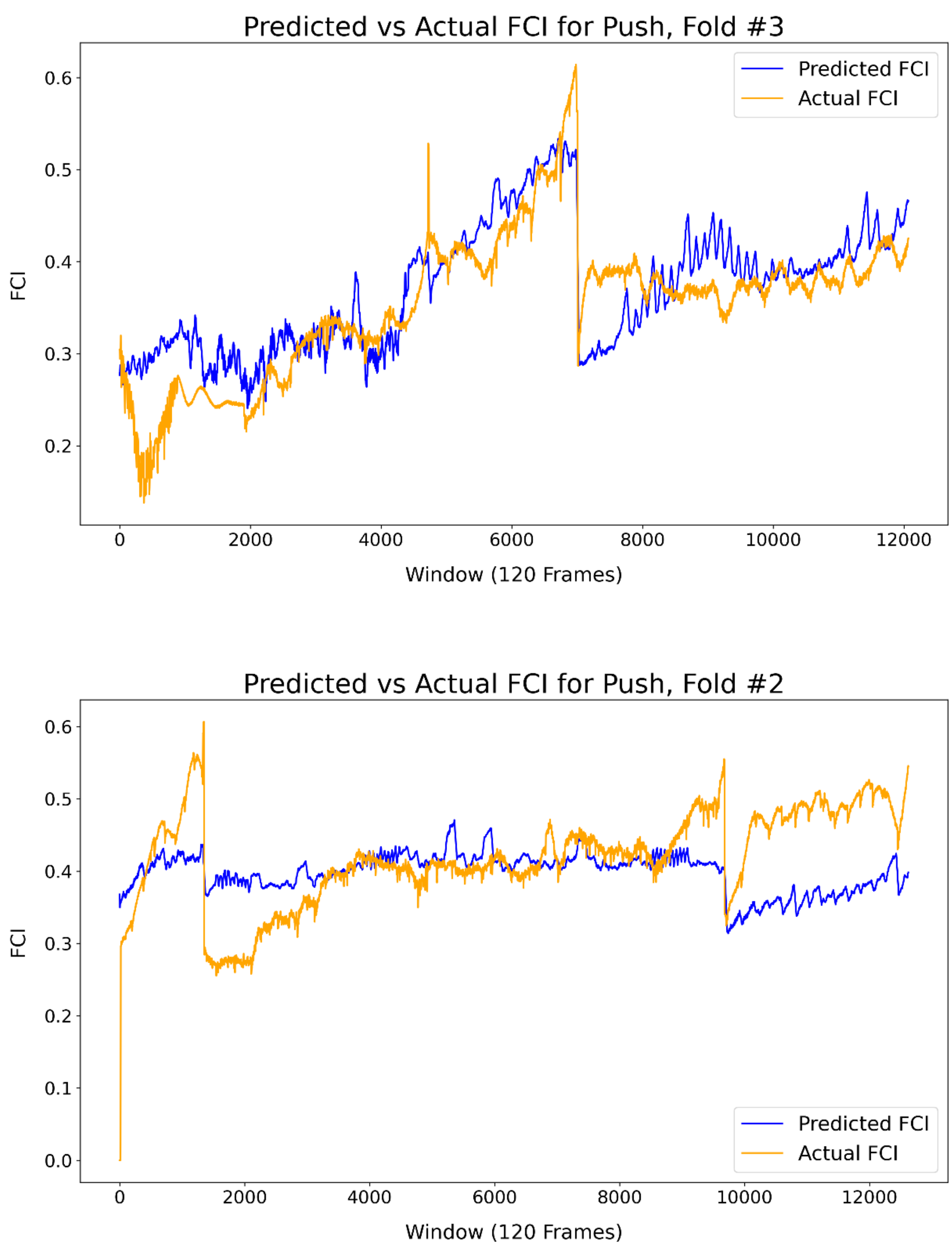


Fig. 6.30. The predicted versus actual fatigue composite index (FCI) for the T8 sensor combination during the push task. The best (fold 3, MAE = 0.047, r = 0.83) and worst (fold 2, MAE = 0.057, r = 0.052) folds are pictured at the top and bottom, respectively.

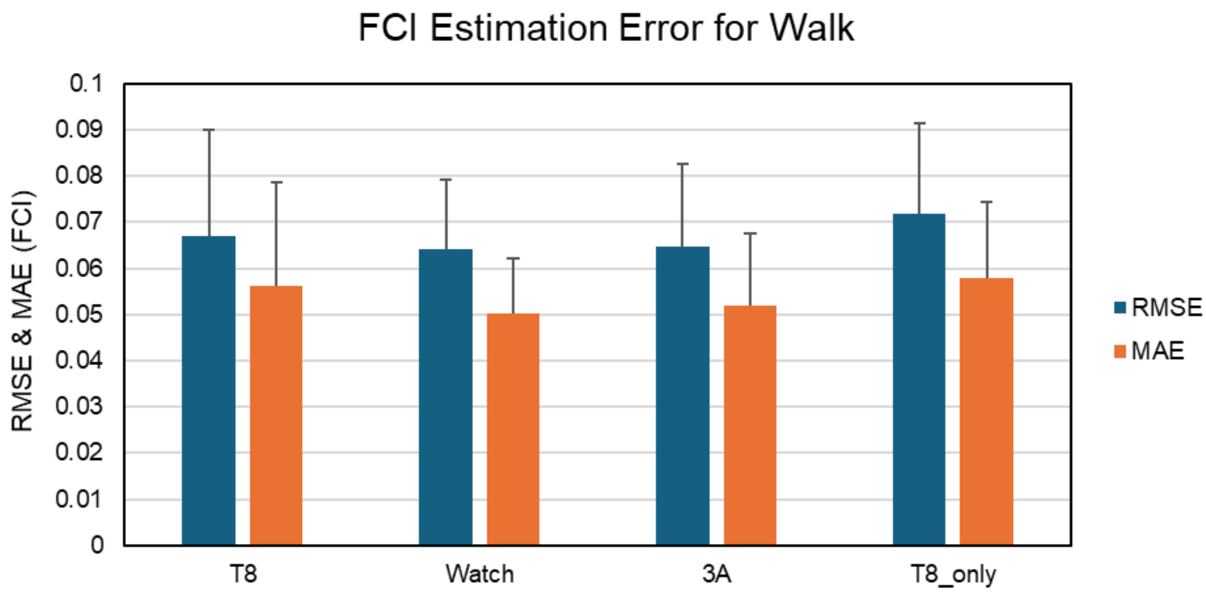


Fig. 6.31. The root mean squared error (RMSE) and mean absolute error (MAE) for the estimations of the fatigue composite index (FCI) during the walk task when averaged across all folds. The error bars indicate the standard deviation.

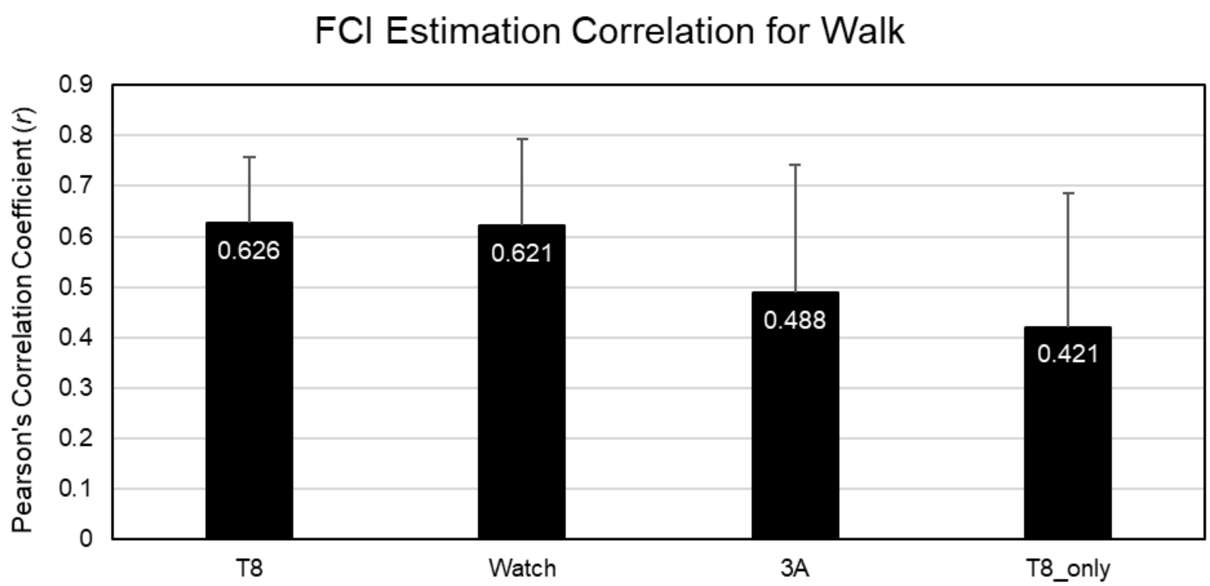
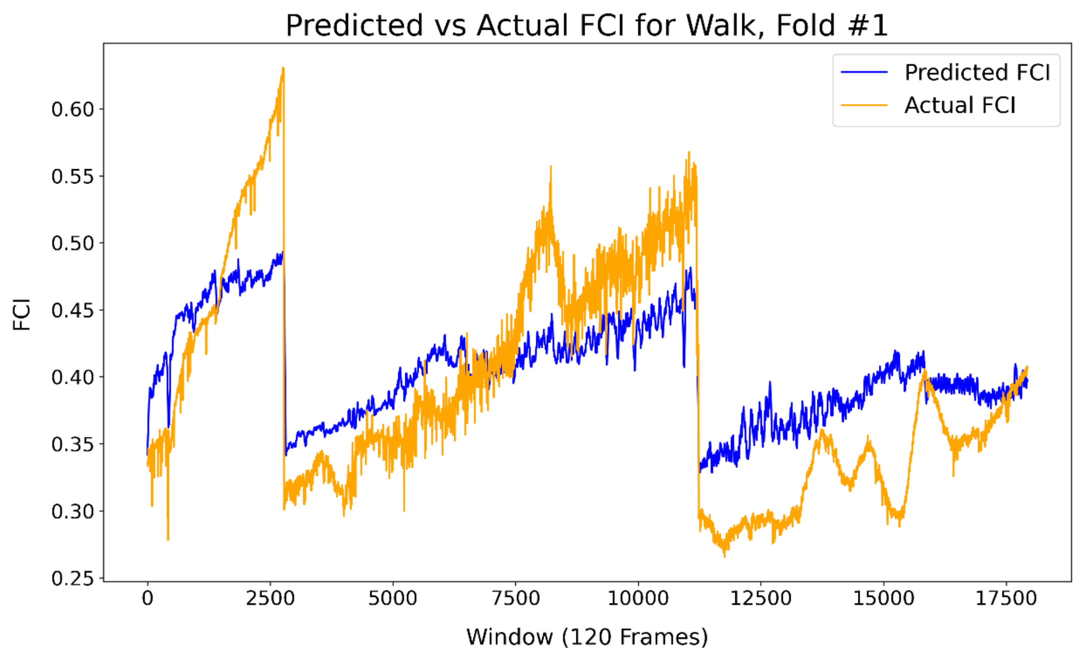


Fig. 6.32. Pearson's correlation coefficient (*r*) between the estimated and actual fatigue composite index (FCI) during the walk task when averaged across all folds. The error bars indicate the standard deviation.

Table 6.13. Training and prediction times of the artificial neural network for the walk task when averaged across all folds, for all sensor combinations. Values are presented as mean (standard deviation).

Sensor combination	Training time (s)	Prediction time (s)
T8	1378.76 (1090.54)	3.67 (1.03)
Watch	1030.05 (932.49)	1.88 (0.46)
3A	3325.83 (3719.16)	4.17 (1.18)
T8_only	1405.51 (1216.60)	3.61 (0.96)



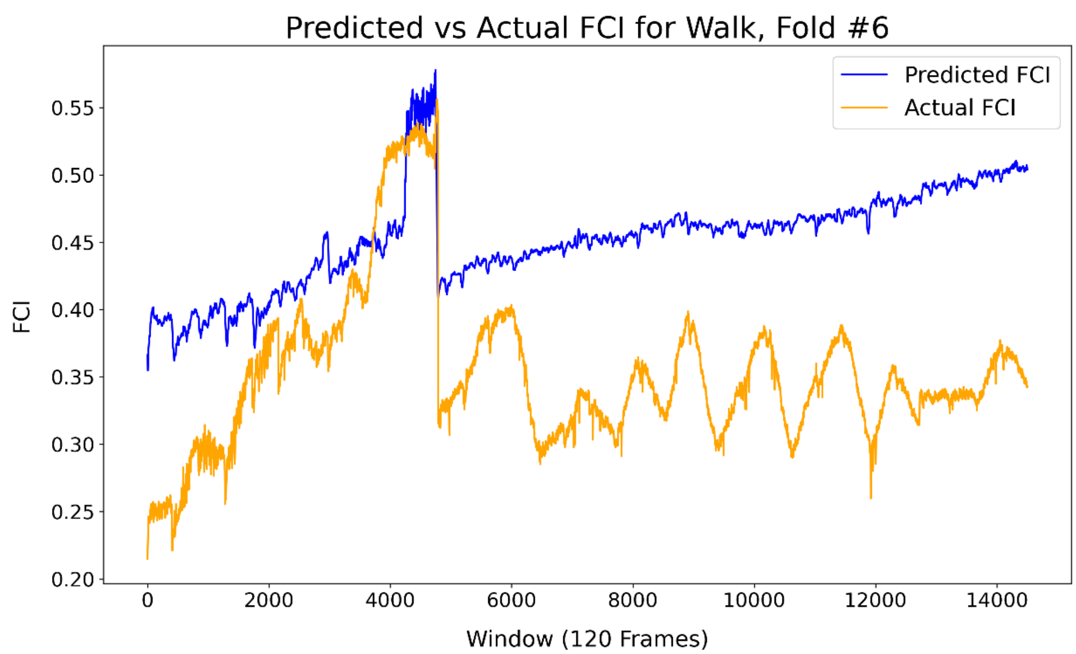


Fig. 6.33. The predicted versus actual fatigue composite index (FCI) for the T8 sensor combination during the walk task. The best (fold 1, MAE = 0.046, r = 0.88) and worst (fold 6, MAE = 0.10, r = 0.45) folds are pictured at the top and bottom, respectively.

6.4 Discussion

The objectives of this study were to develop a continuous FCI that provided a holistic representation of fatigue during a simulated manual handling protocol, then train and evaluate task-specific ANNs to estimate the FCI. The results showed that the FCI increased from the beginning to the end of the protocol for all 24 participants, where it increased more in those who stopped due to the stopping criteria than the four participants who voluntarily stopped. The task-specific ANNs successfully estimated FCI throughout the trials with and without HR data; the T8 combination (T8 IMU with HR) estimated the FCI with the lowest average error (RMSE = 0.06, MAE = 0.05) and the highest average correlation ($r = 0.59$). Furthermore, the FCI was estimated similarly well across all manual handling tasks. This work represents a component of a system that can adapt to the manual handling task being performed using HAR, then estimate holistic fatigue as a continuous variable in near real-time.

The variables comprising the FCI were selected because of their previous association with one or more components of fatiguability and their use in biomechanical research. The only variables that (visibly) changed in a manner consistent with literature were fatigue VAS, MLS, and HRinst. While many participants exhibited changes in LFHFratio, EMG MDF, and movement jerk at the individual level, no trends were observed on average. It was speculated that changes in the LFHFratio did not occur in response to fatigue because LFHFratio has been proposed to reflect balance between the sympathetic (low HR frequencies) and parasympathetic (high HR frequencies) (Pagani et al., 1986). While sympathetic activity may have increased due to the physical exertion involved, it may have been counter-balanced with increases in parasympathetic activity due to boredom and/or cognitive fatigue resulting from the repetitive experimental protocol (Pattyn et al., 2008). The lack of consistent changes in EMG MDF may be due to the fact that it was assessed during a gross, full-body strength assessment. Although participants were instructed to maintain their self-selected technique throughout the protocol to the best of their ability, it is possible that participants altered their muscle recruitment strategies and/or patterns to maintain force generation (Fedorowich et al., 2013). This may explain the initial (expected) leftward shift in EMG MDF across all seven muscles in the first 10% of protocol completion, and why the shift did not persist afterwards (Fig.

6.13). Lastly, consistent changes in pelvic jerk magnitude were not observed. Changes in jerk can be indicative of fatigue-related alterations in motor control (L. Zhang et al., 2019), a component of performance fatiguability (Enoka & Duchateau, 2016). However, due to the stopping criteria employed in this protocol, it is possible that participants exceeded a criterion that was minimally related to performance fatigability (e.g., perceived fatigability) and thus, stopping the protocol before motor control was measurably affected. Overall, the selected variables were still able to capture the increases in fatigue throughout the manual handling protocol when combined together in the FCI.

While the FCI increased throughout the protocol for all individuals, it stayed within a relatively small range, with a mean FCI difference of 0.21 between the protocol start and end. Conceptionally, the tight range observed in this holistic index is not surprising because the current experimental protocol ceased for any single stopping criteria, where the criteria were associated with different aspects of fatigability. Since the FCI aimed to capture multiple components of fatigability and most participants stopped because of a single criterion only, it is probable that other variables comprising the FCI were low(er) at the time of stopping. This notion was supported by the fact that participants who stopped due to two criteria (i.e., excessive fatigue VAS and MLS decline) had a greater mean FCI difference of 0.28. Thus, the tight range observed in this work provides confidence in the fact that fatigue was captured holistically. While other useful fatigue indices have been presented in literature, they tend to (thoroughly) capture a single component of fatigability. These include: the neurological fatigue index to measure perceived fatiguability in persons with multiple sclerosis (Aldughmi et al., 2017); a fatigue index to measure change in repeated sprint performance (González-Frutos et al., 2021); a dynamic fatigue index that incorporates maximal voluntary contractile force, current exertable maximum force, and external muscle load to model muscle fatigue during manual handling (Ma et al., 2008, 2010); a cumulative fatigue and risk index that incorporates work pattern, duty timing, and job type/breaks for manual work (Folkard et al., 2007); and the NASA task load index that measures the subjective workload experience by using variables that represent mental, physical, temporal demands, frustration, effort, and performance (Hart, 2006). Thus, the composite index presented in the current work is novel and useful because numerous, relevant components of fatiguability are measured, and their time-scale allow them to be suitable for near real-time monitoring of fatigue progression.

The FCI was estimated with satisfactory performance in this study. While the Pearson's correlations were considered moderate on average, strong correlations (as high as $r = 0.94$) were achieved for certain fold × task combinations. Further, the MAEs were low overall, with very small errors on numerous fold × task combinations (RMSE < 0.04, MAE < 0.03). The T8 was the best combination tested and estimated the FCI with similar performance across all tasks, which has been found to be an effective sensor location for classifying gross, whole-body movements such as manual handling tasks (Chapter 5; Chan & Graham, unpublished). However, the Watch combination was marginally better than T8 for the push and stand tasks. This finding may have been explained by the increased relative importance of upper limb movement to these tasks (Gutnik et al., 2005; Rancourt & Hogan, 2001), that could only be captured using upper limb kinematics. That being said, both sensor locations are suitable candidates for implementing this fatigue estimation approach using a single device: chest HR monitors that are mounted at the level of the T8 vertebrae are available with an IMU (e.g., Garmin HRM-Pro Plus), while smartwatches often include both IMU and PPG sensors (e.g., Apple Watch). Overall, as the majority of related work has aimed to classify discrete fatigue states, the correlations and errors on the present regression problems can be perceived as promising.

Relative to other works that have estimated fatigue as a continuous variable (typically quantified using self-reported fatigue only), the ANNs in this study performed reasonably well. For example, using HR reserve and critical power, perceived whole body fatigue was estimated with MAE = 12.8% and Spearman's correlation of $\rho = 0.83$ for construction work (G. Lee et al., 2024). Continuous fatigue was estimated during squats using whole-body kinematics, achieving RMSE = 0.54 on a fatigue scale ranging from 1 – 6 (Karg et al., 2014), while rating of perceived exertion (RPE) was estimated using an HR monitor and four IMUs during part assembly, supply pick up and insertion, and manual material handling with an MAE = 2.16 on a Borg 6 – 20 RPE scale (Maman et al., 2017). In an athletic context, self-reported fatigue was estimated during treadmill running using eight textile strain sensors with a RMSE = 0.06 when estimating the Borg 6-20 RPE scale (Gholami et al., 2020). Thus, the estimation performance of the present approach can likely still be improved, through methods such as obtaining more training samples. An important, additional strength of the present work is that when integrated with HAR models, a system can be developed to adapt to the task being performed prior to estimating continuous, holistic fatigue. These analyses are feasible for near real-time monitoring due the speed of the ANN estimations.

This study was presented with several limitations. First, this work was not developed in a real-world environment, thus potentially limiting the ecological validity of the estimation models. However, a simulated task was selected to ensure a large number of training samples could be present for each task. Furthermore, the data types required to develop the FCI are difficult to employ outside of laboratory-based collections (e.g., EMG, MLS assessments, etc.), warranting a simulated protocol. Regardless, further work is required to implement these FCI estimation models in the workplace. Another limitation is the minimal amount of information on the psychological components of fatigue included in the FCI (e.g., arousal, motivation, wakefulness; Enoka & Duchateau, 2016); in the present work, psychological state could only be (potentially) captured using the LFHFratio and the fatigue VAS. As a calibration procedure was not included when using the fatigue VAS (Abdel-Malek et al., 2021), it is unknown if and how much participants reflected on their psychological state throughout. Future work involving the FCI should include other outcome measures to directly include the psychological component of fatiguability, such as the use of electroencephalography (Tanaka et al., 2012).

6.5 Conclusion

Fatigue can increase WMSD and occupational accident risk, but if successfully monitored, preventative actions can be taken to mitigate this risk. The use of kinematics measured using IMU(s) with and without HR was used to estimate fatigue as a continuous variable in this study, quantified holistically using a composite index (FCI) of six variables previously associated with fatigue. Task-specific ANNs estimated the FCI with small errors and moderate-strength correlations during a simulated manual handling protocol with six tasks. Future development integrating this work with human activity recognition models will lead to a system that adapts to the task being performed, then estimates holistic fatigue level in near real-time, allowing for the provision of preventative feedback to mitigate WMSD and accident risk.

Chapter 7. General Discussion

Work-related musculoskeletal disorders (WMSDs) continue to be a major cause of occupational injury globally, resulting in significant health and economic consequences. Causing over $54 billion in annual costs in the United States alone (Karg et al., 2014), improving primary prevention efforts for WMSDs is a critical priority. An important framework by Van der Beek et al. (2017) outlined the sequence of WMSD prevention, comprised of six steps: 1) quantify the incidence of WMSD; 2) identify risk factors for WMSD; 3) determine the underlying mechanisms; 4) develop interventions that target known risk factors and mechanisms; 5) evaluate the intervention effectiveness; and 6) implement the intervention while considering the process, cost, and fidelity. The recent surge in computing power and improvements to the ease of collecting large amounts of data have created a boom in the use of machine learning (ML), which have been successfully employed to make data-driven decisions in a variety of health fields because of the wide range of applications: unsupervised learning for exploratory or cluster analyses, supervised learning for classifying or performing regression on unseen data, and reinforcement learning to determine optimal action sequences. Perhaps not surprisingly, ML has begun to play an increasingly large role for the primary prevention of WMSDs (Victor C.H. Chan et al., 2022), with most applications in step 4 of the framework. Of the 48 studies identified for intervention development, a notable application was fatigue modelling, especially because fatigue is a known risk factor that can increase the risk for occupational accidents in addition to WMSDs (M. F. Antwi-Afari et al., 2017; Namian et al., 2021).

The ability to model fatigue status during work in near real-time would allow for the optimization of work-rest ratios through the provision of timely breaks and job rotations (Padula et al., 2017) in addition to providing feedback on movement technique (Owlia et al., 2020). While numerous approaches to fatigue modelling have been proposed, this area of work is still in its early stages because existing approaches still have shortcomings. Some limitations across previous work include the dichotomization of fatigue into discrete states (typically one), the reliance on self-reported fatigue ratings as the ground truth variable for fatigue, the inability to adapt to various, non-cyclical tasks, and insufficient data for broader generalization and application. Building on these shortcomings could help progress towards a fatigue modelling system for manual handling work that would be effective at mitigating fatigue-related WMSD and occupation accident risk. Thus, this was a major objective of this dissertation.

The overarching aim of this doctoral dissertation was to mitigate WMSD and occupational accident risk by scoping the literature for target areas for the application of ML techniques, addressing knowledge gaps, and developing an effective intervention strategy by building on the shortcomings of previous fatigue modelling work. Thus, the objectives of each study were to:

1) Conduct a scoping review to determine how ML techniques have been used for primary WMSD prevention in the steps outlined by van der Beek et al. (2017).

   Secondary objective: Identify areas where ML can be effectively leveraged to mitigate WMSD and occupational accident risk.

2) Develop ML models to utilize data from wearable IMUs to automatically recognize manual handling tasks.

   Secondary objective: Compare the performance and training time between models driven by raw IMU signals versus features derived using sensor fusion processing.

3) Perform sensitivity analyses on the quantity and placement of IMUs for automatic recognition of six MMH tasks.

   Secondary objective: Compare the performance of using a shallower (3 hidden layers) versus a deeper (6 hidden layers) artificial neural network on HAR performance.

4) Develop task-specific ML models to estimate continuous fatigue levels during the performance of the six MMH tasks.

   Secondary objective: Develop a fatigue composite index (FCI) to represent a holistic ground truth of fatigue.

These objectives were achieved, and their findings were summarized in subchapter 7.1.

7.1 Summary of Findings

In study 1 (Chapter 3), a scoping review was conducted to determine the role of ML in the primary prevention of WMSD. A thorough literature search strategy was developed and returned 4,639 studies for screening, 130 of which were included in the final analysis. The papers were summarized and organized into the steps outlined by van der Beek et al. (2017). Since 1990, there had been a large increase in the application of ML for WMSD primary prevention, with 23.8% of all studies being published in the 12-months prior to the search conclusion in June 2021. Studies were primarily published by American authors (45.4%), and ANNs were the most common ML technique employed (used in 43.1% of included studies). Supervised machine learning was the most common ML technique employed, where it was most commonly used for classification (59.2% of studies) followed by regression (26.2% of studies). In the steps outlined by van der Beek et al. (2017), ML contributed to each step in the following manner:

1) Incidence and severity of WMSD (14 studies): primarily for autocoding worker compensation claims.
2) Risk factors for WMSD (33 studies): cluster analyses to retrospectively identify subgroups and risk factors; identify physical WMSD risk factor magnitude; classify jobs as low- or high-risk for WMSDs.
3) Underlying mechanisms (29 studies): augment existing biomechanical methods for studying WMSD etiology; identify patterns associated with WMSD risk factors; estimate biomechanical variables without modelling.
4) Development of interventions (48 studies): enable (semi)automated systems to monitor WMSD risk by classifying posture and the time spent in them, risky tasks and the time spent performing them, estimating inputs for ergonomic assessments, and modelling (classifying or estimating) fatigue status; pre-employment screening to assess worker risk.
5) Evaluation of interventions (6 studies): assessing the usability and accessibility of interventions with user feedback; quantifying changes in risk factors post-intervention.
6) Implementation of effective interventions (0 studies): None.

Overall, the contributions of ML were documented and areas for further research, development, and innovation to mitigate WMSD risk were identified.

For studies 2 – 4, a fatiguing manual handling protocol was designed and conducted to collect the required data to develop a fatigue modelling system. The manual handling protocol was comprised of six tasks (pulling, lifting, carrying, and pushing a crate, followed by unloaded walking to the start and occasional standing) was performed by 24 participants with self-selected technique. Each 5 min set was comprised of 20 rounds of the tasks, with fatigue assessments between each set: a fatigue visual analog scale (VAS) and a maximal force generation capacity assessment in the form of an isometric deadlift (MLS). Participants performed manual handling sets followed by fatigue assessment back and forth without breaks this until their fatigue VAS became ≥ 9 cm, their MLS became < 70% of their peak force, they could not maintain pace for 3 consecutive rounds, or they wished to stop. During this experimental protocol, kinematics (using an Xsens inertial motion capture suit and an Apple Watch), muscle activation (using electromyography; EMG), and heart activity (using an electrocardiogram; EKG) were recorded.

In studies 2 and 3 (Chapters 4 and 5, respectively), human activity recognition (HAR) models for fatiguing, manual handling tasks were developed using ML techniques and kinematic data from the Xsens IMUs. Study 2 developed two fully connected neural networks (FC) for comparison: one trained using raw 3D linear acceleration, angular velocity, and magnetometer data (RAW) and another trained using data processed with sensor fusion to obtain orientation and position data (FUSION). Marginal yet significantly better performances ($p < 0.01$) were observed when using RAW (accuracy = 94.93%) than FUSION (accuracy = 94.39%), with no differences in training time. Study 3 utilized these results and used the RAW dataset to develop 36 HAR models: 18 sensor combinations each for fully connected (FC) and DeepConvLSTM neural networks. While a greater number of IMUs to lead to greater classification performance (e.g., accuracy = 94.26% with six IMUs), stand out performances were observed with just the T8 (accuracy = 93.17%) or watch (accuracy = 91.61%) IMUs, or three on the T8 and both feet (3A; accuracy = 94.03%). Marginally better performances were observed with the DeepConvLSTM compared to the FC until 3 or more IMUs were involved, after which the performance of the FC was equal or better.

In study 4 (Chapter 6), a fatigue composite index (FCI) was computed for all participants by combining normalized fatigue VAS, MLS, instantaneous heart rate (HRinst), HR variability (LFHFratio), EMG median frequency (MDF) of seven muscles, and averaged 3D pelvic jerk. While not all individual variables changed throughout the protocol, the FCI increased throughout the protocol for all participants. Task-specific FC models were trained to estimate the FCI using input data from the notable sensor combinations from study 3: T8, Watch, 3A (with HR data), and T8 (without HR data). This resulted in 24 models trained and evaluated using the optimized sliding window sizes from study 3 using a *k*-fold cross-validation procedure. While all sensor combinations achieved satisfactory performance, the FC trained using T8 with HR estimated the FCI marginally better than the other combinations across the tasks, averaging small errors (RMSE = 0.063, MAE = 5.0%) and moderate correlation (r = 0.58) for its estimations.

7.2 Limitations

An important limitation that persists throughout this dissertation is that the hyperparameters and model architectures searched for ML in studies 2 - 4 are not nearly exhaustive. This is because even within predefined hyperparameter search spaces, a large number of possible permutations still existed; for example, 768 permutations existed per sensor combination (18 total) for the DeepConvLSTM model (one of two) in study 3 (Table 8). When considering there exist infinite values for each hyperparameter and vastly different model complexities (Nematzadeh et al., 2022), it is highly unlikely that the hyperparameters and model architectures were perfectly optimized for these problems (i.e., performances across studies 2 – 4 may still be improved). However, when training times can exceed 200 min per trial (Table 12) for a single model and sensor combination, decisions to ensure feasibility were warranted – such as the use of Bayesian hyperparameter optimization instead of exhaustive grid search, limiting the maximum number of trials executed per model, defining smaller search spaces based on previous research, and in some cases, not performing hyperparameter optimization at all for a specific permutation of model and sensor combination. This rationale extends to the number and types of model architecture employed. Despite this, however, the results in this dissertation are still worthwhile because: 1) hyperparameter optimization was performed; 2) the definition of the hyperparameter search spaces were guided by previous research with similar ML problems; 3) model architectures were selected based previous research; and 4) the classification and estimation performances are in-line with other related works. Thus, it is argued that these decisions were justified and do not negatively impact the importance and impact of the present work.

Another limitation that permeates through studies 2 - 4 is that the HAR and fatigue estimation models were not developed in real-world contexts, thus limiting the generalizability of the current models. The decision to conduct a simulated manual handling protocol in a laboratory were based on two primary reasons. First, it allowed for control over the tasks being performed. This ensured that a large number of all tasks were performed, and in succession without breaks to ensure an increase in fatigue. Depending on the work demands, certain tasks of interest may have been over- or under-performed if the data were collected in real-world environments, and breaks may have been taken to avoid fatigue. Furthermore, it would have been substantially more difficult to label the tasks being performed (i.e., for developing the HAR models). As the above reasons could have negatively affected the development of HAR and fatigue estimation models, having control over the protocol ensured that the necessary data could be successfully collected with sufficient quality. The second reason was that the data types that comprised the FCI are difficult to measure in the field, such as EMG and MLS. Thus, to develop the FCI, a laboratory-based collection was warranted. Overall, the decision to conduct the experimental protocol in the laboratory allowed for the HAR and fatigue estimation models to be developed successfully, and prove that this approach to adaptive, continuous fatigue estimation can be effective. Beyond ideal scenarios, this work must move past this limitation and be translated for use in the field – an important follow-up to this dissertation for future research and development.

7.3 Future directions for research and development

Because this work has garnered a large dataset, numerous opportunities exist to leverage the data to further knowledge on the manifestations of fatigue. One important element that has not been touched on in this dissertation is why the neural networks predicted a certain task (i.e., during HAR) and estimated a fatigue level (i.e., for fatigue estimation), which is referred to as model explainability (Carvalho et al., 2019).

Understanding this aspect would provide insight, not only into the models' behaviours, but also contribute to the body of knowledge on the kinematic manifestations of fatigue. Specifically, techniques such as the Shapley Additive Explanation (SHAP) can aid in quantifying the contribution and/or importance of each input feature to the prediction or estimation (Nohara et al., 2022). Future work should implement such techniques to explain model behaviour and quantify the importance of the features.

The envisioned use of this system was to unobtrusively monitor workers' fatigue using kinematic data (with or without HR data), employing HAR models to classify tasks, and estimating continuous fatigue levels. Then, if the rate of fatigue increase was too large and/or if the fatigue level was excessive for an extended period of time, a configurable alert could be triggered to the worker and/or employer to inform breaks, rotations, and provide technique reminders. An example of how an alert could look on an Apple Watch is depicted in Fig. 7.1. In order to arrive at real-world implementation, various procedures must still be performed: cascading HAR and regression models together in an ensemble, model deployment, front-end software development, and the selection of notification thresholds.

Fig. 7.1. An example of how a fatigue notification for a worker triggered by the envisioned system could look on an Apple Watch.

Cascading ensembles refers to the “stacking” of numerous ML models together, to provide outputs not possible using isolated models (García-Pedrajas et al., 2005). Though various configurations are possible, a suitable configuration for the models developed in the present dissertation is proposed in Fig. 7.2. Kinematic (with or without HR) data will be the input with a constant sliding window size (thus making it imperative to use the sliding window sizes optimized in study 3 for FCI estimation), the HAR model will classify the task being performed, the output of the HAR model will inform which task-specific model to employ for fatigue estimation, then the task-specific regression model will estimate the FCI value, representing holistic fatigue level. An example of this integration estimating the FCI on one participant is presented in Fig. 7.3, showcasing the results on participant 5. Using only the T8 IMU, the HAR model classified the task being performed with and accuracy of 93.88% and a weighted-average F1-score of 94.16%. Using these classifications to inform which task-specific model to employ (from Chapter 6), the FCI was estimated using the T8 data supplemented with HR with MAE = 0.048 and a Pearson’s correlation of $r = 0.88$ across all tasks during their manual handling protocol.

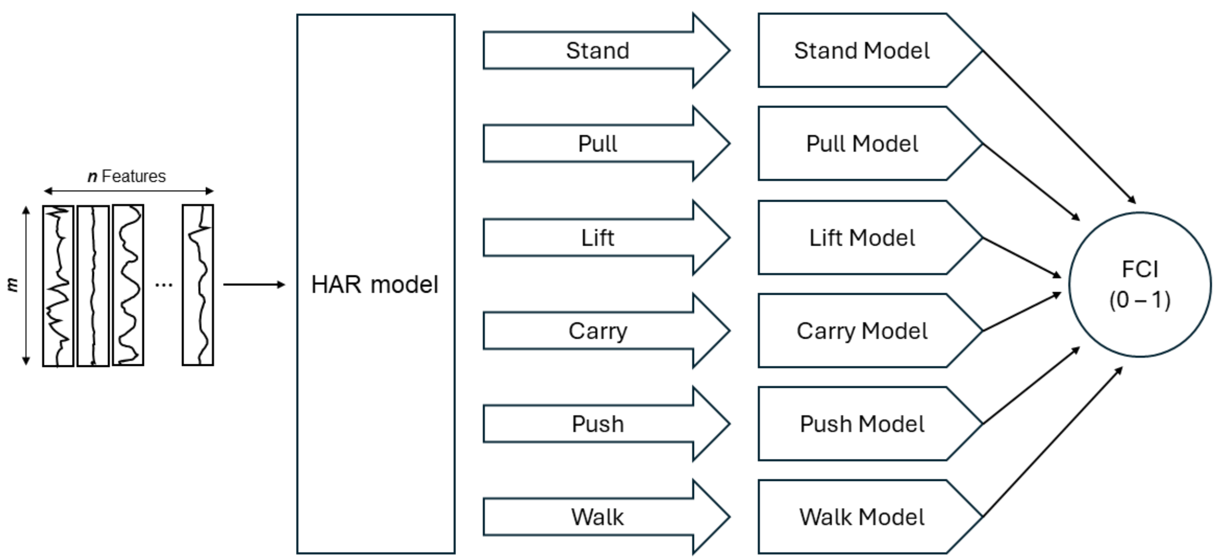


Fig. 7.2. An example of how the models developed in this dissertation could be organized in a cascading ensemble. The input data will have the dimensions equal to *n* number of features × *m* sliding window size. The human activity recognition (HAR) model will classify the windows as being one of six manual handling tasks, then the corresponding regression model will estimate the fatigue composite index (FCI) as a continuous variable ranging from 0 – 1.

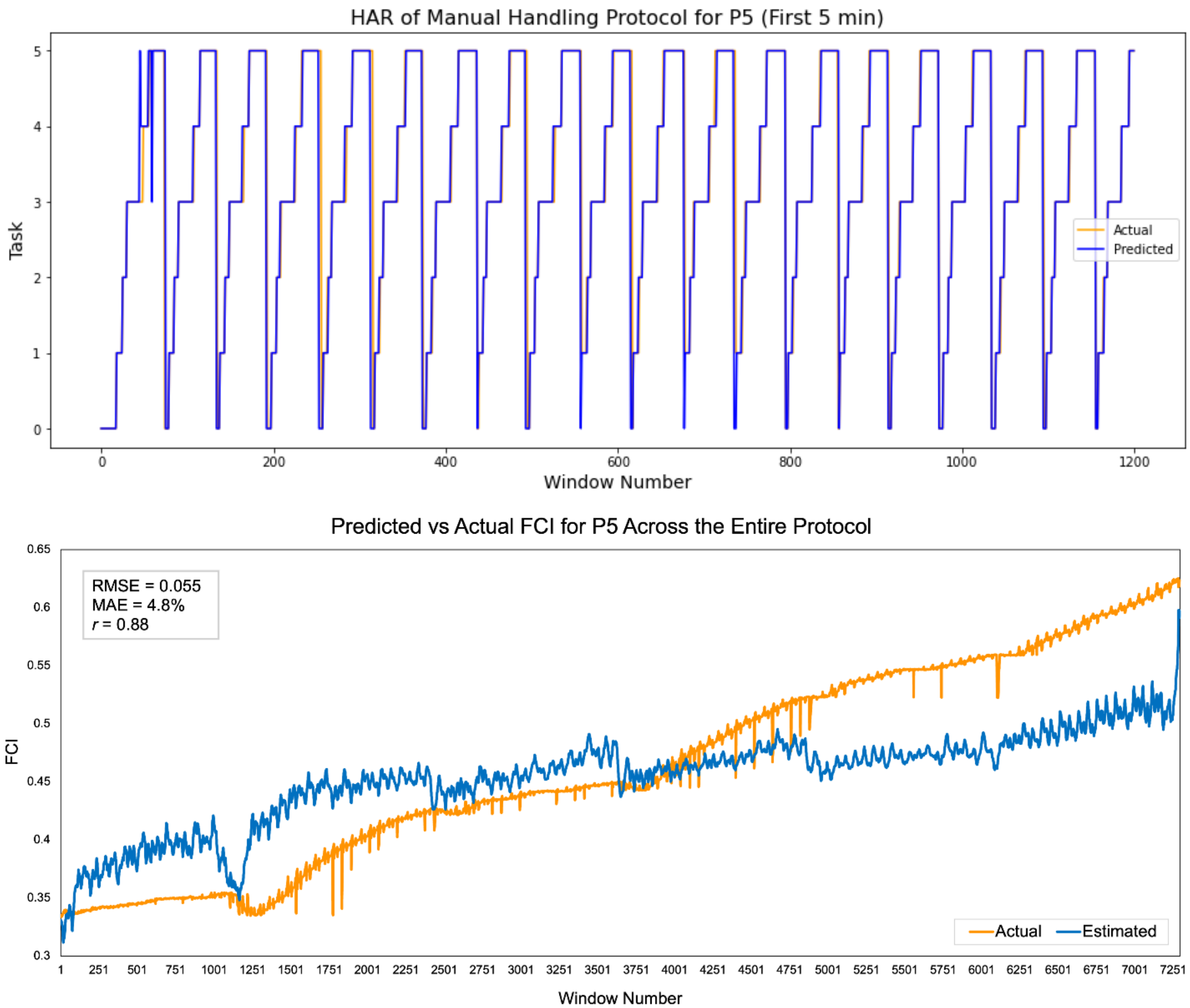


Fig. 7.3. An example of the human activity recognition (HAR; top) and task-specific fatigue composite index (FCI) estimation models working together to estimate the FCI across the entire manual handling protocol for one participant (P5; bottom). The first 5 min of the manual handling protocol HAR performance is depicted, where an accuracy = 93.88% was achieved for the entire protocol. The model performing the FCI estimation for each window (bottom) was dependent on the outputs of the the HAR. Task 0 = stand; 1 = pull; 2 = lift; 3 = carry; 4 = push; 5 = walk; RMSE = root mean squared error; MAE = mean absolute error.

In the context of deploying ML models for on-device mobile inference, model compression is required and refers to the packaging and compression of ML models ensure they are deployable to mobile devices. In short, various procedures are implemented to reduce the size and computational demands of ML models through various methods, such as quantization (reducing numbers' significant figures used in the models), pruning (identifying and removing less important nodes from the model), and weight sharing (identifying parameters that can be shared across the model; Bursa et al., 2023; Manor & Greenberg, 2022). These procedures can be employed using an open-source runtime, such as Google LiteRT (formerly TensorFlow Lite; Alphabet Inc., Mountain View, USA).

Front-end software development refers to the design and creation of a graphical user interface and integration with on-device models combined into software that can be used on mobile devices (e.g., laptops, smart phones, smart watches). This will allow users to interact with the application to setup this system (including pairing devices), configure alert settings, and to receive near real-time alerts from the system about fatigue status via push notifications. An example of how a push notification may appear on an Apple Watch was presented in Fig. 7.1. Overall, these aforementioned steps will allow the work from this dissertation to be employed in the field for fatigue estimation that adapts to the performance of various manual handling tasks. To widen the applicability of this work, future work should also aim to broaden the models by collecting more data.

Another important step for implementation is the determination of appropriate thresholds for triggering notifications. A benefit of estimating holistic fatigue as a continuous value is the ability to monitor excessive FCI values, in addition to large rates in increase. While it is possible to allow end users to determine their own thresholds, one or more default settings need to be defined based on training data. One approach for defining such thresholds would be to retroactively quantify biomechanical risk factors in the dataset, such as peak spine flexion velocity, a risk factor for low back disorders (Marras et al., 1995), or frontal plane knee motion, a risk factor for non-contact anterior cruciate ligament injuries (Boden et al., 2013). Then, the absolute FCI values and the rate of change of FCI values that are associated with the change of established risk factors into the high-risk categories can be noted and used as thresholds. Future work should empirically determine these thresholds using existing and newly collected data.

Many manual handling tasks exist beyond the ones presented, including overhead lifting, pushing and pulling, asymmetrical lifting, carrying, pulling, pushing, etc. To ensure the models can be more generalizable for real-world implementation, more data should be collected on the tasks not yet used in this dissertation to further expand the existing HAR model, while using these data to develop new task-specific

FCI estimation models. Additionally, collecting additional data on tasks already present in the current work will also be important, as it can still yield benefits in improving model performance and generalizability. Finally, this approach can be developed, in theory, for physical tasks specific to other fields as well, such as for construction (Kong et al., 2018) or strength & conditioning applications (Karg et al., 2014).

References

Aaronson, L. S., Teel, C. S., Cassmeyer, V., Neuberger, G. B., Pallikkathayil, L., Pierce, J., Press, A. N., Williams, P. D., & Wingate, A. (1999). Defining and Measuring Fatigue. *Image: The Journal of Nursing Scholarship*, *31*(1), 45–50. https://doi.org/10.1111/j.1547-5069.1999.tb00420.x

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., … Zheng, X. (2016). Tensorflow: A system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation*, *16*, 265–283. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi

Abdel-Malek, D., Foley, R. C. A., Wakeely, F., Graham, J. D., & Delfa, N. J. La. (2021). Calibrating ratings of perceived fatigue relative to objective measures of localised muscle fatigue using a feedback-based familiarisation protocol. *Theoretical Issues in Ergonomics Science*, *22*(5), 511–527.

Abobakr, A., Nahavandi, D., Hossny, M., Iskander, J., Attia, M., Nahavandi, S., & Smets, M. (2019). RGB-D ergonomic assessment system of adopted working postures. *Applied Ergonomics*, *80*(April 2018), 75–88. https://doi.org/10.1016/j.apergo.2019.05.004

Abubakar, A. M., Karadal, H., Bayighomog, S. W., & Merdan, E. (2020). Workplace injuries, safety climate and behaviors: application of an artificial neural network. *International Journal of Occupational Safety and Ergonomics*, *26*(4), 651–661. https://doi.org/10.1080/10803548.2018.1454635

Adam, S. P., Alexandropoulos, S.-A. N., Pardalos, P. M., & Vrahatis, M. N. (2019). No Free Lunch Theorem: A Review. In I. C. Demetriou & P. M. Pardalos (Eds.), *Approximation and Optimization* (pp. 57–82). Springer Nature Switzerland.

Adaskevicius, R. (2014). Method for recognition of the physical activity of human being using a wearable accelerometer. *Elektronika Ir Elektrotechnika*, *20*(5), 127–131. https://doi.org/10.5755/j01.eee.20.5.7113

Aghazadeh, F., Arjmand, N., & Nasrabadi, A. M. (2020). Coupled artificial neural networks to estimate 3D whole-body posture, lumbosacral moments, and spinal loads during load-handling activities. *Journal of Biomechanics*, *102*, 109332. https://doi.org/10.1016/j.jbiomech.2019.109332

Ahn, G., Hur, S., & Jung, M. C. (2020). Bayesian network model to diagnose WMSDs with working characteristics. *International Journal of Occupational Safety and Ergonomics*, *26*(2), 336–347. https://doi.org/10.1080/10803548.2018.1502131

Akanmu, A. A., Olayiwola, J., Ogunseiju, O., & McFeeters, D. (2020). Cyber-physical postural training system for construction workers. *Automation in Construction*, *117*(May), 103272. https://doi.org/10.1016/j.autcon.2020.103272

Akay, D. (2011). Grey relational analysis based on instance based learning approach for classification of risks of occupational low back disorders. *Safety Science*, *49*(8–9), 1277–1282. https://doi.org/10.1016/j.ssci.2011.04.018

Akhavian, R., & Behzadan, A. H. (2016). Smartphone-based construction workers' activity recognition and classification. *Automation in Construction, 71*, 198–209.

Akkas, O., Lee, C. H., Hu, Y. H., Harris Adamson, C., Rempel, D., & Radwin, R. G. (2017). Measuring exertion time, duty cycle and hand activity level for industrial tasks using computer vision. *Ergonomics*, *60*(12), 1730–1738. https://doi.org/10.1080/00140139.2017.1346208

Akkas, O., Lee, C. H., Hu, Y. H., Yen, T. Y., & Radwin, R. G. (2016). Measuring elemental time and duty cycle using automated video processing. *Ergonomics*, *59*(11), 1514–1525. https://doi.org/10.1080/00140139.2016.1146347

Aldughmi, M., Bruce, J., & Siengsukon, C. F. (2017). Relationship between fatigability and perceived fatigue measured using the neurological fatigue index in people with multiple sclerosis. *International Journal of MS Care*, *19*(5), 232–239. https://doi.org/10.7224/1537-2073.2016-059

Ali, M. H., Azmir, N. A., & Yahya, M. N. (2018). Prediction on hand arm vibration exposure cause-effect among grass-cutting workers in Malaysia. *International Journal of Engineering and Technology(UAE)*, *7*(3), 48–54. https://doi.org/10.14419/ijet.v7i3.24.17300

Alpaydin, E. (2020). *Introduction to Machine Learning* (4th ed.). The MIT Press.

Alwasel, A., Sabet, A., Nahangi, M., Haas, C. T., & Abdel-Rahman, E. (2017). Identifying poses of safe and productive masons using machine learning. *Automation in Construction*, *84*(October), 345–355. https://doi.org/10.1016/j.autcon.2017.09.022

Amann, M. (2012). Significance of Group III and IV muscle afferents for the endurance exercising human. *Clinical and Experimental Pharmacology and Physiology*, *39*(9), 831–835. https://doi.org/10.1111/j.1440-1681.2012.05681.x

Ament, W., & Verkerke, G. J. (2009). Exercise and Fatigue. *Sports Medicine*, *39*(5), 389–422. https://doi.org/10.1016/b978-1-4831-7818-9.50021-6

Antwi-Afari, M. F., Li, H., Edwards, D. J., Pärn, E. A., Seo, J., & Wong, A. Y. L. (2017). Biomechanical analysis of risk factors for work-related musculoskeletal disorders during repetitive lifting task in construction workers. *Automation in Construction*, *83*(July), 41–47. https://doi.org/10.1016/j.autcon.2017.07.007

Antwi-Afari, Maxwell Fordjour, Li, H., Umer, W., Yu, Y., & Xing, X. (2020). Construction Activity Recognition and Ergonomic Risk Assessment Using a Wearable Insole Pressure System. *Journal of Construction Engineering and Management*, *146*(7), 04020077. https://doi.org/10.1061/(asce)co.1943-7862.0001849

Antwi-Afari, Maxwell Fordjour, Li, H., Yu, Y., & Kong, L. (2018). Wearable insole pressure system for automated detection and classification of awkward working postures in construction workers. *Automation in Construction*, *96*(September), 433–441. https://doi.org/10.1016/j.autcon.2018.10.004

Arbelaez Garces, G., Rakotondranaivo, A., & Bonjour, E. (2016). An acceptability estimation and analysis methodology based on Bayesian networks. *International Journal of Industrial Ergonomics*, *53*, 245–256. https://doi.org/10.1016/j.ergon.2016.02.005

Arjmand, N., Ekrami, O., Shirazi-Adl, A., Plamondon, A., & Parnianpour, M. (2013). Relative performances of artificial neural network and regression mapping tools in evaluation of spinal loads and muscle forces during static lifting. *Journal of Biomechanics*, *46*(8), 1454–1462. https://doi.org/10.1016/j.jbiomech.2013.02.026

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology: Theory and Practice*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Armstrong, D. P., Ross, G. B., Graham, R. B., & Fischer, S. L. (2019). Considering movement competency within physical employment standards. *Work*, *63*(4), 603–613. https://doi.org/10.3233/WOR-192955

Aryal, A., Ghahramani, A., & Becerik-Gerber, B. (2017). Monitoring fatigue in construction workers using physiological measurements. *Automation in Construction*, *82*, 154–165. https://doi.org/10.1016/j.autcon.2017.03.003

Asadi, H., Zhou, G., Lee, J. J., Aggarwal, V., & Yu, D. (2020). A computer vision approach for classifying isometric grip force exertion levels. *Ergonomics*, *63*(8), 1010–1026. https://doi.org/10.1080/00140139.2020.1745898

Asensio-Cuesta, S., Diego-Mas, J. A., & Alcaide-Marzal, J. (2010). Applying generalised feedforward neural networks to classifying industrial jobs in terms of risk of low back disorders. *International Journal of Industrial Ergonomics*, *40*(6), 629–635. https://doi.org/10.1016/j.ergon.2010.04.007

Asgari, N., Sanjari, M. A., & Esteki, A. (2017). Local dynamic stability of the spine and its coordinated lower joints during repetitive lifting: effects of fatigue and chronic low back pain. *Human Movement Science*, *54*, 339–346. https://doi.org/10.1016/j.humov.2017.06.007

Atal, M. K., Palei, S. K., Chaudhary, D. K., Kumar, V., & Karmakar, N. C. (2020). Occupational exposure of dumper operators to whole-body vibration in opencast coal mines: an approach for risk assessment using a Bayesian network. *International Journal of Occupational Safety and Ergonomics*. https://doi.org/10.1080/10803548.2020.1828551

Baghdadi, A., Cavuoto, L. A., Jones-farmer, A., Rigdon, S. E., Esfahani, E. T., Megahed, F. M., Baghdadi, A., Cavuoto, L. A., Jones-farmer, A., & Rigdon, S. E. (2021). Monitoring worker fatigue using wearable devices : A case study to detect changes in gait parameters. *Journal of Quality Technology*, *53*(1), 47–71. https://doi.org/10.1080/00224065.2019.1640097

Baghdadi, A., Megahed, F. M., Esfahani, E. T., & Cavuoto, L. A. (2018). A machine learning approach to detect changes in gait parameters following a fatiguing occupational task. *Ergonomics*, *61*(8), 1116–1129. https://doi.org/10.1080/00140139.2018.1442936

Bagnall, A., Lines, J., Bostrom, A., Large, J., & Keogh, E. (2017). The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, *31*(3), 606–660. https://doi.org/10.1007/s10618-016-0483-9

Bances, E., Karol, A. M. A., & Schneider, U. (2022). LSTM and CNN Based IMU Sensor Fusion Approach for Human Pose Identification in Manual Handling Activities. *Wearable Robotics: Challenges and Trends: Proceedings of the 5th International Symposium on Wearable Robotics, WeRob2020, and of WearRAcon Europe 2020, October 13–16, 2020*, 461–465.

Bangaru, S. S., Wang, C., Busam, S. A., & Aghazadeh, F. (2021). ANN-based automated scaffold builder activity recognition through wearable EMG and IMU sensors. *Automation in Construction*, *126*(February), 103653. https://doi.org/10.1016/j.autcon.2021.103653

Barkallah, E., Freulard, J., Otis, M. J. D., Ngomo, S., Ayena, J. C., & Desrosiers, C. (2017). Wearable devices for classification of inadequate posture at work using neural networks. *Sensors (Switzerland)*, *17*(9), 1–24. https://doi.org/10.3390/s17092003

Barry, B. K., & Enoka, R. M. (2007). The neurobiology of muscle fatigue: 15 years later. *Integrative and Comparative Biology*, *47*(4), 465–473. https://doi.org/10.1093/icb/icm047

Barshan, B., & Yurtman, A. (2016). Investigating Inter-Subject and Inter-Activity Variations in Activity Recognition Using Wearable Motion Sensors. *Computer Journal*, *59*(9), 1345–1362. https://doi.org/10.1093/comjnl/bxv093

Bastani, K., Kim, S., Kong, Z. J., Nussbaum, M. A., & Huang, W. (2016). Online Classification and Sensor Selection Optimization with Applications to Human Material Handling Tasks Using Wearable Sensing Technologies. *IEEE Transactions on Human-Machine Systems*, *46*(4), 485–497.

https://doi.org/10.1109/THMS.2016.2537747

Bertke, S. J., Meyers, A. R., Wurzelbacher, S. J., Bell, J., Lampl, M. L., & Robins, D. (2012). Development and evaluation of a Naïve Bayesian model for coding causation of workers' compensation claims. *Journal of Safety Research*, *43*(5–6), 327–332. https://doi.org/10.1016/j.jsr.2012.10.012

Bertke, S. J., Meyers, A. R., Wurzelbacher, S. J., Measure, A., Lampl, M. P., & Robins, D. (2016). Comparison of methods for auto-coding causation of injury narratives. *Accident Analysis and Prevention*, *88*, 117–123. https://doi.org/10.1016/j.aap.2015.12.006

Bigland-Ritchie, B., Furbush, F., & Woods, J. J. (1986). Fatigue of intermittent submaximal voluntary contractions: Central and peripheral factors. *Journal of Applied Physiology*, *61*(2), 421–429. https://doi.org/10.1152/jappl.1986.61.2.421

Bigland Ritchie, B., Johansson, R., Lippold, O. C. J., & Woods, J. J. (1983). Contractile speed and EMG changes during fatigue of sustained maximal voluntary contractions. *Journal of Neurophysiology*, *50*(1), 313–324. https://doi.org/10.1152/jn.1983.50.1.313

Boden, B. P., Sheehan, F. T., Torg, J. S., & Hewett, T. E. (2013). Non-contact ACL Injuries: Mechanisms and Risk Factors. *The Journal of the American Academy of Orthopaedic Surgeons*, *18*(9), 520–527.

Bodin, J., Garlantézec, R., Costet, N., Descatha, A., Fouquet, N., Caroly, S., & Roquelaure, Y. (2017). Forms of work organization and associations with shoulder disorders: Results from a French working population. *Applied Ergonomics*, *59*, 1–10. https://doi.org/10.1016/j.apergo.2016.07.019

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. https://doi.org/10.1097/01.BRS.0000087500.70575.45

Bootsman, R., Markopoulos, P., Qi, Q., Wang, Q., & Timmermans, A. A. A. (2019). Wearable technology for posture monitoring at the workplace. *International Journal of Human-Computer Studies*, *132*, 99–111.

Bose, R., Wang, H., Dragomir, A., Thakor, N. V, Bezerianos, A., & Li, J. (2019). Regression-based continuous driving fatigue estimation: Toward practical implementation. *IEEE Transactions on Cognitive and Developmental Systems*, *12*(2), 323–331.

Bota, P., Silva, R., Carreiras, C., Fred, A., & da Silva, H. P. (2024). BioSPPy: A Python toolbox for physiological signal processing. In *SoftwareX* (Vol. 26, p. 101712). Elsevier.

Brooks, B. (2008). Shifting the focus of strategic occupational injury prevention. Mining free-text, workers compensation claims data. *Safety Science*, *46*(1), 1–21. https://doi.org/10.1016/j.ssci.2006.09.006

Brutsaert, T. D., Hernandez-Cordero, S., Rivera, J., Viola, T., Hughes, G., & Haas, J. D. (2003). Iron supplementation improves progressive fatigue resistance during dynamic knee extensor exercise in iron-depleted, nonanemic women. *American Journal of Clinical Nutrition*, *77*(2), 441–448. https://doi.org/10.1093/ajcn/77.2.441

Bursa, S. Ö., Durmaz İncel, Ö., & Işıklar Alptekin, G. (2023). Building lightweight deep learning models with TensorFlow Lite for human activity recognition on mobile devices. *Annals of Telecommunications*, *78*(11), 687–702.

Busto Serrano, N., Suárez Sánchez, A., Sánchez Lasheras, F., Iglesias-Rodríguez, F. J., & Fidalgo Valverde, G. (2020). Identification of gender differences in the factors influencing shoulders, neck and upper limb MSD by means of multivariate adaptive regression splines (MARS). *Applied Ergonomics*, *82*(October 2019). https://doi.org/10.1016/j.apergo.2019.102981

Carvalho, D. V, Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, *8*(8), 832.

Catai, A. M., Pastre, C. M., de Godoy, M. F., da Silva, E., de Medeiros Takahashi, A. C., & Vanderlei, L. C. M. (2020). Heart rate variability: are you using it properly? Standardisation checklist of procedures. *Brazilian Journal of Physical Therapy*, *24*(2), 91–102.

Chan, V C H, Beaudette, S. M., Smale, K. B., Beange, K. H. E., & Graham, R. B. (2020). A subject-specific approach to detect fatigue-related changes in spine motion using wearable sensors. *Sensors (Switzerland)*, *20*(9), 2646. https://doi.org/10.3390/s20092646

Chan, Victor C.H., Ross, G. B., Clouthier, A. L., Fischer, S. L., & Graham, R. B. (2022). The role of machine learning in the primary prevention of work-related musculoskeletal disorders: A scoping review. *Applied Ergonomics*, *98*(September 2021), 103574. https://doi.org/10.1016/j.apergo.2021.103574

Chen, C. L., Kaber, D. B., & Dempsey, P. G. (2000). A new approach to applying feedforward neural networks to the prediction of musculoskeletal disorder risk. *Applied Ergonomics*, *31*(3), 269–282. https://doi.org/10.1016/S0003-6870(99)00055-1

Chen, C. L., Kaber, D. B., & Dempsey, P. G. (2004). Using feedforward neural networks and forward selection of input variables for an ergonomics data classification problem. *Human Factors and Ergonomics In Manufacturing*, *14*(1), 31–49. https://doi.org/10.1002/hfm.10052

Chen, D., Cai, Y., Cui, J., Chen, J., Jiang, H., & Huang, M. C. (2018). Risk factors identification and visualization for work-related musculoskeletal disorders with wearable and connected gait analytics system and kinect skeleton models. *Smart Health*, *7–8*(June), 60–77. https://doi.org/10.1016/j.smhl.2018.05.003

Chen, K., Zhang, D., Yao, L., Guo, B., Yu, Z., & Liu, Y. (2021). Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR)*, *54*(4), 1–40.

Chen, Z., Prosperi, M., Bian, J., Min, J., Wang, M., & Li, C. (2019). Clinical correlates of workplace injury occurrence and recurrence in adults. *PLoS ONE*, *14*(9), 1–10. https://doi.org/10.1371/journal.pone.0222603

Chiarotto, A., Maxwell, L. J., Ostelo, R. W., Boers, M., Tugwell, P., & Terwee, C. B. (2019). Measurement Properties of Visual Analogue Scale, Numeric Rating Scale, and Pain Severity Subscale of the Brief Pain Inventory in Patients With Low Back Pain: A Systematic Review. *Journal of Pain*, *20*(3), 245–263. https://doi.org/10.1016/j.jpain.2018.07.009

Chiu, H., Adeli, E., Wang, B., Huang, D.-A., & Niebles, J. C. (2019). Action-agnostic human pose forecasting. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1423–1432.

Cholewicki, J., & McGill, S. M. (1994). EMG assisted optimization: A hybrid approach for estimating muscle forces in an indeterminate biomechanical model. *Journal of Biomechanics*, *27*(10), 1287–1289. https://doi.org/10.1016/0021-9290(94)90282-8

Christodoulou, C. (2017). Approaches to the measurement of fatigue. In *The Handbook of Operator Fatigue* (pp. 125–138). CRC Press.

Chung, M. K., Lee, I., Kee, D., & Kim, S. H. (2002). A postural workload evaluation system based on a macro-postural classification. *Human Factors and Ergonomics In Manufacturing*, *12*(3), 267–277. https://doi.org/10.1002/hfm.10017

Cifrek, M., Medved, V., Tonković, S., & Ostojić, S. (2009). Surface EMG based muscle fatigue evaluation in biomechanics. *Clinical Biomechanics*, *24*(4), 327–340. https://doi.org/10.1016/j.clinbiomech.2009.01.010

Claudino, J. G., Capanema, D. de O., de Souza, T. V., Serrão, J. C., Machado Pereira, A. C., & Nassis, G. P. (2019). Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a Systematic Review. *Sports Medicine - Open*, *5*(1), 1–12. https://doi.org/10.1186/s40798-019-0202-3

Clouthier, A. L., Ross, G. B., & Graham, R. B. (2020). Sensor data required for automatic recognition of athletic tasks using deep neural networks. *Frontiers in Bioengineering and Biotechnology*, *7*(January), 1–8. https://doi.org/10.3389/fbioe.2019.00473

Colyer, S. L., Evans, M., Cosker, D. P., & Salo, A. I. T. (2018). A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. *Sports Medicine - Open*, *4*(1), 24–38. https://doi.org/10.1186/s40798-018-0139-y

Conforti, I., Mileti, I., Prete, Z. Del, & Palermo, E. (2020). Measuring Biomechanical Risk in Lifting Load Tasks Through Wearable System and Machine-Learning Approach. *Sensors*, *20*(6), 1557.

Connaboy, C., Eagle, S. R., Johnson, C. D., Flanagan, S. D., Mi, Q. I., & Nindl, B. C. (2018). Using Machine Learning to Predict Lower-Extremity Injury in US Special Forces. *Medicine & Science in Sports & Exercise*, *December*, 1073–1079. https://doi.org/10.1249/MSS.0000000000001881

Cortes, N., Onate, J., & Morrison, S. (2014). Differential effects of fatigue on movement variability. *Gait and Posture*, *39*(3), 888–893. https://doi.org/10.1016/j.gaitpost.2013.11.020

Côté, J. N., Raymond, D., Mathieu, P. A., Feldman, A. G., & Levin, M. F. (2005). Differences in multi-joint kinematic patterns of repetitive hammering in healthy, fatigued and shoulder-injured individuals. *Clinical Biomechanics*, *20*(6), 581–590. https://doi.org/10.1016/j.clinbiomech.2005.02.012

Crew, B. (2019). *The top 10 countries for scientific research in 2018*. Nature Index. https://www.natureindex.com/news-blog/top-ten-countries-research-science-twenty-nineteen

Cudlip, A. C., Callaghan, J. P., & Dickerson, C. R. (2015). Effects of sitting and standing on upper extremity physical exposures in materials handling tasks. *Ergonomics*, *58*(10), 1637–1646. https://doi.org/10.1080/00140139.2015.1035763

D'arco, L., Wang, H., & Zheng, H. (2022). Assessing Impact of Sensors and Feature Selection in Smart-Insole-Based Human Activity Recognition. *Methods and Protocols*, *5*(3). https://doi.org/10.3390/mps5030045

Darbandy, M., Rostamnezhad, M., Hussain, S., Khosravi, A., Nahavandi, S., & Sani, Z. (2020). A new approach to detect the physical fatigue utilizing heart rate signals. *Research in Cardiovascular Medicine*, *9*(1), 23. https://doi.org/10.4103/rcm.rcm_8_20

Davoudi Kakhki, F., Freeman, S. A., & Mosher, G. A. (2019). Evaluating machine learning performance in predicting injury severity in agribusiness industries. *Safety Science*, *117*(July 2018), 257–262. https://doi.org/10.1016/j.ssci.2019.04.026

de Kok, J., Vroonhof, P., Snijders, J., Roullis, G., Clarke, M., Peereboom, K., van Dorst, P., & Isusi, I. (2019). *Work-related musculoskeletal disorders: prevalence, costs and demographics in the EU*. https://doi.org/10.2802/66947

Del Rosario, M. B., Lovell, N. H., & Redmond, S. J. (2019). Learning the orientation of a loosely-fixed wearable IMU relative to the body improves the recognition rate of human postures and activities. *Sensors (Switzerland)*, *19*(13). https://doi.org/10.3390/s19132845

Delgado-Bonal, A., & Marshak, A. (2019). Approximate entropy and sample entropy: A comprehensive tutorial. In *Entropy* (Vol. 21, Issue 6). https://doi.org/10.3390/e21060541

Dick, R. B., Hudock, S. D., Lu, M.-L., Waters, T. R., & Putz-Anderson, V. (2016). Manual materials handling. *Physical and Biological Hazards of the Workplace*, 33–52.

Dimitrov, G. V., Arabadzhiev, T. I., Mileva, K. N., Bowtell, J. L., Crichton, N., & Dimitrova, N. A. (2006). Muscle fatigue during dynamic contractions assessed by new spectral indices. *Medicine and Science in Sports and Exercise*, *38*(11), 1971–1979. https://doi.org/10.1249/01.mss.0000233794.31659.6d

Dingwell, J. B. (2006). Lyapunov Exponents. *Wiley Encyclopedia of Biomedical Engineering*, *20*(7). https://doi.org/10.1002/9780471740360.ebs0702

Dolan, P., & Adams, M. A. (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. https://doi.org/10.1016/S0021-9290(98)00086-4

Donisi, L., Cesarelli, G., Coccia, A., Panigazzi, M., Capodaglio, E. M., & Addio, G. D. (2021). Work-Related Risk Assessment According to the Revised NIOSH Lifting Equation : A Preliminary Study Using a Wearable Inertial Sensor and Machine Learning. *Sensors*, *21*, 2593.

Du, B., Boileau, M., Wierts, K., Hignett, S., Fischer, S., & Yazdani, A. (2019). Existing Science on Human Factors and Ergonomics in the Design of Ambulances and EMS Equipment. *Prehospital Emergency Care*, *23*(5), 631–646. https://doi.org/10.1080/10903127.2019.1568651

Duda, R. O., Hart, P. E., Stork, D. G., Hard, P. E., Stork, D. G., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). John Wiley & Sons, Inc.

Eichelberger, P., Ferraro, M., Minder, U., Denton, T., Blasimann, A., Krause, F., & Baur, H. (2016). Analysis of accuracy in optical motion capture – A protocol for laboratory setup evaluation. *Journal of Biomechanics*, *49*(10), 2085–2088. https://doi.org/10.1016/j.jbiomech.2016.05.007

Eksioglu, M., Fernandez, J. E., & Twomey, J. M. (1996). Predicting peak pinch strength: Artificial neural networks vs. regression. *International Journal of Industrial Ergonomics*, *18*(5–6), 431–441. https://doi.org/10.1016/0169-8141(95)00106-9

Engkvist, I.-L., Hagberg, M., Hjelm, E. W., Menckel, E., Ekenvall, L., & Group, P. study. (1998). The accident process preceding overexertion back injuries in nursing personnel. *Scandinavian Journal of Work, Environment & Health*, *24*(5), 367–375.

Engkvist, I. L., Kjellberg, A., Wigaeus, H. E., Hagberg, M., Menckel, E., & Ekenvall, L. (2001). Back injuries among nursing personnel - Identification of work conditions with cluster analysis. *Safety Science*, *37*(1), 1–18. https://doi.org/10.1016/S0925-7535(00)00039-4

Engkvist, Inga Lill. (2004). The accident process preceding back injuries among Australian nurses. *Safety Science*, *42*(3), 221–235. https://doi.org/10.1016/S0925-7535(03)00044-4

Enoka, R. M., & Duchateau, J. (2008). Muscle fatigue: What, why and how it influences muscle function. *Journal of Physiology*, *586*(1), 11–23. https://doi.org/10.1113/jphysiol.2007.139477

Enoka, R. M., & Duchateau, J. (2016). Translating Fatigue to Human Performance. *Medicine and Science in Sports and Exercise*, *48*(11), 2228–2238. https://doi.org/10.1249/MSS.0000000000000929

Escobar-Linero, E., Domínguez-Morales, M., & Luis Sevillano, J. (2022). Worker’s physical fatigue classification using neural networks. *Expert Systems with Applications*, *198*, 116784.

Farzandipour, M., Nabovati, E., Saeedi, S., & Fakharian, E. (2018). Fuzzy decision support systems to diagnose musculoskeletal disorders: A systematic literature review. *Computer Methods and Programs in Biomedicine*, *163*, 101–109. https://doi.org/10.1016/j.cmpb.2018.06.002

Fawaz, I. H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, *33*(4), 917–963. https://doi.org/10.1007/s10618-019-00619-1

Fedorowich, L., Emery, K., Gervasi, B., & Côté, J. N. (2013). Gender differences in neck/shoulder muscular patterns in response to repetitive motion induced fatigue. *Journal of Electromyography and Kinesiology*, *23*(5), 1183–1189.

Ferrari, A., Micucci, D., Mobilio, M., & Napoletano, P. (2020). On the Personalization of Classification Models for Human Activity Recognition. *IEEE Access*, *8*, 32066–32079. https://doi.org/10.1109/ACCESS.2020.2973425

Folkard, S., Robertson, K. A., & Spencer, M. B. (2007). A Fatigue/Risk Index to assess work schedules. *Somnologie*, *11*(3), 177–185.

Friedman, L., & Komogortsev, O. V. (2019). Assessment of the effectiveness of seven biometric feature normalization techniques. *IEEE Transactions on Information Forensics and Security*, *14*(10), 2528–2536.

Fritsch, F. N., & Butland, J. (1984). A method for constructing local monotone piecewise cubic interpolants. *SIAM Journal on Scientific and Statistical Computing*, *5*(2), 300–304.

Fujimaki, G., & Mitsuya, R. (2002). Study of the seated posture for VDT work. *Displays*, *23*(1–2), 17–24. https://doi.org/10.1016/S0141-9382(02)00005-7

Furuya, S., Aoki, T., Nakahara, H., & Kinoshita, H. (2012). Individual differences in the biomechanical effect of loudness and tempo on upper-limb movements during repetitive piano keystrokes. *Human Movement Science*, *31*(1), 26–39. https://doi.org/10.1016/j.humov.2011.01.002

Gallagher, S., & Schall, M. C. (2017). Musculoskeletal disorders as a fatigue failure process: evidence, implications and research needs. *Ergonomics*, *60*(2), 255–269. https://doi.org/10.1080/00140139.2016.1208848

Ganga, G. M. D., Esposto, K. F., & Braatz, D. (2012). Application of discriminant analysis-based model for prediction of risk of low back disorders due to workplace design in industrial jobs. *Work*, *41*(SUPPL.1), 2370–2376. https://doi.org/10.3233/WOR-2012-0467-2370

Gao, Y., González, V. A., Yiu, T. W., & Cabrera-Guerrero, G. (2019). *The Use of Machine Learning and Big Five Personality Taxonomy to Predict Construction Workers' Safety Behaviour*. 1–34.

Gao, Y., González, V. A., Yiu, T. W., & Cabrera-Guerrero, G. (2020). *Non-linearity identification for construction workers ' personality-safety behaviour predictive relationship using neural network and linear regression modelling*. 1–32.

García-Pedrajas, N., Ortiz-Boyer, D., Castillo-Gomariz, R. del, & Hervás-Martínez, C. (2005). Cascade ensembles. *Computational Intelligence and Bioinspired Systems: 8th International Work-Conference on Artificial Neural Networks, IWANN 2005, Vilanova i La Geltrú, Barcelona, Spain, June 8-10, 2005. Proceedings 8*, 598–603.

Gauche, E., Couturier, A., Lepers, R., Michaut, A., Rabita, G., & Hausswirth, C. (2009). Neuromuscular fatigue following high versus low-intensity eccentric exercise of biceps brachii muscle. *Journal of Electromyography and Kinesiology*, *19*(6), e481–e486. https://doi.org/10.1016/j.jelekin.2009.01.006

Gaudez, C., Gilles, M. A., & Savin, J. (2016). Intrinsic movement variability at work. How long is the path from motor control to design engineering? *Applied Ergonomics*, *53*, 71–78. https://doi.org/10.1016/j.apergo.2015.08.014

Georgakis, A., Stergioulas, L. K., & Giakas, G. (2003). Fatigue analysis of the surface EMG signal in isometric constant force contractions using the averaged instantaneous frequency. *IEEE Transactions on Biomedical Engineering*, *50*(2), 262–265. https://doi.org/10.1109/TBME.2002.807641

Gerdle, B., Edström, M., & Rahm, M. (1993). Fatigue in the shoulder muscles during static work at two different torque levels. *Clinical Physiology*, *13*(5), 469–482. https://doi.org/10.1111/j.1475-097X.1993.tb00463.x

Ghasemi, F., & Mahdavi, N. (2020). A new scoring system for the Rapid Entire Body Assessment (REBA) based on fuzzy sets and Bayesian networks. *International Journal of Industrial Ergonomics*, *80*(November 2018), 103058. https://doi.org/10.1016/j.ergon.2020.103058

Gholami, M., Napier, C., Patiño, A. G., Cuthbert, T. J., & Menon, C. (2020). Fatigue monitoring in running using flexible textile wearable sensors. *Sensors*, *20*(19), 5573.

Gholipour, A., & Arjmand, N. (2016). Artificial neural networks to predict 3D spinal posture in reaching and lifting activities; Applications in biomechanical models. *Journal of Biomechanics*, *49*(13), 2946–2952. https://doi.org/10.1016/j.jbiomech.2016.07.008

Giannini, P., Bassani, G., Avizzano, C. A., & Filippeschi, A. (2020). Wearable sensor network for biomechanical overload assessment in manual material handling. *Sensors*, *20*(14), 3877.

Giraudo, M., Bena, A., Leombruni, R., & Costa, G. (2016). Occupational injuries in times of labour market flexibility: The different stories of employment-secure and precarious workers. *BMC Public Health*, *16*(1), 1–11. https://doi.org/10.1186/s12889-016-2834-2

Goes, R. A., Lopes, L. R., Cossich, V. R. A., De Miranda, V. A. R., Coelho, O. N., Do Carmo Bastos, R., Domenis, L. A. M., Guimarães, J. A. M., Grangeiro-Neto, J. A., & Perini, J. A. (2020). Musculoskeletal injuries in athletes from five modalities: A cross-sectional study. *BMC Musculoskeletal Disorders*, *21*(1), 1–9. https://doi.org/10.1186/s12891-020-3141-8

Goh, Y. M., Ubeynarayana, C. U., Wong, K. L. X., & Guo, B. H. W. (2018). Factors influencing unsafe behaviors: A supervised learning approach. *Accident Analysis and Prevention*, *118*(June), 77–85. https://doi.org/10.1016/j.aap.2018.06.002

Gohari, M., Rahman, R. A., Tahmasebi, M., & Nejat, P. (2014). Off-road vehicle seat suspension optimisation, part I: Derivation of an artificial neural network model to predict seated human spine acceleration in vertical vibration. *Journal of Low Frequency Noise Vibration and Active Control*, *33*(4), 429–442. https://doi.org/10.1260/0263-0923.33.4.429

Gomes, P., Margaritoff, P., & Silva, H. (2019). pyHRV: Development and evaluation of an open-source python toolbox for heart rate variability (HRV). *Proc. Int'l Conf. On Electrical, Electronic and Computing Engineering (Icetran)*, 822–828.

González-Frutos, P., Aguilar-Navarro, M., Morencos, E., Mallo, J., & Veiga, S. (2021). Relationships between strength and step frequency with fatigue index in repeated sprint ability. *International Journal of Environmental Research and Public Health*, *19*(1), 196.

González-Izal, M., Malanda, A., Gorostiaga, E., & Izquierdo, M. (2012). Electromyographic models to assess muscle fatigue. *Journal of Electromyography and Kinesiology*, *22*(4), 501–512. https://doi.org/10.1016/j.jelekin.2012.02.019

Google. (2021). *Machine Learning Glossary*. Machine Learning Crash Course. https://developers.google.com/machine-learning/glossary

Gorelick, M., Brown, J. M. M., & Groeller, H. (2003). Short-duration fatigue alters neuromuscular coordination of trunk musculature: Implications for injury. *Applied Ergonomics*, *34*(4), 317–325. https://doi.org/10.1016/S0003-6870(03)00039-5

Granata, K. P., & Gottipati, P. (2008). Fatigue influences the dynamic stability of the torso. *Ergonomics*, *51*(8), 1258–1271. https://doi.org/10.1080/00140130802030722

Greene, R. L., Azari, D. P., Hu, Y. H., & Radwin, R. G. (2017). Visualizing stressful aspects of repetitive motion tasks and opportunities for ergonomic improvements using computer vision. *Applied Ergonomics*, *65*, 461–472. https://doi.org/10.1016/j.apergo.2017.02.020

Greene, R. L., Hu, Y. H., Difranco, N., Wang, X., Lu, M. L., Bao, S., Lin, J. H., & Radwin, R. G. (2019). Predicting Sagittal Plane Lifting Postures From Image Bounding Box Dimensions. *Human Factors*, *61*(1), 64–77. https://doi.org/10.1177/0018720818791367

Gribble, P. A., & Hertel, J. (2004). Effect of lower-extremity muscle fatigue on postural control. *Archives of Physical Medicine and Rehabilitation*, *85*(4), 589–592. https://doi.org/10.1016/j.apmr.2003.06.031

Gross, D. P., Steenstra, I. A., Harrell, F. E., Bellinger, C., & Zaïane, O. (2020). Machine Learning for Work Disability Prevention: Introduction to the Special Series. *Journal of Occupational Rehabilitation*, *30*(3), 303–307. https://doi.org/10.1007/s10926-020-09910-1

Guerrero, F. N., Spinelli, E. M., & Haberman, M. A. (2016). Analysis and Simple Circuit Design of Double Differential EMG Active Electrode. *IEEE Transactions on Biomedical Circuits and Systems*, *10*(3), 787–795. https://doi.org/10.1109/TBCAS.2015.2492944

Guo, S., He, J., Li, J., & Tang, B. (2020). Exploring the impact of unsafe behaviors on building construction accidents using a Bayesian network. *International Journal of Environmental Research and Public Health*, *17*(1). https://doi.org/10.3390/ijerph17010221

Gutnik, B., Mackie, H., Hudson, G., & Standen, C. (2005). How close to a pendulum is human upper limb movement during walking? *Homo*, *56*(1), 35–49.

Hagberg, M. (1981). Electromyographic signs of shoulder muscular fatigue in two elevated arm positions. *American Journal of Physical Medicine*, *60*(3), 111–121.

Hajifar, S., Sun, H., Megahed, F. M., Jones-Farmer, L. A., Rashedi, E., & Cavuoto, L. A. (2021). A forecasting framework for predicting perceived fatigue: Using time series methods to forecast ratings of perceived exertion with features from wearable sensors. *Applied Ergonomics*, *90*(September 2020), 103262. https://doi.org/10.1016/j.apergo.2020.103262

Hales, T. R., & Bernard, B. P. (1996). Epidemiology of work-related musculoskeletal disorders. *Orthopedic Clinics of North America*, *27*(4), 679–709. https://doi.org/10.1016/s0030-5898(20)32117-9

Halilaj, E., Rajagopal, A., Fiterau, M., Hicks, J. L., Hastie, T. J., & Delp, S. L. (2018). Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. *Journal of Biomechanics*, *81*, 1–11. https://doi.org/10.1016/j.jbiomech.2018.09.009

Hall, M. A. (2000). *Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning* (00/08).

Hamill, J., Selbie, W. S., & Kepple, T. M. (2014). Three-Dimensional Kinematics. In *Research Methods in Biomechanics* (2nd ed., pp. 35–59).

Hamilton, P. S. (2002). *Open Source ECG Analysis Software Documentation*. EP Limited.

Hammarskjöld, E., & Harms-Ringdahl, K. (1992). Effect of arm-shoulder fatigue on carpenters at work. *European Journal of Applied Physiology and Occupational Physiology*, *64*(5), 402–409. https://doi.org/10.1007/BF00625058

Hämmerle, P., Eick, C., Blum, S., Schlageter, V., Bauer, A., Rizas, K. D., Eken, C., Coslovsky, M., Aeschbacher, S., Krisai, P., & others. (2020). Heart rate variability triangular index as a predictor of cardiovascular mortality in patients with atrial fibrillation. *Journal of the American Heart Association*, *9*(15), e016075.

Haque, M. A., Irani, R., Nasrollahi, K., & Moeslund, T. (2016). Facial Video based Detection of Physical Fatigue for Maximal Muscle Activity. *IET Computer Vision*, *10*(4), 323 – 330. https://doi.org/10.1049/iet- cvi.2015.0215

Harber, P., & Leroy, G. (2017). Feasibility and Utility of Lexical Analysis for Occupational Health Text. *Journal of Occupational and Environmental Medicine*, *59*(6), 578–587. https://doi.org/10.1097/JOM.0000000000001035

Harrell Jr, F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis* (2nd ed.). Springer. https://doi.org/https://doi.org/10.1007/978-3-319-19425-7

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society*, 904–908. https://doi.org/10.1177/154193120605000909

Heiden, M., Mathiassen, S. E., Garza, J., Liv, P., & Wahlström, J. (2016). A Comparison of Two Strategies for Building an Exposure Prediction Model. *Annals of Occupational Hygiene*, *60*(1), 74–89. https://doi.org/10.1093/annhyg/mev072

Hernandez, N., Lundström, J., Favela, J., McChesney, I., & Arnrich, B. (2020). Literature Review on Transfer Learning for Human Activity Recognition Using Mobile and Wearable Devices with Environmental Technology. *SN Computer Science*, *1*(2), 1–16. https://doi.org/10.1007/s42979-020-0070-4

Hernandez, V., Rezzoug, N., Gorce, P., & Venture, G. (2018). Force feasible set prediction with artificial neural network and musculoskeletal model. *Computer Methods in Biomechanics and Biomedical Engineering*, *21*(14), 740–749. https://doi.org/10.1080/10255842.2018.1516763

Herrmann, C. M., Madigan, M. L., Davidson, B. S., & Granata, K. P. (2006). Effect of lumbar extensor fatigue on paraspinal muscle reflexes. *Journal of Electromyography and Kinesiology*, *16*(6), 637–641. https://doi.org/10.1016/j.jelekin.2005.11.004

Hlucny, S. D., & Novak, D. (2020). Characterizing human box-lifting behavior using wearable inertial motion sensors. *Sensors*, *20*(8), 2323.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hong, S., Ham, Y., Chun, J., & Kim, H. (2023). Productivity Measurement through IMU-Based Detailed Activity Recognition Using Machine Learning: A Case Study of Masonry Work. *Sensors*, *23*(17), 7635.

Hosseinian, S. M., Zhu, Y., Mehta, R. K., Erraguntla, M., & Lawley, M. A. (2019). Static and dynamic work activity classification from a single accelerometer: Implications for ergonomic assessment of manual handling tasks. *IISE Transactions on Occupational Ergonomics and Human Factors*, *7*(1), 59–68.

Hughes, R. E. (2017). Using a Bayesian Network to Predict L5/S1 Spinal Compression Force from Posture, Hand Load, Anthropometry, and Disc Injury Status. *Applied Bionics and Biomechanics*, *2017*. https://doi.org/10.1155/2017/2014961

Hunter, A. M., Gibson, A. S. C., Lambert, M., & Noakes, T. D. (2002). Electromyographic (EMG) normalization method for cycle fatigue protocols. *Medicine & Science in Sports & Exercise*, *34*(5), 857–861. https://doi.org/10.1097/00005768-200205000-00020

Hwang, S., Seo, J. O., Jebelli, H., & Lee, S. H. (2016). Feasibility analysis of heart rate monitoring of construction workers using a photoplethysmography (PPG) sensor embedded in a wristband-type activity tracker. *Automation in Construction*, *71*(Part 2), 372–381. https://doi.org/10.1016/j.autcon.2016.08.029

Islam, M. M., Nooruddin, S., Karray, F., & Muhammad, G. (2022). Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges, and future prospects. *Computers in Biology and Medicine*, 106060.

Ito, K., & Hotta, Y. (2012). EMG-based detection of muscle fatigue during low-level isometric contraction by recurrence quantification analysis and monopolar configuration. *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, *2012*, 4237–4241. https://doi.org/10.1109/EMBC.2012.6346902.

Jacobs, J. V, Hettinger, L. J., Huang, Y.-H., Jeffries, S., Lesch, M. F., Simmons, L. A., Verma, S. K., & Willetts, J. L. (2019). Employee acceptance of wearable technology in the workplace. *Applied Ergonomics*, *78*, 148–156.

Jeklin, A. T., Perrotta, A. S., Davies, H. W., Bredin, S. S. D., Paul, D. A., & Warburton, D. E. R. (2021). The association between heart rate variability, reaction time, and indicators of workplace fatigue in wildland firefighters. *International Archives of Occupational and Environmental Health*, *94*, 823–831.

Jeong, H., & Park, W. (2020). Developing and Evaluating a Mixed Sensor Smart Chair System for Real-time Posture Classification: Combining Pressure and Distance sensors. *IEEE Journal of Biomedical and Health Informatics*, *2194*(c), 1–9. https://doi.org/10.1109/JBHI.2020.3030096

Jia, B., & Nussbaum, M. A. (2018). Influences of continuous sitting and psychosocial stress on low back kinematics, kinetics, discomfort, and localized muscle fatigue during unsupported sitting activities. *Ergonomics*, *61*(12), 1671–1684. https://doi.org/10.1080/00140139.2018.1497815

Jimale, A. O., & Mohd Noor, M. H. (2023). Subject variability in sensor-based activity recognition. *Journal of Ambient Intelligence and Humanized Computing*, *14*(4), 3261–3274. https://doi.org/10.1007/s12652-021-03465-6

Juel, C. (1988). Muscle action potential propagation velocity changes during activity. *Muscle & Nerve*, *11*(7), 714–719. https://doi.org/10.1002/mus.880110707

Kakhki, F. D., Freeman, S. A., & Mosher, G. A. (2019). Use of neural networks to identify safety prevention priorities in agro-manufacturing operations within commercial grain elevators. *Applied Sciences (Switzerland)*, *9*(21), 1–16. https://doi.org/10.3390/app9214690

Kallenberg, L. A. C., Schulte, E., Disselhorst-Klug, C., & Hermens, H. J. (2007). Myoelectric manifestations of fatigue at low contraction levels in subjects with and without chronic pain. *Journal of Electromyography and Kinesiology*, *17*(3), 264–274. https://doi.org/10.1016/j.jelekin.2006.04.004

Kamen, G. (2014). Electromyographic Kinesiology. In *Research Methods in Biomechanics* (2nd ed., pp. 179–201). Human Kinetics.

Kang, D., Kim, Y. K., Kim, E. A., Kim, D. H., Kim, I., Kim, H. R., Min, K. B., Jung-Choi, K., Oh, S. S., & Koh, S. B. (2014). Prevention of Work-Related Musculoskeletal Disorders. *Annals of Occupational and Environmental Medicine*, *26*(1), 9–10. https://doi.org/10.1186/2052-4374-26-14

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. https://doi.org/10.1109/TNSRE.2013.2291327

Karthick, P. A., Ghosh, D. M., & Ramakrishnan, S. (2018). Surface electromyography based muscle fatigue detection using high-resolution time-frequency methods and machine learning algorithms. *Computer Methods and Programs in Biomedicine*, *154*, 45–56. https://doi.org/10.1016/j.cmpb.2017.10.024

Karvekar, S., Abdollahi, M., & Rashedi, E. (2021a). Smartphone-based human fatigue level detection using machine learning approaches. *Ergonomics*, *64*(5), 600–612. https://doi.org/10.1080/00140139.2020.1858185

Karvekar, S., Abdollahi, M., & Rashedi, E. (2021b). Smartphone-based human fatigue level detection using machine learning approaches. *Ergonomics*, *64*(5), 600–612. https://doi.org/10.1080/00140139.2020.1858185

Karwowski, W., Gaweda, A., Marras, W. S., Davis, K., Zurada, J. M., & Rodrick, D. (2006). A fuzzy relational rule network modeling of electromyographical activity of trunk muscles in manual lifting based on trunk angels, moments, pelvic tilt and rotation angles. *International Journal of Industrial Ergonomics*, *36*(10), 847–859. https://doi.org/10.1016/j.ergon.2006.06.006

Kazemi, Z., Mazloumi, A., Arjmand, N., Keihani, A., Karimi, Z., Ghasemi, M. S., & Kordi, R. (2021). A Comprehensive Evaluation of Spine Kinematics, Kinetics, and Trunk Muscle Activities During Fatigue-Induced Repetitive Lifting. *Human Factors*, *00*(0), 1–16. https://doi.org/10.1177/0018720820983621

Khatun, M. A., Yousuf, M. A., Ahmed, S., Uddin, M. Z., Alyami, S. A., Al-Ashhab, S., Akhdar, H. F., Khan, A., Azad, A., & Moni, M. A. (2022). Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor. *IEEE Journal of Translational Engineering in Health and Medicine*, *10*, 1–16.

Khatun, M. A., Yousuf, M. A., & Moni, M. A. (2023). Deep cnn-gru based human activity recognition with automatic feature extraction using smartphone and wearable sensors. *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 1–6.

Kilby, J., & Hosseini, H. G. (2004). Wavelet analysis of Surface Electromyography signals. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, *26 I*, 384–387. https://doi.org/10.1109/iembs.2004.1403174

Kim, K., & Cho, Y. K. (2020). Effective inertial sensor quantity and locations on a body for deep learning-based worker's motion recognition. *Automation in Construction*, *113*, 103126.

Kim, S., & Nussbaum, M. A. (2014). An evaluation of classification algorithms for manual material handling tasks based on data obtained using wearable technologies. *Ergonomics*, *57*(7), 1040–1051.

Kingma, I., Baten, C. T. M., Dolan, P., Toussaint, H. M., Van Dieën, J. H., de Looze, M. P., & Adams, M.

A. (2001). Lumbar loading during lifting: A comparative study of three measurement techniques. *Journal of Electromyography and Kinesiology*, *11*(5), 337–345. https://doi.org/10.1016/S1050-6411(01)00011-6

Kluger, B. M., Krupp, L. B., & Enoka, R. M. (2013). Fatigue and fatigability in neurologic illnesses: Proposal for a unified taxonomy. *Neurology*, *80*(4), 409–416. https://doi.org/https://doi.org/10.1212/WNL.0b013e31827f07be

Kong, L., Li, H., Yu, Y., Luo, H., Skitmore, M., & Antwi-Afari, M. F. (2018). Quantifying the physical intensity of construction workers, a mechanical energy approach. *Advanced Engineering Informatics*, *38*(July), 404–419. https://doi.org/10.1016/j.aei.2018.08.005

Krishna, O. B., Maiti, J., Ray, P. K., & Mandal, S. (2015). Assessment of Risk of Musculoskeletal Disorders among Crane Operators in a Steel Plant: A Data Mining-Based Analysis. *Human Factors and Ergonomics In Manufacturing*, *25*(5), 559–572. https://doi.org/10.1002/hfm.20575

Laflamme, L. (1997). Overexertion-injury types among female Swedish nurses and nursing auxiliaries: An age-related problem? *Safety Science*, *27*(2–3), 129–139. https://doi.org/10.1016/S0925-7535(97)00075-1

Landsittel, D. P., Gardner, L. I., & Arena, V. C. (1998). Identifying populations at high risk for occupational back injury with neutral networks. *Human and Ecological Risk Assessment (HERA)*, *4*(6), 1337–1352. https://doi.org/10.1080/10807039891284703

Längkvist, M., Karlsson, L., & Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, *42*(1), 11–24. https://doi.org/10.1016/j.patrec.2014.01.008

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, G., Bae, J., Jacobs, J. V, & Lee, S. (2024). Wearable heart rate sensing and critical power-based whole-body fatigue monitoring in the field. *Applied Ergonomics*, *121*, 104358.

Lee, H. M., Liau, J. J., Cheng, C. K., Tan, C. M., & Shih, J. T. (2003). Evaluation of shoulder proprioception following muscle fatigue. *Clinical Biomechanics*, *18*(9), 843–847. https://doi.org/10.1016/S0268-0033(03)00151-7

Lee, H., Yang, K., Kim, N., & Ahn, C. R. (2020). Detecting excessive load-carrying tasks using a deep learning network with a Gramian Angular Field. *Automation in Construction*, *120*(August), 103390. https://doi.org/10.1016/j.autcon.2020.103390

Lee, W., Karwowski, W., Marras, W. S., & Rodrick, D. (2003). A neuro-fuzzy model for estimating electromyographical activity of trunk muscles due to manual lifting. *Ergonomics*, *46*(1–3), 285–309. https://doi.org/10.1080/00140130303520

Lehto, M., Marucci-Wellman, H., & Corns, H. (2009). Bayesian methods: A useful tool for classifying injury narratives into cause groups. *Injury Prevention*, *15*(4), 259–265. https://doi.org/10.1136/ip.2008.021337

Leijon, O., Härenstam, A., Waldenström, K., Alderling, M., & Vingård, E. (2006). Target groups for prevention of neck/shoulder and low back disorders: An exploratory cluster analysis of working and living conditions. *Work*, *27*(2), 189–204.

Leise, T. L. (2017). Analysis of Nonstationary Time Series for Biological Rhythms Research. *Journal of Biological Rhythms*, *32*(3), 187–194. https://doi.org/10.1177/0748730417709105

Leleu, T. D., Jacobson, I. G., Leardmann, C. A., Smith, B., Foltz, P. W., Amoroso, P. J., Derr, M. A., Ryan, M. A., & Smith, T. C. (2011). Application of latent semantic analysis for open-ended responses in a large, epidemiologic study. *BMC Medical Research Methodology*, *11*. https://doi.org/10.1186/1471-2288-11-136

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: advancing the methodology. *Implementation Science*, *5*(1), 69. https://doi.org/10.1017/cbo9780511814563.003

Li, L., Martin, T., & Xu, X. (2020). A novel vision-based real-time method for evaluating postural risk factors associated with musculoskeletal disorders. *Applied Ergonomics*, *87*(March). https://doi.org/10.1016/j.apergo.2020.103138

Lim, S., & D'Souza, C. (2020). A narrative review on contemporary and emerging uses of inertial sensing in occupational ergonomics. *International Journal of Industrial Ergonomics*, *76*(November 2019), 102937. https://doi.org/10.1016/j.ergon.2020.102937

Lim, S., & Souza, C. D. (2019). Statistical prediction of load carriage mode and magnitude from inertial sensor derived gait kinematics. *Applied Ergonomics*, *76*, 1–11. https://doi.org/10.1016/j.apergo.2018.11.007

Lin, F., Song, C., Xu, X., Cavuoto, L., & Xu, W. (2017). Patient handling activity recognition through pressure-map manifold learning using a footwear sensor. *Smart Health*, *1–2*(May 2017), 77–92. https://doi.org/10.1016/j.smhl.2017.04.005

Lin, F., Wang, A., Cavuoto, L., & Xu, W. (2017). Toward Unobtrusive Patient Handling Activity Recognition for Injury Reduction Among At-Risk Caregivers. *Journal of Biomedical and Health Informatics (JBHI)*, *21*(3), 682–695.

Ling, C. F., Radin Umar, R. Z., & Ahmad, N. (2020). Development of a predictive model for work-relatedness of MSDs among semiconductor back-end workers. *International Journal of Occupational Safety and Ergonomics*, *0*(0), 1–11. https://doi.org/10.1080/10803548.2020.1840116

Liu, B., Li, Y., Zhang, S., & Ye, X. (2017). Healthy human sitting posture estimation in RGB-D scenes using object context. *Multimedia Tools and Applications*, *76*(8), 10721–10739. https://doi.org/10.1007/s11042-015-3189-x

Looft, J. M., & Frey-Law, L. A. (2020). Adapting a fatigue model for shoulder flexion fatigue: Enhancing recovery rate during intermittent rest intervals. *Journal of Biomechanics*, *106*, 109762. https://doi.org/10.1016/j.jbiomech.2020.109762

Lu, C. M., & Ferrier, N. J. (2004). Repetitive Motion Analysis: Segmentation and Event Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*(2), 258–263. https://doi.org/10.1109/TPAMI.2004.1262196

Ma, L., Bennis, F., Chablat, D., & Zhang, W. (2008). Framework for dynamic evaluation of muscle fatigue in manual handling work. *Proceedings of the IEEE International Conference on Industrial Technology*, 1–6. https://doi.org/10.1109/ICIT.2008.4608546

Ma, L., Chablat, D., Bennis, F., Zhang, W., & Guillaume, F. (2010). A new muscle fatigue and recovery model and its ergonomics application in human simulation. *Virtual and Physical Prototyping*, *5*(3), 123–137.

Maman, Z. S., Chen, Y. J., Baghdadi, A., Lombardo, S., Cavuoto, L. A., & Megahed, F. M. (2020). A data analytic framework for physical fatigue management using wearable sensors. *Expert Systems with Applications*, *155*, 113405. https://doi.org/10.1016/j.eswa.2020.113405

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. https://doi.org/10.1016/j.apergo.2017.02.001

Manor, E., & Greenberg, S. (2022). Custom hardware inference accelerator for tensorflow lite for microcontrollers. *IEEE Access*, *10*, 73484–73493.

Marfia, G., & Roccetti, M. (2017). A practical computer based vision system for posture and movement sensing in occupational medicine. *Multimedia Tools and Applications*, *76*(6), 8109–8129. https://doi.org/10.1007/s11042-016-3469-0

Marotta, L., Buurke, J. H., van Beijnum, B.-J. F., & Reenalda, J. (2021). Towards machine learning-based detection of running-induced fatigue in real-world scenarios: Evaluation of IMU sensor configurations to reduce intrusiveness. *Sensors*, *21*(10), 3451.

Márquez Gómez, M. (2020). Prediction of work-related musculoskeletal discomfort in the meat processing industry using statistical models. *International Journal of Industrial Ergonomics*, *75*(March 2018). https://doi.org/10.1016/j.ergon.2019.102876

Marras, W. S., Lavender, S. A., Leurgans, S. E., Fathallah, F. A., Ferguson, S. A., Allread, W. G., & Rajulu, S. L. (1995). Biomechanical risk factors for occupationally related low back disorders. *Ergonomics*, *38*(2), 377–410. https://doi.org/10.1080/00140139508925111

Marras, W. S., Lavender, S. A., Leurgans, S. E., Rajulu, S. L., Allread, G. W., Fathallah, F. A., & Ferguson, S. A. (1993). The Role of Dynamic Three-Dimensional Trunk Motion in Occupationally-Related Low Back Disorders. *Spine*, *18*(5), 617–628.

Martin, P. G., Weerakkody, N., Gandevia, S. C., & Taylor, J. L. (2008). Group III and IV muscle afferents differentially affect the motor cortex and motoneurones in humans. *Journal of Physiology*, *586*(5), 1277–1289. https://doi.org/10.1113/jphysiol.2007.140426

Marucci-Wellman, H., Lehto, M., & Corns, H. (2011). A combined Fuzzy and Naïve Bayesian strategy can be used to assign event codes to injury narratives. *Injury Prevention*, *17*(6), 407–414. https://doi.org/10.1136/ip.2010.030593

Marucci-Wellman, H. R., Corns, H. L., & Lehto, M. R. (2017). Classifying injury narratives of large administrative databases for surveillance—A practical approach combining machine learning ensembles and human review. *Accident Analysis and Prevention*, *98*, 359–371. https://doi.org/10.1016/j.aap.2016.10.014

Marucci-Wellman, H. R., Lehto, M. R., & Corns, H. L. (2015). A practical tool for public health surveillance: Semi-automated coding of short injury narratives from large administrative databases using Naïve Bayes algorithms. *Accident Analysis and Prevention*, *84*, 165–176. https://doi.org/10.1016/j.aap.2015.06.014

Masci, F., Rosecrance, J., Mixco, A., Cortinovis, I., Calcante, A., Mandic-Rajcevic, S., & Colosio, C. (2020). Personal and occupational factors contributing to biomechanical risk of the distal upper limb among dairy workers in the Lombardy region of Italy. *Applied Ergonomics*, *83*(June 2018), 102796. https://doi.org/10.1016/j.apergo.2018.12.013

Massiris Fernández, M., Fernández, J. Á., Bajo, J. M., & Delrieux, C. A. (2020). Ergonomic risk assessment based on computer vision and machine learning. *Computers and Industrial Engineering*, *149*, 106816. https://doi.org/10.1016/j.cie.2020.106816

Mathiassen, S. E., & Aminoff, T. (1997). Motor control and cardiovascular responses during isoelectric contractions of the upper trapezius muscle: Evidence for individual adaptation strategies. *European*

Journal of Applied Physiology and Occupational Physiology, *76*(5), 434–444. https://doi.org/10.1007/s004210050273

Matijevich, E. S., Volgyesi, P., & Zelik, K. E. (2021). A promising wearable solution for the practical and accurate monitoring of low back loading in manual material handling. *Sensors*, *21*(2), 1–25. https://doi.org/10.3390/s21020340

Mavor, M. P., Chan, V. C. H., Gruevski, K. M., Bossi, L. L. M., Karakolis, T., & Graham, R. B. (2023). Assessing the Soldier Survivability Tradespace Using a Single IMU. *IEEE Access*, *11*, 69762–69772. https://doi.org/10.1109/ACCESS.2023.3286305

Mavor, M. P., Ross, G. B., Clouthier, A. L., Karakolis, T., & Graham, R. B. (2020). Validation of an IMU suit for military-based tasks. *Sensors*, *20*(15), 1–14. https://doi.org/10.3390/s20154280

McGill, S. M. (1997). The biomechanics of low back injury: Implications on current practice in industry and the clinic. *Journal of Biomechanics*, *30*(5), 465–475. https://doi.org/10.1016/S0021-9290(96)00172-8

McKinney, W., & others. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, *445*, 51–56.

Mehrizi, R., Peng, X., Metaxas, D. N., Xu, X., Zhang, S., & Li, K. (2019). Predicting 3-D lower back joint load in lifting: A deep pose estimation approach. *IEEE Transactions on Human-Machine Systems*, *49*(1), 85–94. https://doi.org/10.1109/THMS.2018.2884811

Mehrizi, R., Peng, X., Xu, X., Zhang, S., Metaxas, D., & Li, K. (2018). A computer vision based method for 3D posture estimation of symmetrical lifting. *Journal of Biomechanics*, *69*, 40–46. https://doi.org/10.1016/j.jbiomech.2018.01.012

Mehrizi, R., Xu, X., Zhang, S., Pavlovic, V., Metaxas, D., & Li, K. (2017). Using a marker-less method for estimating L5/S1 moments during symmetrical lifting. *Applied Ergonomics*, *65*, 541–550. https://doi.org/10.1016/j.apergo.2017.01.007

Mehta, R. K., & Agnew, M. J. (2012). Effects of physical and mental demands on shoulder muscle fatigue. *Work*, *41*(SUPPL.1), 2897–2901. https://doi.org/10.3233/WOR-2012-0541-2897

Merriaux, P., Dupuis, Y., Boutteau, R., Vasseur, P., & Savatier, X. (2017). A study of vicon system positioning performance. *Sensors*, *17*(7), 1591.

Metzger, J. M. (1996). Effects of phosphate and ADP on shortening velocity during maximal and submaximal calcium activation of the thin filament in skeletal muscle fibers. *Biophysical Journal*, *70*(1), 409–417. https://doi.org/10.1016/S0006-3495(96)79584-X

Meyers, A. R., Al-Tarawneh, I. S., Wurzelbacher, S. J., Bushnell, P. T., Lampl, M. P., Bell, J. L., Bertke, S. J., Robins, D. C., Tseng, C. Y., Wei, C., Raudabaugh, J. A., & Schnorr, T. M. (2018). Applying machine learning to workers' compensation data to identify industry-specific ergonomic and safety prevention priorities: Ohio, 2001 to 2011. *Journal of Occupational and Environmental Medicine*, *60*(1), 55–73. https://doi.org/10.1097/JOM.0000000000001162

Michie, S., Thomas, J., Johnston, M., Aonghusa, P. Mac, Shawe-Taylor, J., Kelly, M. P., Deleris, L. A., Finnerty, A. N., Marques, M. M., Norris, E., O'Mara-Eves, A., & West, R. (2017). The Human Behaviour-Change Project: Harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. *Implementation Science*, *12*(1), 1–12. https://doi.org/10.1186/s13012-017-0641-5

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial*

Intelligence, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Mim, T. R., Amatullah, M., Afreen, S., Yousuf, M. A., Uddin, S., Alyami, S. A., Hasan, K. F., & Moni, M. A. (2023). GRU-INC: An inception-attention based approach using GRU for human activity recognition. *Expert Systems with Applications*, *216*, 119419.

Min, J.-Y., Song, S.-H., Kim, H., & Min, K.-B. (2019). Mining Hidden Knowledge About Illegal Compensation for Occupational Injury: Topic Model Approach. *JMIR Medical Informatics*, *7*(3), e14763. https://doi.org/10.2196/14763

Molnar, C. (2020). *Interpretable machine learning: A Guide for Making Black Box Models Explainable*. Leanpub.

Mörl, F., Wagner, H., & Blickhan, R. (2005). Lumbar spine intersegmental motion analysis during lifting. *Pathophysiology*, *12*(4), 295–302. https://doi.org/10.1016/j.pathophys.2005.09.003

Morse, S., Talty, K., Kuiper, P., Scioletti, M., Heymsfield, S. B., Atkinson, R. L., & Thomas, D. M. (2020). Machine learning prediction of combat basic training injury from 3D body shape images. *PLoS ONE*, *15*(6 June), 1–12. https://doi.org/10.1371/journal.pone.0235017

Mousavi, S. H., Sayyaadi, H., & Arjmand, N. (2020). Prediction of the thorax/pelvis orientations and L5–S1 disc loads during various static activities using neuro-fuzzy. *Journal of Mechanical Science and Technology*, *34*(8), 3481–3485. https://doi.org/10.1007/s12206-020-0740-0

Muller, A., Pontonnier, C., & Dumont, G. (2020). Motion-based prediction of hands and feet contact efforts during asymmetric handling tasks. *IEEE Transactions on Biomedical Engineering*, *67*(2), 344–352. https://doi.org/10.1109/TBME.2019.2913308

Murad, A., & Pyun, J. Y. (2017). Deep recurrent neural networks for human activity recognition. *Sensors*, *17*(11). https://doi.org/10.3390/s17112556

Murillo-Escobar, J., Jaramillo-Munera, Y. E., Orrego-Metaute, D. A., Delgado-Trejos, E., & Cuesta-Frau, D. (2020). Muscle fatigue analysis during dynamic contractions based on biomechanical features and Permutation Entropy. *Mathematical Biosciences and Engineering*, *17*(3), 2592–2615. https://doi.org/10.3934/mbe.2020142

Namian, M., Taherpour, F., Ghiasvand, E., & Turkan, Y. (2021). Insidious Safety Threat of Fatigue: Investigating Construction Workers' Risk of Accident Due to Fatigue. *Journal of Construction Engineering and Management*, *147*(12), 04021162. https://doi.org/10.1061/(asce)co.1943-7862.0002180

Nanda, G., Grattan, K. M., Chu, M. T., Davis, L. K., & Lehto, M. R. (2016). Bayesian decision support for coding occupational injury data. *Journal of Safety Research*, *57*, 71–82. https://doi.org/10.1016/j.jsr.2016.03.001

Nath, N. D., Chaspari, T., & Behzadan, A. H. (2018). Automated ergonomic risk monitoring using body-mounted sensors and machine learning. *Advanced Engineering Informatics*, *38*(September), 514–526. https://doi.org/10.1016/j.aei.2018.08.020

Nazarahari, M., & Rouhani, H. (2021). Sensor fusion algorithms for orientation tracking via magnetic and inertial measurement units: An experimental comparison survey. *Information Fusion*, *76*, 8–23. https://doi.org/10.1016/j.inffus.2021.04.009

Nematzadeh, S., Kiani, F., Torkamanian-Afshar, M., & Aydin, N. (2022). Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: A bioinformatics study on biomedical and biological cases. *Computational Biology and Chemistry*, *97*, 107619.

Nocella, M., Colombini, B., Benelli, G., Cecchi, G., Bagni, M. A., & Bruton, J. (2011). Force decline during fatigue is due to both a decrease in the force per individual cross-bridge and the number of cross-bridges. *The Journal of Physiology*, *589*(13), 3371–3381.

Nohara, Y., Matsumoto, K., Soejima, H., & Nakashima, N. (2022). Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Computer Methods and Programs in Biomedicine*, *214,* 106584.

Nurwulan, N. R., & Selamaj, G. (2021). Human daily activities recognition using decision tree. *Journal of Physics: Conference Series*, *1833*(1). https://doi.org/10.1088/1742-6596/1833/1/012039

O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., & Invernizzi, L. (2019). Keras tuner. *Retrieved May*, *21*, 2020. https://github.com/keras-team/keras-tuner

Oda, S., & Kida, N. (2001). Neuromuscular fatigue during maximal concurrent hand grip and elbow flexion or extension. *Journal of Electromyography and Kinesiology*, *11*(4), 281–289. https://doi.org/10.1016/S1050-6411(01)00004-9

Ordóñez, F. J., & Roggen, D. (2016). Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors*, *16*(1). https://doi.org/10.3390/s16010115

Ortega-Auriol, P. A., Besier, T. F., Byblow, W. D., & McMorland, A. J. C. (2018). Fatigue Influences the Recruitment, but Not Structure, of Muscle Synergies. *Frontiers in Human Neuroscience*, *12*(June), 1–12. https://doi.org/10.3389/fnhum.2018.00217

Owlia, M., Kamachi, M., & Dutta, T. (2020). Reducing lumbar spine flexion using real-time biofeedback during patient handling tasks. *Work*, *66*(1), 41–51. https://doi.org/10.3233/WOR-203149

Padula, R. S., Comper, M. L. C., Sparer, E. H., & Dennerlein, J. T. (2017). Job rotation designed to prevent musculoskeletal disorders and control risk in manufacturing industries: A systematic review. In *Applied Ergonomics* (Vol. 58, pp. 386–397). Elsevier Ltd. https://doi.org/10.1016/j.apergo.2016.07.018

Pagani, M., Lombardi, F., Guzzetti, S., Rimoldi, O., Furlan, R., Pizzinelli, P., Sandrone, G., Malfatto, G., Dell'Orto, S., & Piccaluga, E. (1986). Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. *Circulation Research*, *59*(2), 178–193.

Paillard, T. (2012). Effects of general and local fatigue on postural control: A review. *Neuroscience and Biobehavioral Reviews*, *36*(1), 162–176. https://doi.org/10.1016/j.neubiorev.2011.05.009

Parkinson, R. J., & Callaghan, J. P. (2009). The use of artificial neural networks to reduce data collection demands in determining spine loading: A laboratory based analysis. *Computer Methods in Biomechanics and Biomedical Engineering*, *12*(5), 511–522. https://doi.org/10.1080/10255840902740620

Patel, D. R., Yamasaki, A., & Brown, K. (2017). Epidemiology of sports-related musculoskeletal injuries in young athletes in United States. In *Translational Pediatrics* (Vol. 6, Issue 3, pp. 160–166). AME Publishing Company. https://doi.org/10.21037/tp.2017.04.08

Pattyn, N., Neyt, X., Henderickx, D., & Soetens, E. (2008). Psychophysiological investigation of vigilance decrement: boredom or cognitive fatigue? *Physiology & Behavior*, *93*(1–2), 369–378.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning*

Research, *12*, 2825–2830.

Perez, M. A., & Nussbaum, M. A. (2008). A neural network model for predicting postures during non-repetitive manual materials handling tasks. *Ergonomics*, *51*(10), 1549–1564. https://doi.org/10.1080/00140130802220570

Persona, A., Battini, D., Faccio, M., Bevilacqua, M., & Ciarapica, F. E. (2006). Classification of occupational injury cases using the regression tree approach. *International Journal of Reliability, Quality, and Safety Engineering*, *13*(2), 171–191.

Phinyomark, A., Thongpanja, S., Hu, H., Phukpattaranont, P., & Limsakul, C. (2012). The Usefulness of Mean and Median Frequencies in Electromyography Analysis. In G. R. Naik (Ed.), *Computational Intelligence in Electromyography Analysis* (1st ed., pp. 195–220). InTech. https://doi.org/http://dx.doi.org/10.5772/50639

Pistolesi, F., & Lazzerini, B. (2020). Assessing the Risk of Low Back Pain and Injury via Inertial and Barometric Sensors. *IEEE Transactions on Industrial Informatics*, *16*(11), 7199–7208. https://doi.org/10.1109/TII.2020.2992984

Place, N., Yamada, T., Bruton, J. D., & Westerblad, H. (2010). Muscle fatigue: From observations in humans to underlying mechanisms studied in intact single muscle fibres. *European Journal of Applied Physiology*, *110*(1), 1–15. https://doi.org/10.1007/s00421-010-1480-0

Porta, M., Kim, S., Pau, M., & Nussbaum, M. A. (2021). Classifying diverse manual material handling tasks using a single wearable sensor. *Applied Ergonomics*, *93*, 103386.

Qiu, S., Zhao, H., Jiang, N., Wang, Z., Liu, L., An, Y., Zhao, H., Miao, X., Liu, R., & Fortino, G. (2022). Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. *Information Fusion*, *80*, 241–265.

Qu, X., & Yeo, J. C. (2011). Effects of load carriage and fatigue on gait characteristics. *Journal of Biomechanics*, *44*(7), 1259–1263. https://doi.org/10.1016/j.jbiomech.2011.02.016

Raenović, T. (2023). ANALYSIS OF THE ACCIDENTS AT WORK IN THE EUROPEAN UNION. *Facta Universitatis, Series: Working and Living Environmental Protection*, 157–166.

Ramamurthy, S. R., & Roy, N. (2018). Recent trends in machine learning for human activity recognition—A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), 1–11. https://doi.org/10.1002/widm.1254

Ramanujam, E., Perumal, T., & Padmavathi, S. (2021). Human Activity Recognition with Smartphone and Wearable Sensors using Deep Learning Techniques: A Review. *IEEE Sensors Journal*, *21*(12), 13029–13040. https://doi.org/10.1109/JSEN.2021.3069927

Rampichini, S., Vieira, T. M., Castiglioni, P., & Merati, G. (2020). Complexity analysis of surface electromyography for assessing the myoelectric manifestation of muscle fatigue: A review. *Entropy*, *22*(5). https://doi.org/10.3390/E22050529

Rancourt, D., & Hogan, N. (2001). Dynamics of pushing. *Journal of Motor Behavior*, *33*(4), 351–362.

Ray, S. J., & Teizer, J. (2012). Real-time construction worker posture analysis for ergonomics training. *Advanced Engineering Informatics*, *26*(2), 439–455. https://doi.org/10.1016/j.aei.2012.02.011

Reme, S. E., Shaw, W. S., Steenstra, I. A., Woiszwillo, M. J., Pransky, G., & Linton, S. J. (2012). Distressed, immobilized, or lacking employer support? a sub-classification of acute work-related low back pain. *Journal of Occupational Rehabilitation*, *22*(4), 541–552. https://doi.org/10.1007/s10926-

012-9370-4

Resnick, M. (1996). Postural changes due to fatigue. *Computers and Industrial Engineering*, *31*(1–2), 491–494. https://doi.org/10.1016/0360-8352(96)00182-9

Robert-Lachaine, X., Mecheri, H., Larue, C., & Plamondon, A. (2017). Validation of inertial measurement units with an optoelectronic system for whole-body motion analysis. *Medical and Biological Engineering and Computing*, *55*(4), 609–619. https://doi.org/10.1007/s11517-016-1537-2

Roberts, D., Torres Calderon, W., Tang, S., & Golparvar-Fard, M. (2020). Vision-Based Construction Worker Activity Analysis Informed by Body Posture. *Journal of Computing in Civil Engineering*, *34*(4), 04020017. https://doi.org/10.1061/(asce)cp.1943-5487.0000898

Robertson, D. G. E., Caldwell, G. E., Hamill, J., Kamen, G., & Whittlesey, S. N. (2014). *Research methods in biomechanics* (2nd ed.). Human Kinetics.

Rodríguez, F. A. R., Flores, L. G., & Vitón-Castillo, A. A. (2022). Artificial intelligence and machine learning: present and future applications in health sciences. *Seminars in Medical Writing and Education*, *1*, 9.

Roetenberg, D., Luinge, H., & Slycke, P. (2013). Xsens MVN: full 6DOF human motion tracking using miniature inertial sensors. *Xsens Motion Technologies BV*, *April 3*, 1–7. http://www.xsens.com/images/stories/PDF/MVN_white_paper.pdf

Roth, E., Möncks, M., Bohné, T., & Pumplun, L. (2020). Context-aware cyber-physical assistance systems in industrial systems: A human activity recognition approach. *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, 1–6.

Rube, N., & Secher, N. H. (1991). Effect of training on central factors in fatigue following two- and one-leg static exercise in man. *Acta Physiologica Scandinavica*, *141*(1), 87–95. https://doi.org/10.1111/j.1748-1716.1991.tb09048.x

Ruddy, J. D., Cormack, S. J., Whiteley, R., Williams, M. D., Timmins, R. G., Opar, D. A., Aughey, R. J., Edouard, P., Smoliga, J. M., & Ruddy, J. D. (2019). Modeling the Risk of Team Sport Injuries: A Narrative Review of Different Statistical Approaches. *Frontiers in Physiology*, *10*(July), 1–16. https://doi.org/10.3389/fphys.2019.00829

Rudin, C., & Radin, J. (2019). Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From An Explainable AI Competition. *Harvard Data Science Review*, *Fall 2019*(1.2), 1–9. https://doi.org/10.1162/99608f92.5a8a3a3d

Rusakov, D. A. (2023). A misadventure of the correlation coefficient. *Trends in Neurosciences*, *46*(2), 94–96.

Ryu, J., Seo, J., Jebelli, H., & Lee, S. (2019). Automated action recognition using an accelerometer-embedded wristband-type activity tracker. *Journal of Construction Engineering and Management*, *145*(1), 4018114.

Sadeghi, H., Mohandes, S. R., Hosseini, M. R., Banihashemi, S., Mahdiyar, A., & Abdullah, A. (2020). Developing an ensemble predictive safety risk assessment model: Case of Malaysian construction projects. *International Journal of Environmental Research and Public Health*, *17*(22), 1–25. https://doi.org/10.3390/ijerph17228395

Saito, K. (1999). Measurement of fatigue in industries. *Industrial Health*, *37*(2), 134–142.

Samani, A., Mathiassen, S. E., & Madeleine, P. (2013). Cluster-based exposure variation analysis. *BMC*

Medical Research Methodology, *13*(1). https://doi.org/10.1186/1471-2288-13-54

Sarkar, S., & Maiti, J. (2020). Machine learning in occupational accident analysis: A review using science mapping approach with citation network analysis. *Safety Science*, *131*(September 2019), 104900. https://doi.org/10.1016/j.ssci.2020.104900

Sasikumar, V., & Binoosh, S. champakkadayil A. basheer. (2020). A model for predicting the risk of musculoskeletal disorders among computer professionals. *International Journal of Occupational Safety and Ergonomics*, *26*(2), 384–396. https://doi.org/10.1080/10803548.2018.1480583

Schall Jr, 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*, *60*(3), 351–362.

Schepers, M., Giuberti, M., & Bellusci, G. (2018). Xsens MVN: Consistent tracking of human motion using inertial sensing. *Xsens Technologies B.V.*, *March*, 1–8. https://doi.org/10.13140/RG.2.2.22099.07205

Scheurer, S., Tedesco, S., O'Flynn, B., & Brown, K. N. (2020). Comparing Person-Specific and Independent Models on Subject-Dependent and Independent Human Activity Recognition Performance. *Sensors*, *20*(13), 3647. https://doi.org/https://doi.org/10.3390/s20133647

Schmitt, L., Regnard, J., Desmarets, M., Mauny, F., Mourot, L., Fouillot, J. P., Coulmy, N., & Millet, G. (2013). Fatigue Shifts and Scatters Heart Rate Variability in Elite Endurance Athletes. *PLoS ONE*, *8*(8). https://doi.org/10.1371/journal.pone.0071588

Schmitt, L., Regnard, J., & Millet, G. P. (2015). Monitoring fatigue status with HRV measures in elite athletes: An avenue beyond RMSSD? *Frontiers in Physiology*, *6*(NOV), 2013–2015. https://doi.org/10.3389/fphys.2015.00343

Sedighi, A., & Nussbaum, M. A. (2019). Exploration of Different Classes of Metrics to Characterize Motor Variability during Repetitive Symmetric and Asymmetric Lifting Tasks. *Scientific Reports*, *9*(1), 9821. https://doi.org/10.1038/s41598-019-46297-3

Seo, J., Moon, M., & Lee, S. (2015). Construction operation simulation reflecting workers' muscle fatigue. In *Computing in Civil Engineering 2015* (pp. 515–522).

Shair, E. F., Ahmad, S. A., Marhaban, M. H., Tamrin, S. B. M., & Abdullah, A. R. (2017). EMG processing based measures of fatigue assessment during manual lifting. *BioMed Research International*, *2017*. https://doi.org/10.1155/2017/3937254

Sharma, L. K., & Majumder, J. (2017). Data mining approach to pre-screening of manual material handlers based on hand grip strength. *Journal of Ecophysiology and Occupational Health*, *17*(1–2), 57–62. http://www.informaticsjournals.com/index.php/JEOH/article/viewFile/17926/15920%0Ahttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed18&AN=620205773%0Ahttp://bf4dv7zn3u.search.serialssolutions.com.myaccess.library.utoronto.ca/?url_ver

Sivak-Callcott, J. A., Diaz, S. R., Ducatman, A. M., Rosen, C. L., Nimbarte, A. D., & Sedgeman, J. A. (2011). A survey study of occupational pain and injury in ophthalmic plastic surgeons. *Ophthalmic Plastic and Reconstructive Surgery*, *27*(1), 28–32. https://doi.org/10.1097/IOP.0b013e3181e99cc8

Smallman, C. L. W., Graham, R. B., & Stevenson, J. M. (2013). The effect of an on-body assistive device on transverse plane trunk coordination during a load carriage task. *Journal of Biomechanics*, *46*(15), 2688–2694. https://doi.org/10.1016/j.jbiomech.2013.07.034

Smith, P., LaMontagne, A. D., Lilley, R., Hogg-Johnson, S., & Sim, M. (2020). Are there differences in the return to work process for work-related psychological and musculoskeletal injuries? A longitudinal path analysis. *Social Psychiatry and Psychiatric Epidemiology*, 1–11. https://doi.org/10.1007/s00127-020-01839-3

Solomon Jr, O. M. (1991). PSD computations using Welch's method. *NASA STI/Recon Technical Report N*, *92*, 23584.

Solovieva, S., Vehmas, T., Riihimäki, H., Takala, E.-P., Murtomaa, H., Luoma, K., & Leino-Arjas, P. (2006). Finger Osteoarthritis and Differences in Dental Work Tasks. *Journal of Dental Research*, *85*(4), 344–348.

Sparto, P. J., Parnianpour, M., Reinsel, T. E., & Simon, S. (1997). The Effect of Fatigue on Multijoint Kinematics, Coordination, and Postural Stability During a Repetitive Lifting Test. *Journal of Orthopaedic & Sports Physical Therapy*, *25*(1), 3–12. https://doi.org/10.2519/jospt.1997.25.1.3

Spector, J. T., Lieblich, M., Bao, S., McQuade, K., & Hughes, M. (2014). Automation of workplace lifting hazard assessment for musculoskeletal injury prevention. *Annals of Occupational and Environmental Medicine*, *26*(1), 1–8. https://doi.org/10.1186/2052-4374-26-15

Stefana, E., Marciano, F., Rossi, D., Cocca, P., & Tomasoni, G. (2021). Wearable devices for ergonomics: A systematic literature review. *Sensors (Switzerland)*, *21*(3), 1–24. https://doi.org/10.3390/s21030777

Suárez Sánchez, A., Iglesias-Rodríguez, F. J., Riesgo Fernández, P., & de Cos Juez, F. J. (2014). Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders. *International Journal of Industrial Ergonomics*, *52*, 92–99. https://doi.org/10.1016/j.ergon.2015.09.012

Sung, P. C., Hsu, C. C., Lee, C. L., Chiu, Y. S. P., & Chen, H. L. (2014). Formulating grip strength and key pinch strength prediction models for Taiwanese: a comparison between stepwise regression and artificial neural networks. *Journal of Ambient Intelligence and Humanized Computing*, *6*(1), 37–46. https://doi.org/10.1007/s12652-014-0245-8

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). The MIT Press.

Swaen, G. M. H., Van Amelsvoort, L. G. P. M., Bültmann, U., & Kant, I. J. (2003). Fatigue as a risk factor for being injured in an occupational accident: Results from the Maastricht Cohort Study. *Occupational and Environmental Medicine*, *60*(SUPPL. 1), 88–92. https://doi.org/10.1136/oem.60.suppl_1.i88

Taha, Z., & Nazaruddin. (2005). Grip strength prediction for Malaysian industrial workers using artificial neural networks. *International Journal of Industrial Ergonomics*, *35*(9), 807–816. https://doi.org/10.1016/j.ergon.2004.11.006

Tanaka, M., Shigihara, Y., Ishii, A., Funakura, M., Kanai, E., & Watanabe, Y. (2012). Effect of mental fatigue on the central nervous system: an electroencephalography study. *Behavioral and Brain Functions*, *8*, 1–8.

Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Circulation*, *93*(5), 1043–1065.

Taylor, J. L., & Gandevia, S. C. (2008). A comparison of central aspects of fatigue in submaximal and maximal voluntary contractions. *Journal of Applied Physiology*, *104*(2), 542–550. https://doi.org/10.1152/japplphysiol.01053.2007

Thanathornwong, B., & Suebnukarn, S. (2015). The Improvement of Dental Posture Using Personalized Biofeedback. *Studies in Health Technology and Informatics*, *216*, 756–760. https://doi.org/10.3233/978-1-61499-564-7-756

Thanathornwong, B., Suebnukarn, S., & Ouivirach, K. (2014). A system for predicting musculoskeletal disorders among dental students. *International Journal of Occupational Safety and Ergonomics*, *20*(3), 463–475. https://doi.org/10.1080/10803548.2014.11077063

Tickoo, O., & Iyer, R. (2017). *Making Sense of Sensors: End-to-End Algorithms and Infrastructure Design from Wearable-Devices to Data Center* (1st ed.). Apress. https://doi.org/10.1007/978-1-4302-6593-1

Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2016). Application of machine learning to construction injury prediction. *Automation in Construction*, *69*, 102–114. https://doi.org/10.1016/j.autcon.2016.05.016

Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2017). Construction Safety Clash Detection: Identifying Safety Incompatibilities among Fundamental Attributes using Data Mining. *Automation in Construction*, *74*, 39–54. https://doi.org/10.1016/j.autcon.2016.11.001

Trafimow, J. H., Schipplein, O. D., Novak, G. J., & Andersson, G. B. J. (1993). The Effects of Quadriceps Fatigue on the Technique of Lifting. *Spine*, *18*(3), 364–367. https://doi.org/10.1097/00007632-199303000-00011

Tran, Y., Wijesuriya, N., Tarvainen, M., Karjalainen, P., & Craig, A. (2009). The relationship between spectral changes in heart rate variability and fatigue. *Journal of Psychophysiology*, *23*(3), 143–151. https://doi.org/10.1027/0269-8803.23.3.143

U.S. Bureau of Labor Statistics. (2023). *Employer-Reported Workplace Injuries and Illnesses, 2021-2022*.

Ueda, T., Nabetani, T., & Teramoto, K. (2006). Differential perceived exertion measured using a new visual analogue scale during pedaling and running. *Journal of Physiological Anthropology*, *25*(2), 171–177.

Umer, W., Li, H., Yantao, Y., Antwi-Afari, M. F., Anwer, S., & Luo, X. (2020). Physical exertion modeling for construction tasks using combined cardiorespiratory and thermoregulatory measures. *Automation in Construction*, *112*(November 2018), 103079. https://doi.org/10.1016/j.autcon.2020.103079

Vallmuur, K. (2015). Machine learning approaches to analysing textual injury surveillance data: A systematic review. *Accident Analysis and Prevention*, *79*, 41–49. https://doi.org/10.1016/j.aap.2015.03.018

van der Beek, A. J., Dennerlein, J. T., Huysmans, M. A., Mathiassen, S. E., Burdorf, A., Van Mechelen, W., Van Dieën, J. H., Frings-Dresen, M. H. W., Holtermann, A., Janwantanakul, P., Van Der Molen, H. F., Rempel, D., Straker, L., Walker-Bone, K., & Coenen, P. (2017). A research framework for the development and implementation of interventions preventing work-related musculoskeletal disorders. *Scandinavian Journal of Work, Environment and Health*, *43*(6), 526–539. https://doi.org/10.5271/sjweh.3671

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual* (3.8.8). CreateSpace.

Varrecchia, T., De Marchis, C., Rinaldi, M., Draicchio, F., Serrao, M., Schmid, M., Conforto, S., & Ranavolo, A. (2018). Lifting activity assessment using surface electromyographic features and neural networks. *International Journal of Industrial Ergonomics*, *66*, 1–9. https://doi.org/10.1016/j.ergon.2018.02.003

Vu, L. Q., Kim, H., Schulze, L. J. H., & Rajulu, S. L. (2020). Evaluating Lumbar Shape Deformation With Fabric Strain Sensors. *Human Factors*. https://doi.org/10.1177/0018720820965302

Wan, S., Qi, L., Xu, X., Tong, C., & Gu, Z. (2020). Deep learning models for real-time human activity recognition with smartphones. *Mobile Networks and Applications*, *25*, 743–755.

Wang, G., Ren, X. M., Li, L., & Wang, Z. Z. (2007). Multifractal analysis of surface EMG signals for assessing muscle fatigue during static contractions. *Journal of Zhejiang University: Science A*, *8*(6), 910–915. https://doi.org/10.1631/jzus.2007.A0910

Wang, J., Sun, S., & Sun, Y. (2021). A Muscle Fatigue Classification Model Based on LSTM and Improved Wavelet Packet Threshold. *Sensors (Basel, Switzerland)*, *21*(19), 6369.

Wang, M., Zhao, C., Barr, A., Fan, H., Yu, S., Kapellusch, J., & Harris Adamson, C. (2021). Hand Posture and Force Estimation Using Surface Electromyography and an Artificial Neural Network. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *00*(0), 1–21. https://doi.org/10.1177/00187208211016695

Wang, W., Li, H., Kong, D., Xiao, M., & Zhang, P. (2020). A novel fatigue detection method for rehabilitation training of upper limb exoskeleton robot using multi-information fusion. *International Journal of Advanced Robotic Systems*, *17*(6). https://doi.org/10.1177/1729881420974295

Warburton, D., Jamnik, V., Bredin, S., Shephard, R., & Gledhill, N. (2021). The 2021 physical activity readiness questionnaire for everyone (PAR-Q+) and electronic physical activity readiness medical examination (ePARmed-X+): 2021 PAR-Q+. *The Health & Fitness Journal of Canada*, *14*(1), 83–87.

Westerblad, H., Bruton, J. D., & Katz, A. (2010). Skeletal muscle: Energy metabolism, fiber types, fatigue and adaptability. *Experimental Cell Research*, *316*(18), 3093–3099. https://doi.org/10.1016/j.yexcr.2010.05.019

Williamson, A., & Friswell, R. (2013). Fatigue in the workplace: causes and countermeasures. *Fatigue: Biomedicine, Health and Behavior*, *1*(1–2), 81–98. https://doi.org/10.1080/21641846.2012.744581

Wittmann, F., Lambercy, O., & Gassert, R. (2019). Magnetometer-based drift correction during rest in IMU arm motion tracking. *Sensors (Switzerland)*, *19*(6), 13–17. https://doi.org/10.3390/s19061312

Xie, H. B., Guo, J. Y., & Zheng, Y. P. (2010). Fuzzy approximate entropy analysis of chaotic and natural complex systems: Detecting muscle fatigue using electromyography signals. *Annals of Biomedical Engineering*, *38*(4), 1483–1496. https://doi.org/10.1007/s10439-010-9933-5

Yadav, S. K., Tiwari, K., Pandey, H. M., & Akbar, S. A. (2021). A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. *Knowledge-Based Systems*, *223*, 106970. https://doi.org/10.1016/j.knosys.2021.106970

Yan, Y., Fan, H., Li, Y., Hoeglinger, E., Wiesinger, A., Barr, A., O'Connell, G. D., & Harris-Adamson, C. (2021). Applying wearable technology and a deep learning model to predict occupational physical activities. *Applied Sciences*, *11*(20), 9636.

Yang, K., Ahn, C. R., & Kim, H. (2020). Deep learning-based classification of work-related physical load levels in construction. *Advanced Engineering Informatics*, *45*(March), 101104. https://doi.org/10.1016/j.aei.2020.101104

Yassierli, Nussbaum, M. A., Iridiastadi, H., & Wojcik, L. A. (2007). The influence of age on isometric endurance and fatigue is muscle dependent: A study of shoulder abduction and torso extension. *Ergonomics*, *50*(1), 26–45. https://doi.org/10.1080/00140130600967323

Zehr, J. D., Howarth, S. J., & Beach, T. A. C. (2018). Using relative phase analyses and vector coding to quantify Pelvis-Thorax coordination during lifting—A methodological investigation. *Journal of Electromyography and Kinesiology*, *39*(September 2017), 104–113. https://doi.org/10.1016/j.jelekin.2018.02.004

Zemková, E., Poór, O., & Pecho, J. (2019). Peak Rate of Force Development and Isometric Maximum Strength of Back Muscles Are Associated With Power Performance During Load-Lifting Tasks. *American Journal of Men's Health*, *13*. https://doi.org/10.1177/1557988319828622

Zemp, R., Tanadini, M., Plüss, S., Schnüriger, K., Singh, N. B., Taylor, W. R., & Lorenzetti, S. (2016). Application of Machine Learning Approaches for Classifying Sitting Posture Based on Force and Acceleration Sensors. *BioMed Research International*, *2016*. https://doi.org/10.1155/2016/5978489

Zhang, H., Yan, X., & Li, H. (2018). Ergonomic posture recognition using 3D view-invariant features from single ordinary camera. *Automation in Construction*, *94*(June), 1–10. https://doi.org/10.1016/j.autcon.2018.05.033

Zhang, L., Diraneyya, M. M., Ryu, J., Haas, C. T., & Abdel-Rahman, E. M. (2019). Jerk as an indicator of physical exertion and fatigue. *Automation in Construction*, *104*, 120–128.

Zhao, J., & Obonyo, E. (2020). Convolutional long short-term memory model for recognizing construction workers' postures from wearable inertial measurement units. *Advanced Engineering Informatics*, *46*(August), 101177. https://doi.org/10.1016/j.aei.2020.101177

Zurada, J., Karwowski, W., & Marras, W. S. (1997). A neural network-based system for classification of industrial jobs with respect to risk of low back disorders due to workplace design. *Applied Ergonomics*, *28*(1), 49–58. https://doi.org/10.1016/S0003-6870(96)00034-8

Appendix A. Machine learning glossary

A glossary of the machine learning (ML) terminology used throughout this scoping review. Definitions were adapted from sci-kit learn documentation (Pedregosa et al., 2011), Google Machine Learning Glossary (Google, 2021), and the following citations (Harrell Jr, 2015; Miller, 2019). A confusion matrix for binary classification is presented in Fig. A.1.

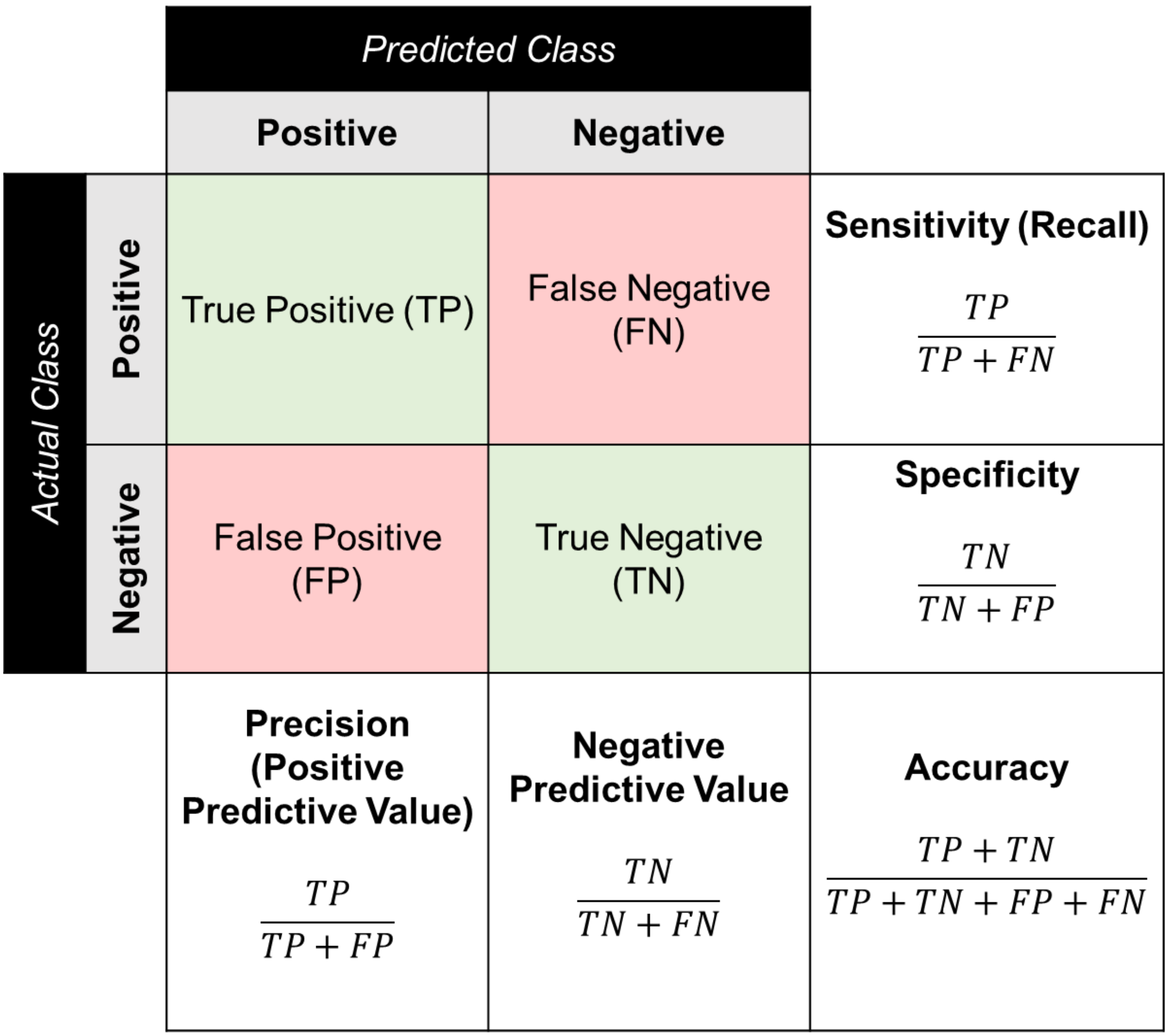


Fig. A.1. A binary classification confusion matrix. Accuracy, sensitivity, specificity, negative predictive value, and precision are performance metrics used to evaluate classification performance.

Accuracy: A performance metric representing the proportion of correct predictions out of the total number of predictions.

Area under the receiver-operator curve (AUC): A performance metric representing the probability that a classifier will be more confident that a randomly chosen positive example is actually positive, than that a randomly chosen negative example is positive.

Artificial neural network (ANN): A model that, taking inspiration from the brain, is composed of layers (at least one of which is hidden) consisting of simple connected units or neurons followed by nonlinearities. Architectures include deep (e.g., convolutional, hourglass), feedforward (e.g., multilayer perceptron), fuzzy, radial basis function, recurrent (e.g., long short-term memory), self-organizing map, amongst others.

Autocoding: The use of ML models to classify free-text into predefined categories. Often used for automatically classifying injury narratives into causation groups.

Bayesian network (BN): A type of probabilistic graphic model that uses Bayesian inference for probability computations. Can capture conditionally dependent and independent relationships between random variables. Also known as Bayesian belief network, belief network, or Bayes network.

Class imbalanced dataset: A dataset used for classification that has significantly different frequencies between the classes.

Classification: A type of model that predicts discrete values.

Cluster analysis: The use of unsupervised machine learning algorithms to automatically group similar samples into sets/clusters. Examples of ML algorithms used for cluster analyses include *k*-means clustering and hierarchical cluster analysis.

Computer vision: A field of study that enables computers to obtain information and understanding from images and/or videos. Discriminative approaches to computer vision utilize ML by relying on models trained using large databases of body parts, joints, or pose examples.

Cross-validation: The technique of testing a model against one or more non-overlapping data subsets withheld from the training set to estimate how well a model will generalize to new data.

Decision tree (DT): A non-parametric supervised learning method for both classification and regression. Numerous algorithms are based on decision trees, such as: alternating decision trees (ADTree); classification and regression trees (CART); chi-square automatic interaction detection (CHAID); or functional trees. Additionally, numerous commonly used ensemble algorithms use DTs, such as random forest (RF), bagged trees, and stochastic gradient tree boosting (SGTB).

F1-score: A performance metric that represents the harmonic mean of precision and sensitivity. Also known as balanced F-score or F-measure.

Features: An input variable used to make predictions.

Gaussian processes (GP): A generic supervised learning method for solving regression and probabilistic classification problems.

Hidden Markov model (HMM): A generative probabilistic model, in which a sequence of an observable variable is generated by a sequence of internal hidden states that cannot be directly observed. The transitions between the hidden states are assumed to have a form of a first-order Markov chain.

Hold-out testing: The technique of allocating a subset of the dataset for testing (i.e., the test set) to estimate how well a learned model will generalize to new data. Testing occurs after training and validation.

Interpretability: The degree to which a human can understand the cause of a decision.

***k*-nearest neighbours (*k*-NN):** A non-parametric algorithm for (un)supervised learning suitable for classification or regression purposes. The predefined number of training samples closest in distance to the new sample are used to predict their label/value.

Label: An output variable of interest.

Linear discriminant analysis: A classifier with a linear decision boundary, generated by fitting class conditional densities to the data.

Logistic regression: A linear model for classification where the probabilities describing the possible outcomes of a single trial are modelled using a logistic function.

Mean absolute error (MAE): The average absolute difference between the expected and predicted values across all training examples.

Multivariate Adaptive Regression Splines (MARS): An ensemble algorithm for regression that finds a set of simple linear functions for modeling complex non-linear relationships.

Naïve bayes (NB): A set of supervised learning algorithms based on applying Bayes' theorem with the "naïve" assumption of conditional independence between every pair of features given the value of the class variable.

Natural language processing (NLP): A field of machine learning to enable computers to understand written or spoken language.

Negative predictive value (NPV): A performance metric for classification models that represents the frequency with which a model was correct when predicting the negative class.

Precision: A performance metric for classification models that represents the frequency with which a model was correct when predicting the positive class. Also referred to as positive predictive value (PPV).

Regression: A type of model that predicts continuous values.

Reinforcement learning: A family of algorithms that learn an optimal policy, whose goal is to maximize return when interacting with an environment.

Root mean square error (RMSE): The square root of the average squared loss per example.

Sensitivity: A performance metric for classification models that represents the proportion of positive labels that the model correctly classified out of all truly positive labels. Also known as recall.

Specificity: A performance metric for classification models that represents the proportion of negative labels that the model correctly classified out of all truly negative labels.

Supervised learning: Training a model from input data and its corresponding labels.

Support vector machine (SVM): A classification algorithm that finds a hyperplane to separate positive and negative classes and seeks to maximize the margin between these classes. A kernel trick can be used to map input data vectors to a higher dimensional space to separate non-linearly separable classes.

Topic models: Models used for discovering topics occurring in text documents, enabled by techniques such as latent dirichlet allocation and latent semantic analysis.

Training set: A proportion of the dataset used for learning a ML model only. In supervised learning contexts, training data will include features and labels for a sample.

Unsupervised learning: Training a model to find patterns in a (typically unlabelled) dataset.

Validation: A process used during training to evaluate the quality of an ML model using a validation set.

Appendix B. Literature search strategies

Search strategy for Medline and CENTRAL, accessed using OVID:

exp Machine Learning/ or pattern recognition.ti,ab,kw. or pattern classification.ti,ab,kw. or supervised learning.ti,ab,kw. or unsupervised learning.ti,ab,kw. or semi-supervised learning.ti,ab,kw. or reinforcement learning.ti,ab,kw. or deep learning.ti,ab,kw. or transfer learning.ti,ab,kw. or cluster analys.ti,ab,kw. or data min*.ti,ab,kw. or neural network*.ti,ab,kw. or decision tree*.ti,ab,kw. or support vector machine*.ti,ab,kw. or naive bayes*.ti,ab,kw. or k-nearest neighbo?r*.ti,ab,kw. or k-means cluster*.ti,ab,kw. or random forest*.ti,ab,kw. or gradient boost*.ti,ab,kw. or latent dirichlet allocation.ti,ab,kw. or text min*.ti,ab,kw. or discriminate analys*.ti,ab,kw. or optical character recognition*.ti,ab,kw. or chi-square automatic interaction detection*.ti,ab,kw. or computer vision.ti,ab,kw. or marker?less motion capture*.ti,ab,kw. or adaptive logic network*.ti,ab,kw. or gaussian mixture*.ti,ab,kw. or hierarchical cluster*.ti,ab,kw. or hidden markov*.ti,ab,kw. or multilayer perceptron*.ti,ab,kw. or self-organizing map*.ti,ab,kw. or mean-shift cluster*.ti,ab,kw. or density-based spatial clustering of applications with noise*.ti,ab,kw. or expectation-maximization cluster*.ti,ab,kw. or autoencoder*.ti,ab,kw. or long short-term memory.ti,ab,kw. or deep deterministic policy gradient*.ti,ab,kw. or state-action-reward-state-action*.ti,ab,kw. or deep belief network*.ti,ab,kw. or generative adversarial network*.ti,ab,kw. or restricted Boltzmann machine*.ti,ab,kw. or Skip-gram*.ti,ab,kw. or Q-learning*.ti,ab,kw. or deep Q network*.ti,ab,kw. or pose estimation*.ti,ab,kw. or Markov decision process*.ti,ab,kw. AND exp Cumulative Trauma Disorders/ or exp Occupational Injuries/ or exp Biomechanical Phenomena/ or musculoskeletal disorder*.ti,ab,kw or overuse Injur*.ti,ab,kw. or repetiti* strain injur*.ti,ab,kw. or repetiti* motion disorder*.ti,ab,kw. or repetiti* strain Injur*.ti,ab,kw. or overuse syndrome*.ti,ab,kw. or repetiti* stress injur*.ti,ab,kw. or cumulative trauma disorder*.ti,ab,kw. or soft tissue injur*.ti,ab,kw. or injur*.ti,ab,kw. or injury surveillance.ti,ab,kw. or injury rate*.ti,ab,kw. or injury incidence*.ti,ab,kw. or injury prevalence*.ti,ab,kw. or musculoskeletal disorder* incidence*.ti,ab,kw. or musculoskeletal disorder* prevalence*.ti,ab,kw. or injury epidemiolog*.ti,ab,kw. or musculoskeletal disorder* epidemiolog*.ti,ab,kw. or injury predict*.ti,ab,kw. or musculoskeletal disorder* predict*.ti,ab,kw. or injury risk factor*.ti,ab,kw. or musculoskeletal disorder* risk factor*.ti,ab,kw. or injury prevent*.ti,ab,kw. or musculoskeletal disorder* prevent*.ti,ab,kw or joint load*.ti,ab,kw. or joint moment*.ti,ab,kw. or joint torque*.ti,ab,kw. or joint compression*.ti,ab,kw. or spin* load*.ti,ab,kw. or spin* compression*.ti,ab,kw. or lumbosacral moment.ti,ab,kw. or elbow torque.ti,ab,kw. or knee torque*.ti,ab,kw. or fatigu*.ti,ab,kw. or safe* behavio?r*.ti,ab,kw. AND exp Work/ or exp Ergonomics/ or exp Occupations/ or occupation*.ti,ab,kw. or work-related.ti,ab,kw. or job-relate*.ti,ab,kw. or on-the-job.ti,ab,kw. or military*.ti,ab,kw. or police*.ti,ab,kw. or paramedic*.ti,ab,kw. or firefight*.ti,ab,kw. or electrician*.ti,ab,kw. or painter*.ti,ab,kw. or construction*.ti,ab,kw. or maintenance work*.ti,ab,kw. or driver*.ti,ab,kw. or air* pilot*.ti,ab,kw. or healthcare professional*.ti,ab,kw. or doctor*.ti,ab,kw. or physician*.ti,ab,kw. or nurs*.ti,ab,kw. or therapist*.ti,ab,kw. or physiotherapist*.ti,ab,kw. or factory work*.ti,ab,kw. or industrial work*.ti,ab,kw. or equipment operat*.ti,ab,kw. or labo?rer*.ti,ab,kw. or manual material? handl*.ti,ab,kw. or load* handl*.ti,ab,kw. or repair work*.ti,ab,kw. or farm*.ti,ab,kw. or ranch*.ti,ab,kw. or iron work*.ti,ab,kw. or steel work*.ti,ab,kw. or garbage collect*.ti,ab,kw. or recycling collect*.ti,ab,kw. or compost collect*.ti,ab,kw. or roofer*.ti,ab,kw. or logging work*.ti,ab,kw. or fishing work*.ti,ab,kw. or office work*.ti,ab,kw. or seated work*.ti,ab,kw. or desk work*.ti,ab,kw. or mining work*.ti,ab,kw. AND limit to (english language and yr="1990 -Current")*

Search strategy for Embase, accessed using OVID:

exp Machine Learning/ or pattern recognition.ti,ab,kw. or pattern classification.ti,ab,kw. or supervised learning.ti,ab,kw. or unsupervised learning.ti,ab,kw. or semi-supervised learning.ti,ab,kw. or reinforcement learning.ti,ab,kw. or deep learning.ti,ab,kw. or transfer learning.ti,ab,kw. or cluster analys.ti,ab,kw. or data min*.ti,ab,kw. or neural network*.ti,ab,kw. or decision tree*.ti,ab,kw. or support vector machine*.ti,ab,kw. or naive bayes*.ti,ab,kw. or k-nearest neighbo?r*.ti,ab,kw. or k-means cluster*.ti,ab,kw. or random forest*.ti,ab,kw. or gradient boost*.ti,ab,kw. or latent dirichlet allocation.ti,ab,kw. or text min*.ti,ab,kw. or discriminate analys*.ti,ab,kw. or optical character recognition*.ti,ab,kw. or chi-square automatic interaction detection*.ti,ab,kw. or computer vision.ti,ab,kw. or marker?less motion capture*.ti,ab,kw. or adaptive logic network*.ti,ab,kw. or gaussian mixture*.ti,ab,kw. or hierarchical cluster*.ti,ab,kw. or hidden markov*.ti,ab,kw. or multilayer perceptron*.ti,ab,kw. or self-organizing map*.ti,ab,kw. or mean-shift cluster*.ti,ab,kw. or density-based spatial clustering of applications with noise*.ti,ab,kw. or expectation-maximization cluster*.ti,ab,kw. or autoencoder*.ti,ab,kw. or long short-term memory.ti,ab,kw. or deep deterministic policy gradient*.ti,ab,kw. or state-action-reward-state-action*.ti,ab,kw. or* *deep belief network*.ti,ab,kw. or generative adversarial network*.ti,ab,kw. or restricted Boltzmann machine*.ti,ab,kw. or Skip-gram*.ti,ab,kw. or Q-learning*.ti,ab,kw. or deep Q network*.ti,ab,kw. or pose estimation*.ti,ab,kw. or Markov decision process*.ti,ab,kw. AND exp Cumulative Trauma Disorder/ or exp Occupational accident/ or exp Biomechanics/ or musculoskeletal disorder*.ti,ab,kwor overuse Injur*.ti,ab,kw. or repetiti* strain injur*.ti,ab,kw. or repetiti* motion disorder*.ti,ab,kw. or repetiti* strain Injur*.ti,ab,kw. or overuse syndrome*.ti,ab,kw. or repetiti* stress injur*.ti,ab,kw. or cumulative trauma disorder*.ti,ab,kw. or soft tissue injur*.ti,ab,kw. or injur*.ti,ab,kw. or injury surveillance.ti,ab,kw. or injury rate*.ti,ab,kw. or injury incidence*.ti,ab,kw. or injury prevalence*.ti,ab,kw. or musculoskeletal disorder* incidence*.ti,ab,kw. or musculoskeletal disorder* prevalence*.ti,ab,kw. or injury epidemiolog*.ti,ab,kw. or musculoskeletal disorder* epidemiolog*.ti,ab,kw. or injury predict*.ti,ab,kw. or musculoskeletal disorder* predict*.ti,ab,kw. or injury risk factor*.ti,ab,kw. or musculoskeletal disorder* risk factor*.ti,ab,kw. or injury prevent*.ti,ab,kw. or musculoskeletal disorder* prevent*.ti,ab,kw. or joint load*.ti,ab,kw. or joint moment*.ti,ab,kw. or joint torque*.ti,ab,kw. or joint compression*.ti,ab,kw. or spin* load*.ti,ab,kw. or spin* compression*.ti,ab,kw. or lumbosacral moment.ti,ab,kw. or elbow torque.ti,ab,kw. or knee torque*.ti,ab,kw. or fatigu*.ti,ab,kw. or safe* behavio?r*.ti,ab,kw. AND exp Work/ or exp Ergonomics/ or exp Occupation/or occupation*.ti,ab,kw. or work-related.ti,ab,kw. or job-relate*.ti,ab,kw. or on-the-job.ti,ab,kw. or military*.ti,ab,kw. or police*.ti,ab,kw. or paramedic*.ti,ab,kw. or firefight*.ti,ab,kw. or electrician*.ti,ab,kw. or painter*.ti,ab,kw. or construction*.ti,ab,kw. or maintenance work*.ti,ab,kw. or driver*.ti,ab,kw. or air* pilot*.ti,ab,kw. or healthcare professional*.ti,ab,kw. or doctor*.ti,ab,kw. or physician*.ti,ab,kw. or nurs*.ti,ab,kw. or therapist*.ti,ab,kw. or physiotherapist*.ti,ab,kw. or factory work*.ti,ab,kw. or industrial work*.ti,ab,kw. or equipment operat*.ti,ab,kw. or labo?rer*.ti,ab,kw. or manual material? handl*.ti,ab,kw. or load* handl*.ti,ab,kw. or repair work*.ti,ab,kw. or farm*.ti,ab,kw. or ranch*.ti,ab,kw. or iron work*.ti,ab,kw. or steel work*.ti,ab,kw. or garbage collect*.ti,ab,kw. or recycling collect*.ti,ab,kw. or compost collect*.ti,ab,kw. or roofer*.ti,ab,kw. or logging work*.ti,ab,kw. or fishing work*.ti,ab,kw. or office work*.ti,ab,kw. or seated work*.ti,ab,kw. or desk work*.ti,ab,kw. or mining work*.ti,ab,kw. AND limit to (english language and yr="1990 -Current")*

Search strategy for PsycInfo, accessed using OVID:

exp Machine Learning/ or pattern recognition.tw or pattern classification.tw or supervised learning.tw or unsupervised learning.tw or semi-supervised learning.tw or reinforcement learning.tw or deep learning.tw or transfer learning.tw or cluster analys.tw or data min*.tw or neural network*.tw or decision tree*.tw or support vector machine*.tw or naive bayes*.tw or k-nearest neighbo?r*.tw or k-means cluster*.tw or random forest*.tw or gradient boost*.tw or latent dirichlet allocation.tw or text min*.tw or discriminate analys*.tw or optical character recognition*.tw or chi-square automatic interaction detection*.tw or computer vision.tw or marker?less motion capture*.tw or adaptive logic network*.tw or gaussian mixture*.tw or hierarchical cluster*.tw or hidden markov*.tw or multilayer perceptron*.tw or self-organizing map*.tw or mean-shift cluster*.tw or density-based spatial clustering of applications with noise*.tw or expectation-maximization cluster*.tw or autoencoder*.tw or long short-term memory.tw or deep deterministic policy gradient*.tw or state-action-reward-state-action*.tw or deep belief network*.tw or generative adversarial network*.tw or restricted Boltzmann machine*.tw or Skip-gram*.tw or Q-learning*.tw or deep Q network*.tw or pose estimation*.tw or Markov decision process*.tw AND exp Musculoskeletal Disorders/ or exp Injuries/ or exp Biomechanics/ or musculoskeletal disorder*.tw or overuse Injur*.tw or repetiti* strain injur*.tw or repetiti* motion disorder*.tw or repetiti* strain Injur*.tw or overuse syndrome*.tw or repetiti* stress injur*.tw or cumulative trauma disorder*.tw or soft tissue injur*.tw or injur*.tw or injury surveillance.tw or injury rate*.tw or injury incidence*.tw or injury prevalence*.tw or musculoskeletal disorder* incidence*.tw or musculoskeletal disorder* prevalence*.tw or injury epidemiolog*.tw or musculoskeletal disorder* epidemiolog*.tw or injury predict*.tw or musculoskeletal disorder* predict*.tw or injury risk factor*.tw or musculoskeletal disorder* risk factor*.tw or injury prevent*.tw or musculoskeletal disorder* prevent*.tw or joint load*.tw or joint moment*.tw or joint torque*.tw or joint compression*.tw or spin* load*.tw or spin* compression*.tw or lumbosacral moment.tw or elbow torque.tw or knee torque*.tw or fatigu*.tw or safe* behavio?r*.tw AND exp Human Factors Engineering/ or exp Occupations/or occupation*.tw or work-related.tw or job-relate*.tw or on-the-job.tw or military*.tw or police*.tw or paramedic*.tw or firefight*.tw or electrician*.tw or painter*.tw or construction*.tw or maintenance work*.tw or driver*.tw or air* pilot*.tw or healthcare professional*.tw or doctor*.tw or physician*.tw or nurs*.tw or therapist*.tw or physiotherapist*.tw or factory work*.tw or industrial work*.tw or equipment operat*.tw or labo?rer*.tw or manual material? handl*.tw or load* handl*.tw or repair work*.tw or farm*.tw or ranch*.tw or iron work*.tw or steel work*.tw or garbage collect*.tw or recycling collect*.tw or compost collect*.tw or roofer*.tw or* *logging work*.tw or fishing work*.tw or office work*.tw or seated work*.tw or desk work*.tw or mining work*.tw AND limit to (english language and yr="1990 -Current")*

Search strategy for IEEE Xplore (searching metadata, journals only, English only, 1990-current):

("All Metadata":"machine learning" OR "neural network" OR "cluster analys*") AND*
("All Metadata":"injur" OR "musculoskeletal disorders" OR "cumulative trauma disorders" OR "joint load*" OR "fatigue" OR "safety behavior") AND*
("All Metadata":"occupation" OR "work-related" OR "ergonomic*" OR "handl*")*

Search strategy for SPORTDiscus, performed four times to search titles, abstracts, subjects, and headings (Boolean/Phrase searching; applying related words, applying equivalent subjects, English only, peer-reviewed only, from January 1990 and on):

machine learning OR "neural network" OR "cluster analys*" AND*
injur OR "musculoskeletal disorder*" OR "cumulative trauma disorder*" OR "joint load*" OR "fatigu*" OR "safety behavio?r" AND*
"occupation" OR "work-related" OR "ergonomic*" OR "handl*"*

Search strategy for Scopus:

TITLE-ABS-KEY(“Machine Learning” OR “pattern recognition” OR “pattern classification” OR “supervised learning” OR “unsupervised learning” OR “semi-supervised learning” OR “reinforcement learning” OR “deep learning” OR “transfer learning” OR “cluster analys” OR “data min*” OR “neural network*” OR “decision tree*” OR “support vector machine*” OR “naive bayes*” OR “k-nearest neighbo?r*” OR “k-means cluster*” OR “random forest*” OR “gradient boost*” OR “latent dirichlet allocation” OR “text min*” OR “discriminate analys*” OR “optical character recognition*” OR “chi-square automatic interaction detection*” OR “computer vision” OR “marker?less motion capture*” OR “adaptive logic network*” OR “gaussian mixture*” OR “hierarchical cluster*” OR “hidden markov*” OR “multilayer perceptron*” OR “self-organizing map*” OR “mean-shift cluster*” OR “Density-Based Spatial Clustering of Applications with Noise*” OR “Expectation-Maximization Cluster*” OR “Autoencoder*” OR “Long Short-Term Memory” OR “Deep Deterministic Policy Gradient*” OR “State-Action-Reward-State-Action*” OR “Deep Belief Network*” OR “Generative Adversarial Network*” OR “Restricted Boltzmann Machine*” OR “Skip-gram*” OR “Q-learning*” OR “deep Q network*” OR “pose estimation*” OR “Markov Decision Process*”) AND TITLE-ABS-KEY(“Cumulative Trauma Disorders” OR “Occupational Injuries” OR “Biomechanical Phenomena” OR “musculoskeletal disorder*” OR “Overuse Injur*” OR “Repetiti* strain injur*” OR “Repetiti* motion disorder*” OR “Repetiti* Strain Injur*” OR “Overuse Syndrome*” OR “Repetiti* Stress Injur*” OR “Cumulative Trauma Disorder*” OR “soft tissue injur*” OR “injur*” OR “injury surveillance” OR “injury rate*” OR “injury incidence*” OR “injury prevalence*” OR “musculoskeletal disorder* incidence*” OR “musculoskeletal disorder* prevalence*” OR “injury epidemiolog*” OR “musculoskeletal disorder* epidemiolog*” OR “injury predict*” OR “musculoskeletal disorder* predict*” OR “injury risk factor*” OR “musculoskeletal disorder* risk factor*” OR “injury prevent*” OR “musculoskeletal disorder* prevent*” OR “joint load*” OR “joint moment*” OR “joint torque*” OR “joint compression*” OR “spin* load*” OR “spin* compression*” OR “lumbosacral moment” OR “elbow torque” OR “knee torque*” OR “fatigu*” OR “safe* behavio?r*”) AND TITLE-ABS-KEY(“Work” OR “Ergonomics” OR “occupation*” OR “work-related” OR “job-relate*” OR “on-the-job” OR “military*” OR “police*” OR “paramedic*” OR “firefight*” OR “electrician*” OR “painter*” OR “construction*” OR “maintenance work*” OR “driver*” OR “air* pilot*” OR “healthcare professional*” OR “doctor*” OR “physician*” OR “nurs*” OR “therapist*” OR “physiotherapist*” OR “factory work*” OR “industrial work*” OR “equipment operat*” OR “labo?rer*” OR “manual material? handl*” OR “load* handl*” OR “repair work*” OR “farm*” OR “ranch*” OR “iron work*” OR “steel work*” OR “garbage collect*” OR “recycling collect*” OR “compost collect*” OR “roofer*” OR “logging work*” OR “fishing work*” OR “office work*” OR “seated work*” OR “desk work*” OR “mining work*”) AND PUBYEAR AFT 1989 AND (LIMIT-TO (PUBSTAGE , "final")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SRCTYPE , "j"))*

Appendix C. Machine learning study summaries tables

Table C.1. Studies where machine learning (ML) techniques were applied to understand the incidence and severity of work-related musculoskeletal disorders (WMSDs; step 1 in the van der Beek et al., 2017 research framework). **Bolded** machine learning (ML) technique(s) indicates selection as best (evaluated by respective authors). ANN = artificial neural network; NB = naïve Bayes classifier; NLP = natural language processing; RF = random forest; SGTB = stochastic gradient tree boosting; SVM = support vector machine; ✓ = indicates code or data are publicly available (potentially through request).

Citation	Purpose(s) of ML	Sample	Data	ML technique(s) used	Application	Outcome	Code	Data
Bertke et al. (2012)	Develop and evaluate an autocoding method to aid the manual coding of the Ohio Bureau of Workers' Compensation as musculoskeletal disorder, slip trips or falls, or other.	10,132 claims from OBWC-insured, single location employers.	2,240 claims used for training; 160 claims removed (disagreed & not classified); 7,732 claims used as a test set; 800 claims used for evaluation.	**NB**	Multiclass classification	It took less than 5 minutes to auto-code 7,732 claims. When using only the unstructured text narrative, the auto-coding method predicted 88.4% of claims correctly, when using structured injury category code, the accuracy improved to 89.9%.	-	-
Bertke et al. (2016)	1) Compare performance of ML techniques to code injury cause in free-text narratives from workers' compensation claims; 2) investigate the performance of adding 2-word sequences into a single model; 3) demonstrate	7,200 claims from 2001-2009 from the OHBWC database.	6,200 claims used as a training set; 1,000 claims used as a test set.	NB, **regularized logistic regression**	Multiclass classification	The logistic regression model was 5% better than NB for coding the 2-digits for common event/exposures. Using 2-word keywords improved performance of both models. Limited evaluation showed that outside researchers achieved similar performance without learning their own weights.	-	-

	the feasibility of sharing the pre-trained model with an outsider researcher.							
Brooks (2008)	Apply the SAS Text Miner to the free-text component of worker compensation claims data with reference to the Victorian System.	Wood industry workers and car and delivery drivers.	Two sets of 2500-3000 claims for workers in Victoria, Australia from 1992-2002.	**Expectation maximization clustering**	Clustering	Claims records could be grouped using only the free-text describing the accidents and injuries that occurred themselves. The noise was in the order of 6-8%.	-	-
Davoudi Kakhki et al. (2019)	Test the performance of ML techniques in modeling and predicting occupational incidents severity in agribusiness industries using workers' compensation claims.	33,458 claims from a USA insurance company.	16 variables from 33,458 claims including severity (> or < $10,000); 70% used for training (23,421); 30% used for the testing (10,037).	**SVM** (**RBF,** linear, quadratic kernels**)**, boosted trees, NB	Binary classification	The SVM with RBF kernel achieved the greatest performance with ≥ 89.55% for recall, specificity, precision, F-score, and overall accuracy. Using demographic information and injury details from workers' compensation claims data, this model predicted injury severity of a large number agrindustry claims.	-	-
Harber and Leroy (2017)	Assess feasibility and potential utility of NLP for storing and analyzing occupational health data.	89,084 occupational injury and illness records from 2002-2014.	MSHA Form 7000 with fixed format fields and free text fields for injury and workplace factor description.	**Stanford part-of-speech tagger**	Multiclass classification	Employing a part-of-speech tagger for NLP enabled the automated analyses of free text in occupational injury records.	-	✓

Landsittel et al. (1998)	Investigate the accuracy of ML techniques for identifying populations at high risk for occupational back injury.	Simulated data sets (100,000 samples) based on known associations between the presence of back injury and the predictor variables (2 sets: random and non-linear associations).	Simulated data with four different predictor variables (experience in years, BMI, % time lifting, % time in non-neutral trunk postures).	ANN, **logistic regression** (**main effects**, fully-specified)	Binary classification	Logistic regression model with main effects led to the greatest accuracy on the test set (71%). This method could be used to identify populations at high risk for occupational back injuries using 4 variables.	-	-
Lehto et al. (2009)	Compare two Bayesian methods for classifying injury narratives in administrative datasets into event cause groups.	14,000 injury narratives extracted from claims filed with a workers' compensation insurance provider.	Manually-coded injury narratives; 3000 used as a test set.	**Fuzzy bayes, NB**	Multiclass classification	NB led to slightly more accurate predictions than Fuzzy bayes; one-digit BLS codes were predicted more accurately than two-digit codes. This work supports the use of semi-automatic approaches to classify injury narratives to improve speed and accuracy.	-	-
Marucci-Wellman et al. (2011)	1) Develop models to automatically assign event codes to injury narratives; and 2) compare coding accuracy of two, semi-computerized strategies.	14,000 injury narratives extracted from claims filed with a workers' compensation insurance provider.	Manually-coded injury narratives; 3000 used as a test set.	**Fuzzy bayes, NB**	Multiclass classification	Models agreed on 64% of narratives; best sensitivities when leaving 50% of coded data to manual review, selected based on lack of prediction strength.	-	-

Marucci-Wellman et al. (2017)	1) Test and compare the practicality and performance of a human-machine combined approach to classify short injury narratives; and 2) test the utility of NLP rules to identify additional narratives in categories.	Workers who filed a worker compensation claim between January 1 and December 31, 2007	30,000 worker compensation insurance claims.	**NB, SVM,** single word and Bi-gram models, and logistic regression	Multiclass classification	NLP used on MSD reports in a human-machine combined approach can classify injuries according to a specified format for the Bureau of Labor Statistics and reduce manual review down to 41% of the narratives. Alone sensitivity = 0.89; together sensitivity = 0.93.	✓	-
Marucci-Wellman et al. (2015)	Develop a human-machine system that could be tailored to classify injury narratives routinely and accurately from large administrative databases.	30,000 records randomly extracted from claims filed with a workers' compensation insurance provider between January 1 and December 31, 2007.	15,000 records for training; 15,000 records for testing.	**NB** (×2): single word and sequence word models	Multiclass classification	NB models can be used to accurately classify the events leading to injury. A benefit of using NB is that low confidence can be filtered out for manual review.	-	-
Meyers et al. (2018)	Improve and apply automated methods to code 1.2 million Ohio Bureau of Workers' compensation	8600 worker compensation claims in Ohio from 2001-2011.	Worker compensation claims with a 57-diagnosis category variable assigned to each claim based on	**NB**	Multiclass classification	The NB algorithm achieved accuracy = 90%. Their methods were used to auto-code more than 1.2 million worker compensation claims, determine the sections they occurred, and prioritize ergonomic and safety intervention efforts.	-	-

	claims into three intervention categories.		the Optimal RTW ICD-9-CM diagnosis code.					
Min et al. (2019)	Analyze social media data using topic modeling to explore illegal compensation for occupational injuries related to underreporting (i.e., gong-sang) in South Korea.	2210 documents from social media data using the keyword "gong-sang" between January 1, 2006-Devember 31, 2017.	Gong-sang related documents.	**Latent Dirichlet allocation**	Multiclass classification	Social media documents could be classified into 4 categories with 10 topics extracted using the latent dirichlet allocation model. Providing insight into gong-sang practices provides a more accurate assessment of occupational injury occurrence.	-	-
Nanda et al. (2016)	Develop a system to autocode occupational lost work-day cases, filter cases requiring manual review, and assist human coders by suggesting the top five choices for possible events.	Workers from the USA with injury resulting in low workdays.	50,000 lost workday cases from the US Bureau of Labor Statistics (2011 SOII data); 10,000 used for testing.	Single-word NB, **two-word sequence NB**	Multiclass classification	The developed, two-word sequence NB-based system results in a less manual work with acceptable accuracy and flagging reports for manual review with suggestions.	-	-
Tixier et al. (2016)	To extract attributes and categorical safety outcomes from a large pool of textual construction	4,398 injury reports gathered from > 470 contractors in industrial, energy, infrastructure,	Injury report data with unstructured text.	**RF, SGTB**	Multiclass classification	Using construction injury reports, a model based on RF and SGTB predicted injury and energy type, and body part (rank probability skill score = 0.236 - 0.436), but not injury severity. Injury severity in this dataset may have been random or insufficient data. NLP performance: accuracy =	-	-

	injury reports using NLP and employ ML to predict safety outcomes.	and mining work from around the world.		96%, precision = 95%, recall = 97% for attributes, error rate = 5.7% for energy type and injury code.

Table C.2. Studies where machine learning (ML) techniques were applied to understand the risk factors for work-related musculoskeletal disorders (WMSDs; step 2 in the van der Beek et al., 2017 research framework). **Bolded** machine learning (ML) technique(s) indicates selection as best (evaluated by respective authors). ADTree = alternating decision tree; ANN = artificial neural network; BN = Bayesian network; CART = classification and regression tree; CHAID = chi-square automatic interaction detection; DT = decision tree; EMG = electromyography; HCA = hierarchical cluster analysis; LDA = linear discriminant analysis; MARS = multivariate adaptive regression splines; MLP = multilayer perceptron; MSD = musculoskeletal disorder; NN = neural network; RBF = radial basis function; RF = random forest; SVM = support vector machine; ✓ = indicates code or data are publicly available (potentially through request).

Citation	**Purpose(s) of ML**	**Sample**	**Data**	**ML technique(s) used**	**Application**	**Outcome**	**Code**	**Data**
Abubakar et al. (2020)	Model relationship between safety climate, safety behaviour, and workplace injuries in metal casting industry employees from central Anatolia.	306 metal casting industry employees from central Anatolia.	Questionnaire data regarding safety climate, safety behaviour, workplace injuries, demographics, formal education, and frequency of accidents; 75% and 25% used for training and testing, respectively.	**ANN**	Regression	Safety climate negatively predicts workplace injuries, and safety behaviours strengthen that relationship and negatively predict workplace injuries on their own.	✔	-
Ahn et al. (2020)	Analyze the impact of working characteristics on WMSDs and develop a diagnosis model.	22,262 injury records from the fourth Korean working conditions survey from the Korea	Injury records with working characteristic data.	**BN**, ANN, SVM, DT	Multiclass classification	A BN established the effect of various working characteristics on WMSD risk (e.g., working hours, pace). Furthermore, a BN led to the strongest	-	-

		Occupational Safety and health Agency.				performance in probabilistically diagnosing WMSD.		
Akay (2011)	Classify industrial jobs as low or high risk for WMSDs.	235 MMH tasks from different manufacturing industries.	During the performance of various MMH jobs, lift rate, peak twist velocity average, peak moment, peak sagittal angle, and peak lateral velocity maximum. 148 and 87 jobs in training and test sets, respectively.	**Grey relational analysis***, logistic regression, DT, ANN, ensemble model	Binary classification	Classifying industrial jobs by their risk of LBDs can help identify which occupations require the implementation of interventions. Grey relational analysis was able to correctly classify jobs as low or high risk with at least 10% greater accuracy than select ML classifiers.	-	-
Asensio-Cuesta et al. (2010)	Classify industrial jobs as low or high risk for WMSDs.	235 MMH tasks from different manufacturing industries (Marras et al., 1993).	During the performance of various MMH jobs, lift rate, peak twist velocity average, peak moment, peak sagittal angle, and peak lateral velocity maximum. 94, 54, and 87 jobs for training, validation, and testing, respectively.	**Feedforward NN**	Binary classification	Classifying industrial jobs by their risk of LBDs can help identify which occupations require the implementation of interventions. Feedforward NN classified jobs as low or high risk with accuracy = 81.6%.	-	-
Atal et al. (2020)	Explore the use of a Bayesian Network to predict the risk factors for whole	79 dumper operators of two opencast coal	Measurement of WBV exposure, interview with a questionnaire	**BN**	Regression	The causal dependence of variables contributing to WBV-related WMSD risk were uncovered using a	-	-

	body vibration in dumper operators for the development of WMSD symptoms.	mines in Northern India.	regarding personal factors, anthropometrics, machine related data.			BN. Vibration in the vertical direction was a dominant factor; relationships with other variables were also established.		
Bodin et al. (2017)	1) Identify forms of work organization characterized by patterns of organizational and psychosocial variables in French workers; 2) compare symptomatic and clinically-diagnosed shoulder disorders in the different work organization forms.	Self-administered questionnaires from 3710 workers French workers.	Questionnaires with 16 organizational variables; presence of non-specific shoulder pain in the past 12-months.	**HCA**	Clustering	5 work organization clusters were identified, and logistic regression showed differences between the groups. The group with low decision latitude with pace constraints had more shoulder pain than the other clusters. This approach identified work organization factors associated with MSDs.	-	-
Busto Serrano et al. (2020)	Construct models to identify variables with the strongest effect on shoulder, neck, and upper limb WMSD.	43,500 employees (21,578F/21,922M) from the Eurofound dataset conducted in 2015.	51 variables from questionnaires, organized in 5 groups: individual factors and health status, physical, psychosocial, and other work-related risk factors, and activities outside working hours.	**MARS** (×2): one each for males and females	Regression	The variables most associated with shoulder, neck, and upper limb MSDs were different between male and female employees: males were affected most by physical and other work-related risk factors, whereas females were most affected by psychosocial risk factors and activities outside working hours.	-	✔
Chen et al. (2019)	Investigate which clinical consequences are associated with	3 groups of patients aged 18 years and older: workplace injuries	The SID, SASD and SEDD, containing anonymized,	**Logistic regression**, RF, ADTree	Binary classification	Associations been workplace injuries and clinical morbidities were assessed using regression;	-	✔

	workplace injury and to explore how sociodemographic and clinical variables modify the risk of injury recurrence.	(80,712); injury from another place (161,424); and no injuries (161,424).	longitudinally-linked inpatient, outpatient, and emergency room visit data, including patients' demographics, insurance, diagnoses, and procedures for each hospital visit.			then predictive models were developed to predict injury recurrence (AUROC = 0.60, c-index = 0.62).		
Chen et al. (2000)	Predict the level of low-back disorder risk associated with industrial jobs.	235 MMH tasks from different manufacturing industries (Marras et al., 1993).	During the performance of various MMH jobs, lift rate, peak twist velocity average, peak moment, peak sagittal angle, and peak lateral velocity maximum.	**Feedforward NN**	Binary classification	Their approach to a similar problem showed 4 and 5% classification improvement compared to previous studies.	-	-
Chen et al. (2004)	Develop a method to accurately predict the risk of injuries in industrial jobs based on datasets either not meeting the assumptions of parametric statistical tools or being incomplete.	Dataset developed by Marras et al (1993) of kinematic data of workers and observations on 235 industrial jobs.	Injury risk-level of each job; observation of the industrial jobs; kinematic data of workers; 100 samples for training; 48 samples for validation.	**Feedforward NN**	Binary classification	ML models can map the relationship between risk factors and LBDs. Forward selection of input variables improved the effectiveness of NN when working with missing data, preventing overspecification and/or overfitting an FNN to training data.	-	-
Connaboy et al. (2018)	Investigate if interactions between individual predictors could be	140 Air Force Special Forces Operators.	BOD POD; Parvo Medics metabolic cart; Polar HR	**CHAID**	Regression	Military personnel who sustained a lower-extremity injury over the following year had greater	-	-

	used to develop a model of lower-extremity injury risk specific to Air Force Special Forces Operators.		monitor; ramp treadmill; Velotron cycle ergometer; Biodex Multi-Joint System 3 Pro; inclinometer and/or goniometer;2x force plates sampling at 1200Hz; 8x infrared motion capture cameras (Vicon) sampling at 200Hz to measure kinematics during single-leg landing.			maximum knee flexion angle asymmetry, peak anaerobic power, and smaller left ankle inversion range of motion. The model had an overall injury classification accuracy = 87.143% and AUC = 0.913.		
Engkvist (2004)	Investigate factors involved in the accident process preceding reported over-exertion back in nursing personnel in Melbourne, Australia.	127 nursing personnel with an over-exertion back injury in the past 13-months who filed a staff incident report or a workers' compensation claim.	Structured interviews from 2000-2001; ergonomic checklists.	**HCA** with average linkage-within-groups	Clustering	5 distinct clusters of accident processes were defined, and the factors contributing to the accident processes could be identified. Prominent factors were working in awkward positions despite opportunities to work in safer postures, and lacking knowledge in transfer techniques prior to transferring heavy patients.	-	-
Engkvist et al. (1998)	Investigate factors involved in the accident process preceding overexertion back	130 nursing personnel with an overexertion back injury due to an accident.	Structured interviews with 84 questions; ergonomic checklists.	**HCA** with average linkage-within-groups	Clustering	6 distinct clusters of accident processes were found. The extent to which 22 different factors contributed to each cluster	-	-

	injuries among nursing personnel.					were determined. Most injuries occurred during patient transfer, where transfer devices were seldom used, and a lack of technique training was present in most clusters.		
Engkvist et al. (2001)	Analyze injury history and related exposure factors amongst Swedish nurses.	673 female nursing personnel employed in Stockholm County hospitals from March 1992-December 1994 (220 with LBDs and 453 referents).	Questionnaires with 124 questions regarding exposure factors and injury history.	**HCA** with average linkage-within-groups	Clustering	6 distinct clusters were identified. The proportion of cases differed significantly between clusters (X^2 = 58.08 df = 5, p < 0.001). MSDs amongst Swedish nurses were associated with greater working hours, variable working schedules, and more patient handling.	-	-
Ganga et al. (2012)	Classify industrial jobs as low or high risk for WMSDs.	235 MMH tasks from different manufacturing industries (Marras et al., 1993).	During the performance of various MMH jobs, lift rate, peak twist velocity average, peak moment, peak sagittal angle, and peak lateral velocity maximum.	LDA, **ANN**	Binary classification	Classifying industrial jobs by their risk of LBDs can help identify which occupations require interventions. An LDA provided statistically similar performance to a NN, while being more computationally and resource efficient: only 2 variables, peak moment and peak lateral velocity maximum, were used.	-	-
Giraudo et al. (2016)	Measure the fragmentation and insecurity of careers in response to labour market flexibility and its' relationship with injury risk.	56,760 individuals (7% systemic sample) from the Work and Health Histories Italian Panel, a database of individual work histories; workers 30 years or	Demographic and employment information summarized into 6 variables: entry contract; number of contracts; number of jobs;	**HCA** using average linkage-within-groups criterion	Clustering	6 distinct clusters were identified with clear differences between clusters. A positive association between injury risk and level of career fragmentation existed, indicating precarious	-	✔

		younger from 2000-2005.	number of economic sectors; work intensity; duration of the longest period of non-employment.			workers have higher risks of injuries of all types.		
Heiden et al. (2016)	Use cluster analysis to reduce the number of predictors to build an EMG exposure prediction model for job exposure assessments.	120 computer users	EMG exposures, questionnaires, and workstation measurements from computer users during 2hrs of work.	**HCA**	Clustering	Using cluster analyses to select predictors for an EMG exposure prediction model in computer users was not superior to a stepwise hypothesis testing method.	-	✔
Kakhki et al. (2019)	Classify and predict main sources of safety risks of occupational incidents in grain elevator Agro-manufacturing operations in the USA.	5,400 opened and closed workers' compensation claims from 2008-2016.	Claims data: type of injury, nature of injury, class description, age, experience, gender, and cause of injury. 70 and 30% of data were used for training and testing, respectively.	**MLP NN, RBF NN**	Multiclass classification	The causes of occupational incidents amongst agro-manufacturing operations within commercial grain elevators could be predicted with acceptable accuracy using MLP or RBF ANNs (≈ 60%). For carpal tunnel syndrome, the most likely cause was strain.	-	-
Krishna et al. (2015)	Integrating traditional statistical measures with data mining tools to analyze Nordic Musculoskeletal Questionnaire survey data in crane operators in an Indian steel plant.	76 crane operators in an integrated steel plant in eastern India.	Worker, workplace, and task related variables collected using systemic observation with company records, questionnaire survey, and	**CART**	Regression	The CART model could be used to predict the importance of a predictor relative to MSD occurrence and severity. Crane height was the most important factor associated with MSD occurrence and severity in the lower back, neck, and shoulders (except for lower back	-	-

			anthropometric measurements; MSDs and severity captured using the Nordic Musculoskeletal Questionnaire.			injury severity, which was linked to cabin feature).		
Laflamme (1997)	Examine the age-related risk of types of overexertion injuries among female nursing personnel in Sweden.	6088 injury events experienced by nursing auxiliaries and nurses in Sweden.	Injury records regarding part of the body injured, nature of injury, and main external agent.	***k*-means cluster**	Clustering	Using regression models on the two defined clusters, one cluster had back and neck injuries, while the other had shoulder, arm, or other parts of the body. Age and time were also other factors that differentiated between groups - individual and joint effects on overexertion injury rate ratios were observed.	-	-
Leijon et al. (2006)	Identify subgroups of individuals with similar patterns of working and living conditions and study how these patterns associate with neck/shoulder and low-back disorders.	1,341 referents aged 20-59 years old, working ≥ 17hrs/week, no treatment for MSDs in the preceding 6 months.	Questionnaire with 16 items on risk & protective factors, and pain questions for the neck, shoulder, and low back; interviews by ergonomist and behavioral scientist inquiring about job title, gender, employment type, education, socioeconomic position, estimated	**HCA** (Ward's method); *k*-means cluster	Clustering	11 clusters were identified, with 53-194 individuals per cluster, each with different main characteristics. Three clusters showed increased risk for MSDs, characterized by: "onerous human service job"; "free agent"; and "family burden".	-	-

			income, family situation, health behaviours, weight, and physical activity level.					
Leleu et al. (2011)	Explore the use of latent semantic analysis to analyze open-ended responses from military personnel who participated in the 2001-06 Millennium Cohort Study to investigate important health concerns.	108,129 military personnel from the Millennium Cohort Study, with 27,916 participants providing answers to the open-ended questions.	Questionnaire with 67 questions including open-ended questions.	**Latent semantic analysis, *k*-means cluster** (partitioning around medoids)	Clustering	Individuals who provided answers in an open-ended text field had lower self-reported general health. LSA could be used to determine the topics of discussion, and the 20 clusters were able to identify the area of concerns.	-	-
Ling et al. (2020)	Develop and validate a model to predict the work-relatedness of MSDs among semiconductor workers using MSD diagnoses and ergonomic investigations.	307 ergonomic reports for work compensation claims from 2011-19 amongst back-end semiconductor workers with one MSD diagnosis.	277 ergonomic reports for training, 30 for validation. Reports included demographics, ergonomic risk factors, exposures, work activities and exposures, confounding factors contributing to MSDs, employing time before first MSD, and work-relatedness of MSDs. Report were	**Logistic regression**	Multiclass classification	A predictive logistic regression model was constructed to determine predictors of WMSD using data from ergonomic assessments; this predictive model yielded cross-validation accuracy = 86.2%.	-	-

Márquez Gómez (2020)	Develop models to predict work-related musculoskeletal discomfort in the meat processing industry using physical, psychosocial, and individual variables.	174 meat processing workers from Venezuela.	categorized as low, medium or high exposure. 4 physical exposure assessments: postural overload, repetition of movements, lifting of loads, and pulling/pushing of loads; 6 psychosocial variables: psychological requirements, active work and development opportunities, insecurity, social support and leadership quality, double presence, and esteem; 10 individual variables: gender, age, BMI, smoking, MSD history, exercise, household chores, length of employment, overtime, and job rotation; MSK discomfort questionnaire.	Bayesian network, logistic regression, ANN, decision rules, DT, **functional tree**, logistic classifier, logistic model trees	Binary classification	Meat industry workers face elevated risk for experiencing MSK discomfort, and functional tree was the best model to predict MSK discomfort in 2/3 body regions of interest.	-	-

Masci et al. (2020)	Identify variables affecting the biomechanical overload of the distal upper limb among Italian dairy parlour workers, define risk profiles, and propose interventions to reduce physical exposure.	40 male milking parlor workers using parallel, Herringbone, and Rotary milking parlor systems.	Structured questionnaire to collect personal and occupational information; anthropometric measurements; milking parlor characteristics; environmental parameters; work organizational data; biomechanical risk (measured using Strain Index); muscle activation of distal upper limb measured using sEMG.	**HCA**	Clustering	3 clusters were identified, and ANOVAs were used to compare risk assessment results between groups. The main determinants of risk were identified to be workstation characteristics, work organization, and milking routine.	-	-
Persona et al. (2006)	Analyze occupational injury data to identify specific risk groups and factors using data mining approaches.	156 work-related injury cases in Italy from 2000-2002.	Injury reports with length of illness, type of lesion, disabling after-effects, industry, and 20 other potential predictor variables, such as sex, age, task, job title, place injury occurred, worker training, etc.	**CART**, CHAID, exhaustive CHAID, Quick, unbiased, efficient statistical tree	Multiclass classification	The CART technique generated a decision tree to help determine the most critical work conditions related to each outcome variable. For example, the riskiest activities were load/unload, working at machine, and cleaning/maintenance & control, and the most critical job title was factory worker.	-	-
Reme et al. (2012)	Provide early subgroup classification of	359 volunteer patients looking for treatment for work-	Screening questionnaire covering	***k*-means cluster** with	Clustering	4 clusters were identified with different primary characteristics. Pain	-	-

	workers with acute, work-related low-back pain to guide treatment approaches to prevent work disability.	related back pain in the USA.	functional limitation, pain intensity, depressive symptoms, interpretation of pain, activity avoidance, organizational support, life impact of pain, and recovery expectations.	Euclidean distance		outcome and functional limitations differed significantly between groups after 3 months in the following order: emotional distress > activity limitation > workplace concerns > minimal risk.		
Sadeghi et al. (2020)	Develop a model to assess occupational health and safety risks for construction site workers, analyze risk magnitudes, and evaluate them using an Adaptive Neuro-Fuzzy Inference System.	290 datasets from field observations of workers on a Malaysian construction site.	Datasets consisted of safety risks to Malaysian construction workers from papers, reports, articles, structured interviews, and questionnaires with safety officers, processed by 57 experts in construction engineering and management. 203 and 87 datasets used for training and testing, respectively.	**ANN**	Multiclass classification	Fuzzy-based rules were created using the data from the construction sites, a conclusive occupational health & safety assessment was developed with 4 risk parameters, and magnitude of risks are predicted with high accuracy and were used to propose appropriate evaluations strategies.	-	-
Sivak-Callcott et al. (2011)	Determine factors associated with pain/injury in ophthalmic plastic	130 ophthalmic plastic and reconstructive surgeons reached	29-question survey regarding specialty area, professional	**CHAID**	Binary classification	CHAID analyses were used to determine which factors were significantly associated with surgeons	-	-

	and reconstructive surgeons.	through the American Society of Ophthalmic Plastic and Reconstructive Surgery's listserv.	experience, use of loupes and other devices, exercise practices, surgery behaviors, perceptions of discomfort, pain, and injury, demographics, and attentional data regarding surgical loupes and headlamps during surgery.			modifying their practice to accommodate pain or spinal disorder.		
Solovieva et al. (2006)	Investigate whether the pattern of work tasks is associated with finger joint osteoarthritis or the localization of osteoarthritis in the most-used fingers.	291 middle-aged female dentists.	OA evaluations using hand radiographs; self-administered questionnaires regarding anthropometric measures, occupational exposure, daily manual activities, family history of Heberden's nodes, smoking history, and BMI; work history assessment regarding time spent in last 10 years performing	***k*-means cluster**	Clustering	3 clusters were revealed, primarily distinguished by level of work variation (low, moderate, high). Dentists with less task variation had significantly greater prevalence of OA in thumb, index, and middle fingers.	-	-

			specific dental tasks.				
Suárez Sánchez et al. (2014)	Develop a model to identify factors associated with the prevalence of WMSDs among the general working population.	11,054 male and female workers in Spain from December 2006 - April 2007.	Questionnaire with 78 items regarding working conditions, health damage, and occupational health and safety resources of the company; personal interviews regarding work-related diseases and symptoms resulting from work. 80% used for training.	***k*-NN**	Binary classification	The model achieved WMSD prediction accuracy = 86.79% on the test set, and the strongest predictors were determined: poor workplace lighting conditions, exposure to vibration, and uncomfortable seat.	-	-
Tixier et al. (2017)	Identify safety-critical associations of attributes from large data sets using graph mining and HCA.	4,398 injury reports gathered from >470 contractors in industrial, energy, infrastructure, and mining work from around the world.	Injury report data with unstructured text scanned for 80 attributes using an NLP system.	**Agglomerative HCA** on principal components, *k*-means cluster	Clustering	ML techniques were used in their data mining procedures to identify safety clashes, or attributes combinations that contribute to injuries.	-	-
Zurada et al. (1997)	Develop a diagnostic system to classify industrial jobs according to the risk for low-back disorders due to workplace design.	Workers from 403 industrial lifting jobs (48 manufacturing companies).	Trunk motions (collected using the LMM) and quantified workplace factors during repetitive tasks in high and low risk for LBD workplaces.	**ANN**	Binary classification	Using job characteristics data, lifting jobs were classified into two categories of LBD risk using an ANN-based diagnostic system with 74.7% accuracy.	-	-

*not a ML technique

Table C.3. Studies where machine learning (ML) techniques were applied to understand the underlying mechanisms of work-related musculoskeletal disorders (WMSDs; step 3 in the van der Beek et al., 2017 research framework). **Bolded** machine learning (ML) technique(s) indicates selection as best (evaluated by respective authors). ANN = artificial neural network; EMG = electromyography; MMH = manual material handling; NN = neural network; RF = random forest; RPE = rating of perceived exertion; sEMG = surface electromyography; SVM = support vector machine; ✓ = indicates code or data are publicly available (potentially through request).

Citation	Purpose(s) of ML	Sample	Data	ML technique(s) used	Application	Outcome	Code	Data
Aghazadeh et al. (2020)	Estimate whole-body posture, segmental orientations, and lumbosacral moments directly from kinematics during load-handling activities.	15 healthy, young male individuals with no history of back/knee/hip/surgery or recent pain.	Kinematic and kinetic biomechanical variables during various load-handling tasks derived from a link-segment model driven by motion capture.	**ANN** (×3): one each for posture, body segment orientation, and L5-S1 moment	Regression	Using feed-forward NNs, biomechanical kinematic and kinetics variables can be estimated directly from kinematics with sufficient accuracy (RMSE = 6.7cm, 29.8° and 16.2Nm for posture, orientation, and L5-S1 moment, respectively) during various load-handling tasks without a biomechanical model.	-	-
Ali et al. (2018)	Predict health cause-effect of subjective and objective hand arm vibration exposure during grass-cutting.	204 grass-cutting workers in Malaysia who use hand-held devices.	HAVERA questionnaire, weight, height, hand grip strength using a dynamometer, hand vibration during grass-cutting using accelerometers on the handles and a	Linear regression, **ANN**	Regression	An ANN prediction model resulted in strong regression performance indices for grip strength and vibration exposure. This prediction model can be used to predict vibration exposure using HAVERA	-	-

			human vibration meter.			questionnaire data as input.		
Arjmand et al. (2013)	Map the inputs of a trunk finite element model to spinal load and muscle force outputs during various static lifting activities.	A nonlinear, kinematics-driven finite element model of the thoraco-lumbar spine.	Input (mass of object, anterior and right lateral distances to the L5/S1 joint centre, and load handling technique) and output (muscle forces and spinal load) variables of the thoraco-lumbar spine finite element model during static lifting tasks. Considering all combinations resulted in 10,980 and 5,760 datasets for muscle force and spinal load prediction, respectively.	**ANN** (×2): one each for muscle forces and spine loads, quadratic regression	Regression	A five-layer, fully connected feed-forward ANN mapped the inputs of the finite element model to muscle forces and spinal loads with significantly less RMSE compared to the quadratic regression during static lifting. Since it is easier to implement than the finite element model, ANN may be more practical with sufficient accuracy.	-	-
Asadi et al. (2020)	Develop computer vision techniques to detect isometric grip exertions (2 and 3 class predictions) using facial videos and wearable photoplethysmogram.	18 healthy adults (8M/10F)	GoPro camera footage of individuals' faces; hand grip dynamometer; PPG data during hand grip exertions for 0, 15, 33, 50, 75, and 100% MVC.	**DeepFace** algorithm to process faces in images, **Deep NN**	Binary & Multiclass Classification	Computer vision and DNNs could predict hand force exertions using facial features and PPG to high accuracies.	-	✔
Baghdadi et al. (2021)	Apply multivariate time series clustering to investigate fatigue-related changes in walking gait patterns (between MMH tasks) across different	15 student participants (8M/7F) with some experience performing manual tasks.	During walking between MMH tasks, right ankle kinematics were measured using an IMU; RPE was	**HCA** (with dynamic time warping)	Clustering	Four clusters were identified, and no systematic differences in demographics were found between individuals	✔	✔

	individuals, and to determine if such changes vary systematically based on specific demographic characteristics.		collected every 10-minutes.			clustered based on fatigue-related changes in walking gait.		
Fujimaki and Mitsuya (2002)	Propose NN as an evaluation method for body pressure distribution for seated visual display terminal work.	84 visual display terminal workers from Japan in 2001.	Pressure map from seat and back indicating pressure.	**Self-organizing map, *k*-means cluster**	Clustering	A method is proposed to use SOM to identify pressure patterns, then cluster analyses to identify patterns corresponding to discomfort in visual display terminal workers.	-	-
Furuya et al. (2012)	Determine the effect of loudness and tempo on upper extremity kinematics and muscular activity during repetitive piano keystrokes.	18 right-handed pianists with > 15 years of piano training.	Repetitive keystrokes at 4 levels of loudness and tempo; position sensors to measure peak shoulder, wrist, and finger flexion and elbow extension angular velocities; EMG to measure mean activity of 6 upper-limb muscles and co-activation index of the forearm, upper-arm, and shoulder.	***k*-means cluster**	Clustering	3 clusters were identified with different kinematic and muscular responses to increasing loudness and volume, suggesting that the areas with the greatest risk for repetitive strain injuries in piano playing are individualized.	-	-
Gholipour and Arjmand (2016)	Predict 3D spinal posture in reaching and lifting activities and use this posture to estimate	40 healthy young males without a history of back surgery or recent back, hip, or knee	Spinal posture; 3D position of the hand load relative to the body; and body height. 630	**ANN** (×2): one each for reaching and lifting postures	Regression	S1, T12, and T1 vertebrae orientations were estimated using hand load position	✔	-

	compressive and shear spinal loading.	pain. Half performed reaching only; half performed lifting only.	and 270 samples were used for training and testing, respectively.			and body height in reaching and lifting postures with RMSE < 11°. Postures estimated using the ANNs could be used to estimate spinal loads with sufficient accuracy.		
Gohari et al. (2014)	Optimize off-road vehicle seat suspension, predict spine acceleration from excitation signal, human body mass, and height.	5 healthy males.	Acceleration of the pelvis, body mass, and height. Acceleration measured using accelerometers on the seat, spine, and head. 60%, 20%, and 20%, of data used for training, validation, and testing, respectively.	**ANN**	Regression	Vibration from low-speed vehicle drivers cause spine acceleration, a mechanism for the development of low back disorders. Spine acceleration was estimated using an ANN model with accuracy = 96% using pelvis acceleration, height, and mass.	-	-
Hughes (2017)	Predict L5/S1 compression force using posture, hand load, anthropometry, and disc injury status during heavy occupational lifting.	10 aluminum reduction facility workers.	During heavy lifting of chunks of carbon, these variables were extracted from a single video camera: mass in hands, disc injury status, elbow, shoulder, torso, knee, and ankle angles.	**Bayesian network**	Regression	A deterministic static 2D biomechanical model for L5/S1 shear and compression force was implemented as a Bayesian network. The estimates had errors <1%, but benefitted from allowing uncertainty in model inputs, and allowed for the extension to account for disc injury	-	-

						status. Both L5S1 mean compression and shear force estimates increased when disc injury status was set to true.		
Karwowski et al. (2006)	Model the EMG responses of 10 trunk muscles in lifting tasks using trunk angles, moment, pelvic tilt, and rotation angles.	20 male college students.	EMG activity collected from 5 pairs of trunk muscles; L5/S1 forces and moments computed using a force plate and electrogoniometer (lumbar motion monitor).	**Fuzzy relational rule network**	Regression	This model was able to provide a reasonable level of accuracy in estimating time domain EMG responses for some lifting task conditions.	-	-
Kingma et al. (2001)	Compare 3 techniques to measure L5/S1 joint torque during forward bending and lifting tasks.	8 healthy young males.	Erector spinae muscle activation measured using EMG; sagittal plane trunk tilt angle, angular velocity, and linear acceleration measured using accelerometers; L5/S1 extension torque estimated using a cantilever model, EMG signals, and trunk angle and angular velocity.	**ANN**	Regression	Compared to the LSM, the NN approach underestimated peak L5/S1 joint torque but showed lower average error than an EMG-based model.	-	-
Kong et al. (2018)	Propose a framework to investigate the mechanical energy expenditure of workers using a	1 male subject.	2D video recording; 3D pose estimation model trained using Human Pose and	**Hourglass network**	Regression	A computer vision-based 3D biomechanical model to estimate kinematics, joint	-	✔

	computer vision-based 3D biomechanical model.		Human 3.6 M datasets containing 2D and 3D joint centre coordinates.			forces, and mechanical energy expenditure is presented and provides a non-invasive method to monitor worker injury risk in the workplace.		
Lee et al. (2003)	Develop a system to estimate the magnitude of EMG responses of 10 trunk muscles using lifting task variables.	12 subjects.	Trunk moment and velocity; magnitude of EMG responses of 10 trunk muscles during lateral upper trunk moment exertions (at two moments and three velocities).	**Fuzzy NN**	Regression	Using lifting task trunk motion variables, the magnitude of EMG responses was estimated with a mean absolute error of 8.43%.	-	-
Lu and Ferrier (2004)	Use a multidimensional segmentation algorithm to automatically decompose complex, repetitive motions into a sequence of simple linear dynamic models for ergonomic assessments.	1 subject performing karate.	Human motion kinematic data.	**Cluster analysis** (unspecified)	Clustering	A method to automatically segment motion data was presented, and cluster analysis is used to classify the segment using model parameters. This can be used for classifying repetitive motion in in ergonomic assessments.	-	-
Mathiassen and Aminoff (1997)	Evaluate the use of an isoelectric technique involving the upper trapezius muscle and investigate motor control and cardiovascular responses during such	10 female members of the Institute staff without MSDs in the neck, shoulders, or upper extremities.	Trapezius and deltoideus EMG root mean square-amplitude and zero crossing; supportive force measured using a strain gauge	***k*-means cluster**	Clustering	Three clusters were identified differing in motor control strategies of the upper trapezius during the fatiguing task, providing insight into why	-	-

	contractions in a static fatiguing protocol.		transducer; heart rate and mean arterial blood pressure; and perceived fatigue.			some individuals may develop occupational shoulder and neck disorders.		
Matijevich et al. (2021)	Use data from IMUs and pressure insoles to develop ML models to estimate low back loading during MMH tasks.	10 healthy individuals (7M/3F)	During the performance of about 400 MMH tasks in the laboratory, marker-based motion capture and force plates were used to compute lumbar extension moment; IMU sensors and pressure insoles collected at the same time.	Generalized linear models, SVM, ANN, **gradient boosted decision trees**	Regression	Using signals from a trunk IMU and pressure insoles, a gradient boosted decision tree was developed to monitor lumbar extension moment across a wide variety of MMH tasks with high accuracy (r^2 = 0.89).	-	-
Mehrizi et al. (2019)	Propose a method to estimate 3D pose from multiview images and estimate L5/S1 joint loading during (a)symmetrical lifting tasks.	12 healthy males.	Kinematics during various (a)symmetrical lifting tasks measured using marker-based motion capture and 2 camcorders; L5/S1 joint moment computed using top-down inverse dynamics.	**Deep NN** (×2): one each for 2D pose estimation and 3D pose generation	Regression	Deep NNs were used to estimate 2D and 3D pose from camcorder RGB images. Another deep NN to accurately estimated L5/S1 joint moments and forces compared to kinetics computed using an inverse dynamics model (L5/S1 joint moment time series total absolute error = 9.06 ± 7.60Nm; L5/S1 joint moment peak RMSE = 6.14Nm and r = 0.96; L5/S1 joint	-	-

						force time series RMS < 20N and *r* > 0.80; L5/S1 joint force peak RMSE = 4.45N and *r* = 0.99).		
Mehrizi et al. (2018)	Develop and validate a computer vision based markerless motion capture method to assess 3D joint kinematics of symmetrical lifting tasks.	12 healthy males.	Kinematics of symmetrical lifting tasks captured using marker-based motion capture and 2 camcorders (90° and 135° positions).	**Twin Gaussian process**	Regression	Using a computer vision method with a ML-based discriminative approach led to successful 3D pose reconstruction and kinematic analyses of individuals performing lifting tasks.	-	-
Mehrizi et al. (2017)	Use a computer vision-based method to estimate 3D pose to calculate 3D L5/S1 joint moment during symmetrical lifting.	12 healthy males.	Kinematics of symmetrical lifting measured using marker-based motion capture and x2 camcorders (90° and 135° positions).	**Constrained twin Gaussian process**	Regression	A discriminative computer vision approach enabled by a constrained twin Gaussian process allowed for 3D pose reconstruction. These kinematic data were used to drive top-down and bottom-up inverse dynamic models to reliably predict peak and mean L5/S1 moment.	-	-
Mörl et al. (2005)	Measure the parameters of intersegmental lumbar motion during lifting.	11 healthy subjects with no history of low back pain for individual lifting technique group; 23 healthy subjects for the squat lifting group.	Intersegmental spine motion captured using marker-based motion capture during lifting tasks: individual (freestyle) and	***k*-means cluster**	Clustering	3 clusters were identified across groups: one characterized by limited lumbar motion and greater hip and knee mobility; one	-	-

			squat lifting techniques.			characterized by 3.25 times greater spine range of motion with little/no knee and hip flexion; and one group without a primary characterization. Intersegmental lumbar spine motion was different between individuals and was differentially influenced by the selected lifting technique.		
Mousavi et al. (2020)	Predict 3D L5/S1 segment loads and thorax/pelvis orientations in reaching activities.	20 young male subjects without surgery in the back and no recent pain in the knee or hips.	During reaching tasks at 9 positions × 5 heights: thorax and pelvis orientation measured using IMUs; L5/S1 disc compression and shear loading calculated using an MSK model. 70% used as a training set, 15% each for validation and test sets.	**Recurrent NN** (×2): first to predict thorax and pelvis sagittal flexion angles from load positions and body height; second to predict L5/S1 disc compression and shear loading first model outputs	Regression	An approach employing 2 recurrent NN blocks followed by a fuzzy logic algorithm resulted in more accurate L5/S1 disc loading estimations than a feed-forward NN approach. These tools are easier to implement than the use of biomechanical models.	-	-
Muller et al. (2020)	Detect grip and deposit events as part	13 male participants.	During asymmetric handling tasks, the	**ANN** (to detect grip	Binary Classification	An ANN was used to automatically	-	-

	of a method to predict GRF and moments and load contact forces during asymmetric handling tasks using marker-based motion capture.		kinematics of a marker on left and right hands were measured using marker-based motion capture to determine distance between markers, velocity, and acceleration of both markers.	and deposit events)		detect the grip and deposit events of loads during handling tasks. This was part of a larger system to estimate GRF and moments and load contact forces using kinematic data only.		
Parkinson and Callaghan (2009)	Estimate cumulative spine loading during lifting using kinematics only.	20 participants (10M/10F) with no low back pain in past 12 months.	Kinematic and kinetic biomechanical variables during repetitive, (a)symmetrical lifting derived from a rigid link model driven by motion capture, force plate, and EMG data.	**ANN** (×2): one each for joint moments and forces	Regression	Kinematics and an ANN can be used to estimate cumulative loading during repetitive, (a)symmetrical lifting with sufficient accuracy. Compared to the rigid link model, peak moments and forces tended to be under-predicted, whereas cumulative moments and forces were statistically similar.	-	-
Perez and Nussbaum (2008)	Predict initial and final whole-body lifting postures in 2D and 3D.	20 (10M/10F) healthy volunteers.	During 2, 2-3hr sessions of lifting and transferring objects: 41 anthropometric and strength characteristics reduced using PCA; kinematic data from motion capture and electromagnetic	**ANN** (×2): one each for 2D and 3D	Regression	Initial and final position of whole-body posture in 2D and 3D during lifting and transferring could be predicted using anthropometric, strength, and kinematics regarding hand position with errors	-	✔

			tracking devices were used to derive segment orientations and joint locations - hand position was used.			greater in 3D model.		
Sung et al. (2014)	Formulate prediction models for grip and key pinch strength using easy-to-obtain information.	33 right-handed volunteers from Taiwan without upper extremity MSDs in the past 12 months.	MVC for grip strength and key pinch strength, height, weight, and hand dimensions. 74 and 25 samples used as training and test sets, respectively.	**ANN**, stepwise regression	Regression	ANNs and stepwise regression models were developed to predict maximum grip and key pinch strength using minimal anthropometric measurements. ANNs resulted in smaller RMSE compared to stepwise regression, but estimates were statistically similar.	-	-
Vu et al. (2020)	Develop a model to predict torso body shape changes and lumbar motion during movement in a space suit using fabric strain sensors.	12 male volunteers with BMI between 18.5-24.9 from the Johnson Space Center active test subject pool.	3D torso posture during 8 static postures were measured using 10 body-worn fabric strain sensors and 12 photogrammetry camera modules. Radial basis function interpolation was used to fit 3D human geometry templates to the photogrammetry scans to obtain the ground truth. 96	**Linear regression**	Regression	Participant-specific linear regression models were developed to predict 3D body shape using body-worn fabric strain sensor signals. This enables body shape and lumbar motion analyses within space suits for the development of injury countermeasures.	-	-

			and 60 samples used as training and test sets, respectively.					
Wang et al. (2021)	Develop an approach to predict hand posture (pinch vs. grip) and grasp force using forearm sEMG and ANNs during tasks with varying rate and duty cycle; evaluate accuracy of classifying hand posture and applied force.	14 (9M/5F) college-aged participants with no history of hand or arm injuries.	sEMG of 5 forearm muscles collected during manipulation tasks involving different combinations of grip (pinch and power grips), load, and duty cycle (20%, 80%), and repetition rate (12/min, 20/min). Data was split into 80% training, 10% validation, and 10% testing.	**ANN**	Binary Classification	Individualized ANNs were developed to use forearm sEMG to predict hand posture and grip force, binarized as below or at/above the pre-defined thresholds (1kg for pinch, 4.5kg for power grip). Across all participants, accuracy = 79% and 73% for hand posture and force, respectively.	-	-
Zemp et al. (2016)	Develop an instrumented chair with force and acceleration sensors and automatically identify the users' sitting position.	41 healthy subjects (16F/25M).	During sitting in 7 common sitting positions, force distribution patterns were measured using 16 pressure sensors on the chair, the backrest angle and global chair movements were measured using an IMU attached to the backrest.	SVM, multinomial regression, boosting, NN, **RF**	Multiclass classification	Using leave-one-out cross validation, the RF classifier performed best in classifying 7 common sitting positions (accuracy = 90.9%) using pressure sensor and backrest IMU data. This system can be used to study the effects of prolonged static sitting on WMSD development.	-	-

Table C.4. Studies where machine learning (ML) techniques were applied for the development of interventions to prevent of work-related musculoskeletal disorders (WMSDs; step 4 in the van der Beek et al., 2017 research framework). **Bolded** machine learning (ML) technique(s) indicates selection as best (evaluated by respective authors). ANN = artificial neural network; CART = classification and regression tree; CNN = convolutional neural network; COP = centre of pressure; DT = decision tree; EMG = electromyography; HCA = hierarchical cluster analysis; IMU = inertial measurement unit; *k*-NN: *k*-nearest neighbors; LASSO = least absolute shrinkage and selection operator; LDA = linear discriminant analysis; LSTM = long short-term memory; MAE = mean absolute error; MLP = multilayer perceptron; MMH = manual material handling; NB = naïve Bayes classifier; NN = neural network; QDA = quadratic discriminant analysis; RF = random forest; RMS = root mean square; RULA = rapid upper limb assessment; SVM = support vector machine; ✓ = indicates code or data are publicly available (potentially through request).

Citation	Purpose(s) of ML	Sample	Data	ML technique(s) used	Application	Outcome	Code	Data
Abobakr et al. (2019)	Assess body joint angles in depth images for an ergonomic assessment (RULA).	Dataset of RGB-D images and motion capture kinematic data generated synthetically using 6 virtual human models (2M/2F/2 neutral).	RGB-D images, kinematic data of 3650 sparse postures.	**CNN** (×2): first to segment the body from the background; second to predict body joint angles	Regression	RULA can be performed using RGB-D images using a CNNs, increasing the accessibility of ergonomic assessments.	-	-
Adaskevicius (2014)	Develop a method to identify type, duration, and intensity of movements using wearable sensors to encourage active breaks from prolonged computer usage.	5 subjects (3F/3M).	While performing 6 known tasks, 3D acceleration was measured using a wearable accelerometer in the front pocket of their shirt and the tasks were filmed for labelling.	***k*-NN**	Multiclass classification	The classification scheme was able to achieve accuracy = 78.9% for recognizing computer break tasks and was more successful when using orientation rather than acceleration.	-	-
Akkas et al. (2017)	Develop a computer vision method to automatically estimate exertion time, duty cycle,	50 industrial tasks.	Videos of workers performing 50 industrial tasks; exertion time, duty cycle, and hand activity level	**DT** (determine when an exertion was made), ***k*-NN** (state estimation)	Regression	Unobtrusive ergonomic assessments of exertion time, duty cycle, and hand activity level could	-	-

	and hand activity level using videos of workers performing 50 industrial tasks.		determined manually.			be made for industrial tasks using a computer vision method based on two ML algorithms.		
Akkas et al. (2016)	Develop and compare 2 computer vision methods to automatically estimate average duty cycle (DC), hand activity level (HAL), and elemental time using videos of workers performing a repetitive laboratory task.	19 university student volunteers (6M/13F).	Videos of individuals performing a paced, repetitive reach-grasp-move-release task; DC, HAL, and elemental time determined manually. 41 clips used for training and validation; 87 clips used for testing.	**DT, *k*-NN**	Regression	Unobtrusive assessments of average DC, HAL, and elemental time could be made for a repetitive task using a computer vision method based on ML algorithms. DT slightly outperformed the KNN-based approach, but both were similar to the ground truth with < 3% error in average DC.	-	-
Alwasel et al. (2017)	Present a framework to classify work poses in masons during concrete wall building using IMUs.	21 masons with various levels of expertise (novice to experts).	During concrete wall building, kinematics were collected using body-worn IMUs and a video camera. 70% and 30% of the dataset (945 units) were used for training and testing, respectively.	***k*-means cluster** (pose segmentation), **SVM** (experience classification)	Binary classification	Experience as a mason corresponds to increased productivity, safety, and quality. Using IMUs and a method involving *k*-means clustering for pose segmentation, individuals could be classified as experts or non-experts solely using kinematic data and an SVM.	-	-
Antwi-Afari et al. (2020)	Use acceleration and plantar	2 healthy male participants with	During 4 types of overexertion-	ANN, DT, **RF**, *k*-NN, SVM	Multiclass classification	Construction activities that place	✔	✔

	pressure distribution data captured using a wearable insole pressure system to recognize WMSD risk factors and assess ergonomic risk level in construction activities.	basic construction knowledge and experience. They had no history of pain/injury of the back and upper and lower extremities.	related construction activities (grip force, lift/lower/carry, push/pull, and non-risk), foot acceleration and plantar pressure distribution were measured using a wearable insole pressure system. Video recordings were taken for manual annotation.			individuals at a higher risk for developing WMSDs can be automatically identified using insoles with an accelerometer and a pressure system when used with a ML classifier. RF led to the best classification accuracy and sensitivities ≥ 95% (evaluated using 10-fold cross validation).		
Antwi-Afari et al. (2018)	Develop a method to automatically classify awkward construction working postures using plantar pressure distribution measured using a wearable insole pressure system.	10 healthy, novice participants with no history of pain/injury of the back and upper and lower extremities.	During the performance of tasks in 5 awkward postures, plantar pressure distribution was measured using a wearable insole pressure system and videos were captured for manual annotation.	ANN, DT, *k*-NN, **SVM**	Multiclass classification	5 awkward construction working postures could be identified using a wearable insole pressure system and an SVM, achieving accuracy and sensitivities > 99% (evaluated using 10-fold cross validation).	-	-
Armstrong et al. (2019)	Demonstrate the utility of artificial intelligence to detect risk-potential of different movement strategies within physical	22 participants (9 paramedics, 13 novices) without MSDs in the 6-months prior.	During backboard lifting, peak flexion angles and extensor moment about the lumbar spine were calculated. Based on this exposure, lifts were	**Linear discriminate analysis**	Binary classification	Using the motion data as input, this system classified the exposure-level of lifts correctly with > 85% accuracy (evaluated using leave-one-out cross validation). This	-	-

	employment standards.		characterized as low or high exposure.			work demonstrated that movement competency can be assessed in an objective manner, and candidates requiring training can be proactively identified.		
Baghdadi et al. (2018)	Classify non-fatigued vs. fatigued states following MMH using kinematic data from one IMU.	20 subjects (14M/6F), recruited from local work forces and students with manual work experience.	Walking gait kinematics collected before and after 3Hrs of exhaustive manual material handling tasks using an IMU attached to the right ankle.	**SVM**	Binary classification	After a MMH fatiguing protocol, kinematic data from the right ankle may be sufficient to classify fatigue status during walking when trained using pre- and post- MMH walking kinematics.	✔	✔
Bangaru et al. (2021)	Develop an ANN-based automated construction worker activity recognition method to recognize 15 complex scaffold builder activities using forearm EMG and IMU data.	7 male, right-handed participants with no history of MSDs or scaffold building experience.	38 features from right (dominant) arm EMG and IMU data during the performance of 15 scaffold builder activities. Data from 6 participants were used for training; data from 1 participant was used for testing.	**ANN**, RF, DT, gradient boosting, SVM, NB, logistic regression, LDA, QDA, *k*-NN, MLP	Multiclass classification	The developed ANN model recognized 15 scaffold builder activities with an accuracy = 94% and weighted precision, recall, and F1-score = 0.94. This model represents a fundamental requirement for real-time assessment of worker safety and performance.	-	-
Barkallah et al. (2017)	Develop a methodology to classify prolonged and/or inadequate postures of workers at	1 participant.	During various MMH tasks, head kinematics measured using an IMU attached to the worker's	**ANN**	Multiclass classification	During MMH tasks, an ANN could predict the adequacy of the working posture with accuracy = 90%	-	-

	workstations using COP and IMU data.		helmet and COP using pressure sensing insoles in the shoes. MMH postures were labelled as adequate or inadequate.			using helmet kinematics and COP data from pressure sensing insoles. When pressure data were represented as an image before input into the ANN, accuracy was improved to 95%.		
Bastani et al. (2016)	Develop an online, supervised, task classification algorithm for monitoring and evaluating MMH activities using body-worn IMUs.	10 young participants (5M/5F).	Kinematic data during 8 MMH tasks collected using 17 body-worn IMUs. 500 and 250 samples used for training and validation, respectively.	SVM, QDA, *k*-NN, NB, ANN, **greedy Bayesian method**	Multiclass classification	An online sparse estimation-based classification algorithm outperformed other ML algorithms for task classification, was viable for online classification and sensor selection optimization.	-	-
Chen et al. (2018)	Propose a wearable system to automatically recognize, record, and evaluate WMSD risk factors using plantar pressure.	8 healthy young participants.	During carrying, lifting, bending, pushing, and pulling tasks, plantar pressure measured using insole-shaped plantar pressure sensors, Kinect camera (RGB-D) to record skeleton joint information. 80% and 20% of data used for training and testing, respectively.	**SVM** (quasi-static posture recognition), **LSTM & CNN** (sequential motion recognition)	Multiclass classification	Using a wearable insole pressure monitoring system, quasi-static postures were recognized with 100% accuracy and sequential motions with accuracy $\geq 74\%$. OpenSim models were built using joint data collected from a Kinetic sensor to help visualize risk factors. These contributions can be used to monitor risk factors during MMH	-	-

						work and visualizations allow for other visual-based ergonomic assessments to be conducted.		
Chung et al. (2002)	Develop a computer-based postural workload evaluation system based on a macro-postural classification scheme that uses perceived discomfort of various joint motions.	19 male automobile assembly line workers.	Postural stress level at each joint, estimated using their perceived discomfort ratings of different joints and overall positions	**3-layer MLP** (backpropagation used)	Multiclass classification	Overall whole-body discomfort level was predicted using a NN that uses the postural stress level at several joints during manufacturing tasks. Predicted values had a moderate linear relationship to measured values (r^2 = 0.59)	-	-
Conforti et al. (2020)	Estimate biomechanical risk during lifting tasks via postural pattern recognition using ML driven by kinematic data from wearable sensors.	26 healthy subjects.	Kinematic data from squat (i.e., correct) and stoop (i.e., incorrect) lifting of 1, 2, and 5kg loads measured using 8 body-worn IMUs; trained using full-body and upper-body only datasets	**SVM**	Binary classification	Using a kinematic data from upper and lower body during lifting of various loads, SVM can be used to predict lifting posture: across all loads, full-body dataset accuracy ≥ 98.1%, upper-body only dataset accuracy ≥71.2% during lifting phase). Insufficient performance when an upper-body only dataset is used.	-	-
Darbandy et al. (2020)	Detect workers' physical fatigue	8 participants (5M/3F) from manufacturing	Heart rate signal domain features extracted using	***k*-NN**	Binary classification	Performance of fatigue detection was as high as	-	-

	using heart rate signals.	industries and students with different exposures to physical activities	different entropies and statistical tests; fatigue status.			78.18, 60.96, 82.15, 60.18% for accuracy, sensitivity, specificity, and AUC, respectively with *k* = 15.		
Donisi et al. (2021)	Use ML to classify biomechanical risk according to the revised NIOSH lifting equation.	7 healthy volunteers without hypertension, heart disease, or MSD.	Root mean square, SD, min, and max acceleration and angular velocity of an Opal IMU worn on the pelvis during various repeated lifting tasks; evaluated using leave-one-out cross-validation.	RF, DT, **gradient boosting**, adaptive boosting, k-NN, NB, MLP, SVM, logistic regression	Binary Classification	Using raw data from one pelvis Opal IMU during various lifting tasks, gradient boost achieved excellent performance was achieved to estimate risk/no-risk according to the revised NIOSH lifting equation.	-	✔
Eksioglu et al. (1996)	Develop peak pinch strength using anthropometric and physiological variables and elbow and shoulder flexion angles.	20 healthy males.	Pinch strength at 5 elbow and 7 shoulder flexion angles. Weight, age, standard grip strength, age, and anthropometrics (at the arm, hand, and fingers).	Linear regression, **ANN**	Regression	Pinching is a risk factor for hand/wrist WMSDs, and accurately estimating the capacity of peak pinch strength in various positions allow for job design and ergonomic interventions to reduce WMSD risk. In all combinations of variables used, ANNs predicted peak pinch strength with less error compared to linear regression (RMSE ≥ 0.77kgf).	-	-

Greene et al. (2017)	Demonstrate a method for visualizing task factors comprising hand activity level using computer vision.	A worker performing simulated packing tasks with the conveyor belt.	During the performance of packing tasks, a video is taken with clear view of the hands.	**Bayesian estimation principle**	Regression	Spatiotemporal characteristics pertaining to HAL during packing tasks can be measured from RGB video using a semi-automatic approach.	-	-
Greene et al. (2019)	Develop computer vision software for automatically classifying lifting postures using bounding box dimensions without measuring joint angles or fitting data to skeletal models.	Mannequin poses systematically generated using 3DSSPP for various hand locations and lifting postures.	Sagittal and 30° perspectives of mannequin poses for various hand locations and lifting postures, categorized into squat, stoop, and standing using 3DSSPP. Bounding boxes drawn around images using background subtraction. Tested on videos of workers performing lifting tasks in various manufacturing plants and warehouses.	**CART**	Multiclass classification	Using bounding box dimensions obtained from video to classify lifting postures removes the need for precise measurements of joint angles and biomechanical models. A CART model was trained to classify poses from video into standing, squat, and stoop lifting with 96.66% accuracy and ≥ 90% sensitivity and specificity (evaluated using 10-fold cross validation).	-	✔
Hernandez et al. (2018)	Use ANNs to predict the force feasible set of the upper-limb using joint centre Cartesian positions and anthropometric data.	17 right upper-limb MSK models were created based on anthropometrics of 17 male subjects.	Subject mass, shoulder, elbow, wrist, and end-effector location (relative to the trunk). 70%, 15%, and 15% used for training, validation, and	**ANN** (×5): one per force feasible set	Regression	FFS information is useful in digital ergonomic models to specific the limits of workers' strength. This could be predicted using joint centre positions and anthropometric data with $R^2 > 0.89$.	✔	✔

			testing, respectively.					
Jeong and Park (2020)	Develop a sensor-embedded smart chair system to monitor and classify workers' sitting postures in real time.	36 individuals (21M/15F).	During sitting in 11 specific posture categories, the following measures were taken: trunk distance away from back of chair using 6 backrest-mounted infrared distance sensors, pressure patterns using 6 seat-mounted force-sensing resistors.	***k*-NN**	Multiclass classification	When using distance and pressure sensor data together, an KNN model could classify sitting posture as 1 of 11 posture categories with an accuracy of 92% (evaluated using 10-fold cross validation). This approach can be used wit ha real-time feedback system for preventing sitting-related MSDs.	-	-
Karvekar et al. (2021)	Develop a classification model to identify human fatigue level (2, 3, and 4 classes) based on kinematics collected from a smartphone during walking gait.	24 participants (12M/12F) with no recent history of MSDs or lower-body injuries and exercising 2-3 days per week.	42-kinematic features derived from motion data collected from an iPhone 7+ attached to the right shank during repetitive walking between sets of squats; Borg's RPE.	**SVM** (×3 for 2, 3, 4-class prediction)	Binary & Multiclass Classification	SVMs driven by smart phone kinematic data to predict RPE-level achieved classification accuracy = 91, 78, 64% for 2-, 3-, 4-class SVMs, respectively (evaluated using 5-fold cross validation). A smart-phone attached the shank can be used to assess fatigue from gait motion.	-	-
Lee et al. (2020)	Propose an automatic detection technique for	14 male subjects.	During the carrying of loads (3 masses × 2 postures),	**LSTM**	Binary & Multiclass Classification	Using load carrying, one IMU on the left ankle could classify load mass (3	✔	-

	excessive load carrying and classify load masses and postures using a single IMU on the waist or ankle.		kinematic data from both ankles and the waist were collected. Time series IMU data converted to image data. 70% used for training and validation, 30% used for test set.			categories) with 92.46% accuracy, and one IMU on the right ankle could classify posture (2 categories) with accuracy = 96.33%. This system provides an unobtrusive method of monitoring carrying activities and recognizing posture and excessive loading.		
Li et al. (2020)	Use RGB images and deep-learning to assess posture and compute the RULA activation level (4 levels) in real-time.	Human 3.6 (public human pose dataset with full-body kinematics marker data)	RGB images and associated RULA scores of lifting tasks.	**ANN** (×2): RGB image pose detector and RULA score estimator	Multiclass classification	The algorithm achieved accuracy = 93% and 29fps efficiency for detecting the RULA action level from RGB images.	✔	✔
Lim and Souza (2019)	Develop and assess a statistical prediction algorithm using IMU data to classify load carrying mode and level.	9 healthy men from the university community who were right-footed and handed and screened for pre-existing back injury and chronic pain.	Kinematic data captured during carrying performed in 4 modes using 4x Opal IMUs worn on the T6, S1, and lateral shanks.	**RF** (2-stage, hierarchical classification model with Bayesian inference)	Multiclass classification	Wearable IMUs classified 96.9% of carrying modes and 93.1% of load levels. Coronal thoracic-pelvic coordination and pelvis postural sway were the most relevant predictors.	-	-
Lin et al. (2017a)	Use pressure-map manifold learning and insole pressure sensors to recognize patient handling activities.	8 Subjects (7M/1F)	During 7 patient handling activities (PHA) and sitting, a smart insole system captured plantar pressure; leave-one-out cross-validation	***k*-NN**	Multiclass classification	PHA is associated with WMSDs in the back and shoulders; being able to monitor the frequency in performance of such tasks will help in	-	-

			used for evaluation.			prevention. This framework achieved an overall classification accuracy = 86.6%.		
Lin et al. (2017b)	Use spatio-temporal warping pattern recognition and insole pressure sensors to recognize patient handling activities.	8 Subjects (7M/1F)	During 7 patient handling activities and sitting, a smart insole system captured plantar pressure; leave-one-out cross-validation used for evaluation.	***k*-NN**	Multiclass classification	PHA is associated with WMSDs in the back and shoulders; being able to monitor the frequency in performance of such tasks will help in prevention. This framework achieved an overall classification accuracy = 91.7%.	-	-
Liu et al. (2017)	Develop a framework to estimate sitting posture using RGB-D scenes and a 15-joint representation.	3 subjects working in 3 different office environments.	During sitting in two typical, unhealthy sitting postures, RGB-D scenes were captured using a Kinect sensor, labelled, and represented using a 15-joint model.	**NB**	Binary Classification	A NB classifier was used to classify joints as part of a healthy or unhealthy posture during office desk sitting using RGB-D images.	-	-
Maman et al. (2017)	Examine the use of wearable sensors to detect physical fatigue occurrence in simulated manufacturing tasks and estimate the physical fatigue level over time.	8 participants (5M/3F) from the local community; 2 working in manufacturing.	Heart rate, accelerometer data from the ankle, wrist, hip, and torso during assembly, supply pickup and insertion, and MMH tasks; data from 2 participants with held for testing.	**Penalized logistic regression**, multiple linear regression (LASSO for feature selection)	Binary classification; regression	Features derived from a combination of 5 body-worn sensors were able to detect the presence of fatigue (sensitivity = 1.00, specificity = 0.79) and estimate the level over time (MAE = 2.16).	✔	✔

Maman et al. (2020)	Lay the foundation for a data analytics approach to managing fatigue in physically-demanding workplaces; employ ML to detect fatigue during performance of MMH tasks.	24 healthy individuals: manufacturing workers and students with some physical work experience; 15 used for analysis.	Kinematics of MMH tasks performed over 3hr sessions; RPE.	Logistic regression, penalized logistic regression, DT, NB, *k*-NN, **RF**, bagging, boosting	Binary classification	A data analytic framework for physical fatigue management is presented; one IMU is needed only to detect fatigue with average accuracy ≥ 0.850, RF was best, but the features are task-dependent.	✔	✔
Marfia and Roccetti (2017)	Design and implement a computer vision system to detect overexertion movements to avoid common, associated injuries.	3 workers.	While hitting an object with a hammer and lifting a weight from the ground, webcam and depth-sensing cameras recorded the workers.	***k*-means cluster**	Clustering	*k*-means clustering was used to determine how many hands were used to perform a movement by examining the distance between the clusters – values larger than a threshold indicated two different movements were performed.	-	-
Massiris Fernández et al. (2020)	Develop a method to compute RULA scores using snapshots or digital video and computer vision and ML techniques.	Simulated 3D model, one author, and 5 videos of workers performing wall plastering, hammering, tree cutting, drilling, and marshal signs.	Videos recorded during various working scenarios using a smartphone.	**CNN**	Regression	OpenPose was employed using a two-branch CNN to predict confidence maps for body part detection. These are used to calculate joint angles, fed into the RULA equation. Agreement was 1 for upper arm and grand score, but agreement was 0.44-	-	✔

						0.68 for every other body part.		
Morse et al. (2020)	Predict which individuals may be more susceptible to combat basic training injury using 3D body shape images.	12,985 US army basic training recruits (25.1% female); 97 sustained an injury during the 11-week basic training program.	161 anthropometric measurements using a 3D body imaging scanner; injury status over the 11-week basic training program.	Logistic regression, RF, **ANN**	Binary Classification	Using anthropometric data from 3D body scans taking only 20-seconds, an ANN classifier achieved AUC = 0.70, true positive rate = 69% and false positive rate = 35% for predicting injury in the basic training program. Torso length was the highest weighted variable.	-	-
Murillo-Escobar et al. (2020)	Propose an approach to analyze muscle fatigue during dynamic contractions based on time and spectral domain features, Permutation Entropy (PE), and biomechanical features.	9 healthy young adults (5F/4M).	During seated lateral dumbbell lifting until fatigue for both arms, muscle activation measured using sEMG (burst segmented), upper body kinematics measured using marker-based motion capture.	**HCA**	Clustering	HCA was used to identify 3 clusters using the EMG and kinematic data corresponding to non-fatigue, transition-to-fatigue, and fatigue. The best EMG and kinematic features differentiating these states were found: max angle (r = 0.82) and EMG PE (which was better than mean, median frequency, and RMS).	-	-
Nath et al. (2018)	Determine the best location (arm or waist) for MMH task	2 MMH workers.	Arm kinematics during various MMH tasks using an arm-mounted	**SVM**	Multiclass classification	An arm-mounted smartphone can provide acceptable performance in	-	✔

	recognition (lift/lower and carry), and to accurately estimate WMSD risk during the performance of MMH tasks using a body-mounted smartphone.		smartphone (accelerometer, linear accelerometer, and gyroscope).			recognizing MMH task type. Using task type and frequency, can be fed into an Occupational Safety and Health Administration equation to determine MSD risk with 3-task classification accuracy/recall ≥ 89%, precision ≥ 88%, F1 ≥ 89% when calibrated.		
Pistolesi and Lazzerini (2020)	Present an AI-based system that uses wearable sensors to assess the safety level of workers lifting loads.	30 healthy individuals (15M/15F) for development; 9 workers (5M/4F) for implementation.	During proper (squat), unsafe (stoop), and light object lifting, kinematics measured using 2 chest-mounted IMUs (using magnetometer and accelerometer signals only).	**MLP NN**, *k*-NN, SVM	Multiclass classification	During development, the best classifier (MLP NN) had prediction accuracy = 95.6% of the 3 lifting situations (evaluated using 10-fold cross validation). This was successfully implemented in the workplace and can determine how many times workers perform safe and unsafe lifts during the day to help prevent lifting-related injuries.	-	-
Ray and Teizer (2012)	Monitor worker posture and use predefined rules to categorize activities as ergonomic or non-	3 subjects (to assess classifier performance); 5 subjects (to assess classifier performance with	During the performance of standing, squatting/sitting, stooping or bending, and	**LDA**	Multiclass classification	Mitigation of WMSDs can be achieved by training workers to make them aware of unsafe postures for	-	-

	ergonomic using a Kinect range camera.	unfamiliar subjects)	crawling, kinematic data were collected using a depth-sensing camera (Kinect). 33.3% and 66.66% used for training and testing, respectively.			specific tasks and monitor work behavior on a job site. LDA is used in their unobtrusive system for posture classification and achieved error = 5.2% on the test set (one vs. one voting used). Pose estimation is then performed using OpenNI and then ergonomics were assessed using posture-specific rules/criteria.		
Roberts et al. (2020)	Present activity analysis framework to estimate and track 2D worker pose to output activity labels using RGB video footage of construction worker operation.	27 construction workers.	317 annotated (with activity level) videos of bricklaying and plastering operations.	**CNN**	Multiclass classification	This is a vision-based ergonomics assessment method that can assess the ergonomic viability of individual activities. i3D uses a multistage temporal convolutional network, employing a CNN to extract temporally local features and predict activity label from identified pose. This achieved accuracy = 78.5% across all tasks (evaluated leave-one-out cross-validation).	✔	✔

Sasikumar and Binoosh (2020)	Develop a model to predict the risk of WMSDs among computer professionals.	66 computer professionals working in IT companies.	Modified Nordic MSK questionnaire (demographics, physiological and work-related factors, pain, MSDs, etc.); dynamic posture measured using video cameras & Kinovea; assessed using a weighted RULA score; MSD risk index	**RF, NB**, DT, *k*-NN, ANN, SVM	Multiclass classification	MSD risk index (three class prediction) was predicted using postural, physiological, and work-related factors with an accuracy = 81.25%.	-	-
Sharma and Majumder (2017)	Develop model to predict vulnerability of workers to injury based on hand-grip strength using anthropometric measures as a pre-screening assessment of worker capacity and injury risk.	347 young males from India.	Anthropometrics (mass, skin-fold thickness, BMI, body surface area, lean body mass, hand length), hand-grip strength measured using a hand-grip dynamometer.	***k*-means cluster** (identify subgroups); instance-based, NB, **MLP NN**, SVM, RBF NN (classify workers into identified subgroups)	Clustering; Multiclass Classification	5 distinct clusters were identified (named: very poor, poor, fair, good, excellent), primarily differentiated by hand-grip strength. To classify individuals into these groups, MLP NN had the highest performance of accuracy = 95.10%.	-	-
Spector et al. (2014)	Automate the estimation of parameters for the NIOSH lifting equation using depth camera and skeleton algorithm technology.	6 subjects (4M/2F).	During walking, upper arm movements, random motions, and lifting an object from floor to table, kinematics were recorded using marker-based motion capture and an RGB-D	**Gradient boosted regression trees**	Regression	Gradient boosted regression trees were used to generate an error-correction models for NIOSH parameter estimation compared to marker-based motion capture. This improved parameter estimation than used	-	-

			camera (Kinect); leave-one-out cross-validation used for evaluation.			on its own. ML was used to improve the accuracy of automated NIOSH equation assessments using an RGB-D camera.		
Taha and Nazaruddin (2005)	Obtain grip strength data and develop models to predict grip strength in Malaysian industrial workers.	146 workers (104M/42F) from Malaysia in automotive and electronics industries. They had no history of upper limb pain or MSDs in the past 4 months.	Hand dimensions, weight, wrist circumference, age, and grip strength measured using a hand grip dynamometer;	**ANN**	Regression	Grip strength prediction models can be used to estimate worker capacity using easily obtainable metrics. The estimates of grip-strength from their ANN classifier had no statistical difference with measured values. The factors most associated with hand grip strength for each sub-industry (light and heavy workers) were also identified.	-	-
Umer et al. (2020)	Explore the use of combined cardiorespiratory and thermoregulatory measures to model physical exertion using ML algorithms.	10 male participants without a history of cardio or pulmonary diseases, and no MSDs in the past 12 months.	During the performance of MMH tasks, rating of perceived exertion was collected using the Borg-20 scale, cardiorespiratory and thermoregulatory measures using a vest with a respiration	*k*-NN, SVM, LDA, QDA, DT, **bagged trees**	Multiclass classification	Physical exertion could be modelled using non-invasive, body-worn devices to monitor cardiorespiratory and thermoregulatory measures. Bagged trees led the best performance (accuracy = 95.3%) but was improved when using	-	-

			monitor and ECG, and subjective workload assessment using SWAT.			individualized models (average accuracy = 96.7%; evaluated using leave-one-out cross validation).		
Varrecchia et al. (2018)	Test if ML techniques can map time and frequency EMG features on lifting indices (LI) levels to improve biomechanical risk assessment.	10 male participants without MSDs, upper, lower-limb, or trunk surgery, orthopedic or neurological diseases, vestibular system disorders, visual impairments, or back pain.	During the performance of a 3 crate lifting tasks corresponding to LI = 1,2,3, 12 EMG electrodes placed over 12 trunk muscles. The lifted crate was tracked using marker-based motion capture.	**Feed-forward ANN**	Multiclass classification	LI was predicted using EMG features and ANNs. Modelled using time and frequency features led to the best average performance of 83.10% (average of diagonal of confusion matrix). The erector spinae longissimus was the muscle most sensitive to changes in EMG features. EMG features could be used to estimate biomechanical risk during lifting.	-	-
Yang et al. (2020)	Investigate the feasibility of identifying various physical loading conditions by analyzing workers' lower body movements.	12 subjects (10M/2F).	During the performance of masonry and brick-carrying tasks (4 load levels), kinematics collected using an ankle-worn IMU. Video was recorded for annotation purposes.	**Bi-directional LSTM**	Multiclass classification	Using an ankle-worn IMU and bi-directional LSTM, physical load level can be predicted with accuracies of 74.6-98.6%, depending on the load level comparisons (evaluated using 5-fold cross-validation). The more divisions in	-	-

						load levels (maximum 4), the poorer the performance. Reasonable accuracy could be achieved and can help managers continuously and automatically track physical load posed on workers.		
Zhang et al. (2018)	Propose an ergonomic posture recognition method using 3D view-invariant features from a single 2D camera for construction sites.	Human 3.6 M dataset containing 2D and 3D joint centre coordinates.	2D and 3D joint centre coordinates of workers performing various construction tasks with annotation of arm, back and leg postures.	**CNN** (×2): Convolutional pose machine and 3D joint estimation model	Multiclass classification	A classifier employing multi-stage CNNs was trained to estimate 3D posture for the arm, back, and legs with 94.9%, 93.9%, and 94.6% accuracies, respectively when using camera locations and participants not seen in training. This method allows for reliable postural ergonomic assessments using regular RGB cameras that may be already installed on construction sites.	-	✔
Zhao and Obonyo (2020)	Use ML techniques for recognizing construction worker postures from kinematic	4 experienced construction workers from a residential building project.	During routine construction tasks, kinematics were collected using IMUs attached to the	**CNN** (automated feature learning), **LSTM** (learning sequential patterns), SVM,	Multiclass classification	IMUs can be used with a DNN framework to recognize posture in a manner that outperforms various	✔	✔

	data measured using wearable IMUs.		forehead, chest center, right upper arm, right thigh, and right crus. Video was captured to label postures. 72%, 18%, and 10% of data were used for training, validation, and testing, respectively.	NB, *k*-NN, DT, RF		ML classifiers on their own. Best performance was achieved using two baseline models and using personalized modelling. This approach is a crucial part of a data-driven approach to MSD prevention for construction workers.		

Table C.5. Studies where machine learning (ML) techniques were applied for the evaluation of interventions to prevent work-related musculoskeletal disorders (WMSDs; step 5 in the van der Beek et al., 2017 research framework). **Bolded** machine learning (ML) technique(s) indicates selection as best (evaluated by respective authors). BN = Bayesian network; HMM = hidden Markov model; REBA = rapid entire body assessment. ✓ = indicates code or data are publicly available (potentially through request).

Citation	Purpose(s) of ML	Sample	Data	ML technique(s) used	Application	Outcome	Code	Data
Akanmu et al. (2020)	Describe a cyber-physical postural training environment where workers practice performing work with reduced ergonomic risks, and study feedback effectiveness and user interface.	10 students (6M/4F), 7 with construction industry experience.	Body movement data, risk classification, posture practice, posture knowledge dissemination, and posture assessment and feedback modules during training.	**Actor-critic**	Reinforcement learning	An actor-critic reinforcement learning algorithm is implemented in their postural training environment to diagnose the knowledge and skill gaps of individual trainees while learning to achieve safe postures on the virtual construction sites and adapt the contents and delivery methods according to each individuals' needs. This intervention was evaluated to be convenient without interfering with the workspace.	-	-

Arbelaez Garces et al. (2016)	Propose early-stage development acceptability evaluation for Active Pause: software that prevents the risk of WMSDs during extended computer work by providing break and exercise reminders.	28 prospective users (15F/13M) who use the computer more than 2 hours/day.	Questionnaire with 9 questions and plus one Likert scale question regarding their intention to use the software.	**BN**	Multiclass classification	A BN was implemented to understand the contribution of each factor to the final acceptability of the software. The most important factor influencing acceptability was the understanding that this software would reduce WMSD risk. Global precision = 78.57%. By assessing acceptability earlier in the development process, efforts can be aimed to highlight the identified factors.	-	-
Ghasemi and Mahdavi (2020)	Develop a new scoring system, fuzzy Bayesian network REBA (FBnREBA) to assess WMSD risk that increases sensitivity to input variables.	Case study conducted in a desktop gas cooker assembly line.	Risk factors of WMSDs; REBA	**BN**	Multiclass classification	A FBnREBA was compared to a REBA to make it more sensitive to inputs: it can be used to rank interventions based on their effectiveness; and is less likely to provide the same risk score for different postures.	-	-
Samani et al. (2013)	Investigate the effectiveness of a conventional exposure variation analysis (EVA) to discriminate exposure timelines and compare it with a new, cluster-based method (C-EVA).	Simulated repeated cyclic exposure of right upper arm posture varying within each cycle: low and high exposure; near or far range; low or high	Exposure level.	***k*-means cluster, linear classifier**	Clustering; Binary Classification	C-EVA classified the exposures related to WMSDs better than conventional EVA approaches with accuracy = 52%. However, all methods were poor in their ability to discriminate exposure patterns.	-	-

		velocities; and small or large standard deviation of cycle time.						
Thanathornwong and Suebnukarn (2015)	Describe and evaluate a real-time personalized biofeedback system for dental posture training using vibrotactile feedback.	16 dental students (2M/14F) who performed ≥ 6 hours of dental work per day and ≤ 70% in dental ergonomics.	During dental tasks, the tilt angle of the cervical spine is measured using an accelerometer worn on operators' gown.	**HMM**	Binary Classification	Participants who received vibrotactile feedback (in cases where the log-likelihood of the test movements under WMSD HMM was greater than under the non-WMSD HMM) from the system had lower tilt angles at 10th, 50th, and 90th percentiles of Backx, Backy, and muscular load, compared to those who did not get post-test feedback. This personalized system can be an effective intervention to mitigate risk of WMSD development.	-	-
Thanathornwong et al. (2014)	Develop and evaluate a real-time system for predicting WMSDs and providing knowledge feedback to dental students.	16 dental students (2M/14F) who performed ≥ 6 hours of dental work per day and ≤ 70% in dental ergonomics.	During dental tasks, the tilt angle of the upper back and neck are measured using accelerometers worn on operators' gown and face shield, respectively.	**HMM**	Binary Classification	Participants who received knowledge feedback (in cases where the log-likelihood of the test movements under WMSD HMM was greater than under the non-WMSD HMM) from the system had less extension in the neck and upper back compared to those who did not get feedback. This personalized system can be an effective intervention to mitigate risk of WMSD development.	-	-

Appendix D. PAR-Q+

2021 PAR-Q+

The Physical Activity Readiness Questionnaire for Everyone

The health benefits of regular physical activity are clear; more people should engage in physical activity every day of the week. Participating in physical activity is very safe for MOST people. This questionnaire will tell you whether it is necessary for you to seek further advice from your doctor OR a qualified exercise professional before becoming more physically active.

GENERAL HEALTH QUESTIONS

Please read the 7 questions below carefully and answer each one honestly: check YES or NO.	YES	NO
1) Has your doctor ever said that you have a heart condition ☐ **OR** high blood pressure ☐?	☐	☐
2) Do you feel pain in your chest at rest, during your daily activities of living, **OR** when you do physical activity?	☐	☐
3) Do you lose balance because of dizziness **OR** have you lost consciousness in the last 12 months? Please answer **NO** if your dizziness was associated with over-breathing (including during vigorous exercise).	☐	☐
4) Have you ever been diagnosed with another chronic medical condition (other than heart disease or high blood pressure)? **PLEASE LIST CONDITION(S) HERE:** ______	☐	☐
5) Are you currently taking prescribed medications for a chronic medical condition? **PLEASE LIST CONDITION(S) AND MEDICATIONS HERE:** ______	☐	☐
6) Do you currently have (or have had within the past 12 months) a bone, joint, or soft tissue (muscle, ligament, or tendon) problem that could be made worse by becoming more physically active? Please answer **NO** if you had a problem in the past, but it **does not limit your current ability** to be physically active. **PLEASE LIST CONDITION(S) HERE:** ______	☐	☐
7) Has your doctor ever said that you should only do medically supervised physical activity?	☐	☐

☑ **If you answered NO to all of the questions above, you are cleared for physical activity.**
Please sign the PARTICIPANT DECLARATION. You do not need to complete Pages 2 and 3.

- Start becoming much more physically active – start slowly and build up gradually.
- Follow Global Physical Activity Guidelines for your age (https://www.who.int/publications/i/item/9789240015128).
- You may take part in a health and fitness appraisal.
- If you are over the age of 45 yr and NOT accustomed to regular vigorous to maximal effort exercise, consult a qualified exercise professional before engaging in this intensity of exercise.
- If you have any further questions, contact a qualified exercise professional.

PARTICIPANT DECLARATION
If you are less than the legal age required for consent or require the assent of a care provider, your parent, guardian or care provider must also sign this form.

I, the undersigned, have read, understood to my full satisfaction and completed this questionnaire. I acknowledge that this physical activity clearance is valid for a maximum of 12 months from the date it is completed and becomes invalid if my condition changes. I also acknowledge that the community/fitness center may retain a copy of this form for its records. In these instances, it will maintain the confidentiality of the same, complying with applicable law.

NAME ______ DATE ______

SIGNATURE ______ WITNESS ______

SIGNATURE OF PARENT/GUARDIAN/CARE PROVIDER ______

If you answered YES to one or more of the questions above, COMPLETE PAGES 2 AND 3.

Delay becoming more active if:

- ✓ You have a temporary illness such as a cold or fever; it is best to wait until you feel better.
- ✓ You are pregnant - talk to your health care practitioner, your physician, a qualified exercise professional, and/or complete the ePARmed-X+ at www.eparmedx.com before becoming more physically active.
- ✓ Your health changes - answer the questions on Pages 2 and 3 of this document and/or talk to your doctor or a qualified exercise professional before continuing with any physical activity program.

01-11-2020

2021 PAR-Q+

FOLLOW-UP QUESTIONS ABOUT YOUR MEDICAL CONDITION(S)

1. Do you have Arthritis, Osteoporosis, or Back Problems?

If the above condition(s) is/are present, answer questions 1a-1c — If **NO** ☐ go to question 2

1a.	Do you have difficulty controlling your condition with medications or other physician-prescribed therapies? (Answer **NO** if you are not currently taking medications or other treatments)	YES ☐ NO ☐
1b.	Do you have joint problems causing pain, a recent fracture or fracture caused by osteoporosis or cancer, displaced vertebra (e.g., spondylolisthesis), and/or spondylolysis/pars defect (a crack in the bony ring on the back of the spinal column)?	YES ☐ NO ☐
1c.	Have you had steroid injections or taken steroid tablets regularly for more than 3 months?	YES ☐ NO ☐

2. Do you currently have Cancer of any kind?

If the above condition(s) is/are present, answer questions 2a-2b — If **NO** ☐ go to question 3

2a.	Does your cancer diagnosis include any of the following types: lung/bronchogenic, multiple myeloma (cancer of plasma cells), head, and/or neck?	YES ☐ NO ☐
2b.	Are you currently receiving cancer therapy (such as chemotheraphy or radiotherapy)?	YES ☐ NO ☐

3. Do you have a Heart or Cardiovascular Condition? This includes Coronary Artery Disease, Heart Failure, Diagnosed Abnormality of Heart Rhythm

If the above condition(s) is/are present, answer questions 3a-3d — If **NO** ☐ go to question 4

3a.	Do you have difficulty controlling your condition with medications or other physician-prescribed therapies? (Answer **NO** if you are not currently taking medications or other treatments)	YES ☐ NO ☐
3b.	Do you have an irregular heart beat that requires medical management? (e.g., atrial fibrillation, premature ventricular contraction)	YES ☐ NO ☐
3c.	Do you have chronic heart failure?	YES ☐ NO ☐
3d.	Do you have diagnosed coronary artery (cardiovascular) disease and have not participated in regular physical activity in the last 2 months?	YES ☐ NO ☐

4. Do you currently have High Blood Pressure?

If the above condition(s) is/are present, answer questions 4a-4b — If **NO** ☐ go to question 5

4a.	Do you have difficulty controlling your condition with medications or other physician-prescribed therapies? (Answer **NO** if you are not currently taking medications or other treatments)	YES ☐ NO ☐
4b.	Do you have a resting blood pressure equal to or greater than 160/90 mmHg with or without medication? (Answer **YES** if you do not know your resting blood pressure)	YES ☐ NO ☐

5. Do you have any Metabolic Conditions? This includes Type 1 Diabetes, Type 2 Diabetes, Pre-Diabetes

If the above condition(s) is/are present, answer questions 5a-5e — If **NO** ☐ go to question 6

5a.	Do you often have difficulty controlling your blood sugar levels with foods, medications, or other physician-prescribed therapies?	YES ☐ NO ☐
5b.	Do you often suffer from signs and symptoms of low blood sugar (hypoglycemia) following exercise and/or during activities of daily living? Signs of hypoglycemia may include shakiness, nervousness, unusual irritability, abnormal sweating, dizziness or light-headedness, mental confusion, difficulty speaking, weakness, or sleepiness.	YES ☐ NO ☐
5c.	Do you have any signs or symptoms of diabetes complications such as heart or vascular disease and/or complications affecting your eyes, kidneys, **OR** the sensation in your toes and feet?	YES ☐ NO ☐
5d.	Do you have other metabolic conditions (such as current pregnancy-related diabetes, chronic kidney disease, or liver problems)?	YES ☐ NO ☐
5e.	Are you planning to engage in what for you is unusually high (or vigorous) intensity exercise in the near future?	YES ☐ NO ☐

2021 PAR-Q+

6. Do you have any Mental Health Problems or Learning Difficulties? This includes Alzheimer's, Dementia, Depression, Anxiety Disorder, Eating Disorder, Psychotic Disorder, Intellectual Disability, Down Syndrome

If the above condition(s) is/are present, answer questions 6a-6b — If **NO** ☐ go to question 7

6a.	Do you have difficulty controlling your condition with medications or other physician-prescribed therapies? (Answer **NO** if you are not currently taking medications or other treatments)	YES ☐ NO ☐
6b.	Do you have Down Syndrome **AND** back problems affecting nerves or muscles?	YES ☐ NO ☐

7. Do you have a Respiratory Disease? *This includes Chronic Obstructive Pulmonary Disease, Asthma, Pulmonary High Blood Pressure*

If the above condition(s) is/are present, answer questions 7a-7d — If **NO** ☐ go to question 8

7a.	Do you have difficulty controlling your condition with medications or other physician-prescribed therapies? (Answer **NO** if you are not currently taking medications or other treatments)	YES ☐ NO ☐
7b.	Has your doctor ever said your blood oxygen level is low at rest or during exercise and/or that you require supplemental oxygen therapy?	YES ☐ NO ☐
7c.	If asthmatic, do you currently have symptoms of chest tightness, wheezing, laboured breathing, consistent cough (more than 2 days/week), or have you used your rescue medication more than twice in the last week?	YES ☐ NO ☐
7d.	Has your doctor ever said you have high blood pressure in the blood vessels of your lungs?	YES ☐ NO ☐

8. Do you have a Spinal Cord Injury? *This includes Tetraplegia and Paraplegia*

If the above condition(s) is/are present, answer questions 8a-8c — If **NO** ☐ go to question 9

8a.	Do you have difficulty controlling your condition with medications or other physician-prescribed therapies? (Answer **NO** if you are not currently taking medications or other treatments)	YES ☐ NO ☐
8b.	Do you commonly exhibit low resting blood pressure significant enough to cause dizziness, light-headedness, and/or fainting?	YES ☐ NO ☐
8c.	Has your physician indicated that you exhibit sudden bouts of high blood pressure (known as Autonomic Dysreflexia)?	YES ☐ NO ☐

9. Have you had a Stroke? *This includes Transient Ischemic Attack (TIA) or Cerebrovascular Event*

If the above condition(s) is/are present, answer questions 9a-9c — If **NO** ☐ go to question 10

9a.	Do you have difficulty controlling your condition with medications or other physician-prescribed therapies? (Answer **NO** if you are not currently taking medications or other treatments)	YES ☐ NO ☐
9b.	Do you have any impairment in walking or mobility?	YES ☐ NO ☐
9c.	Have you experienced a stroke or impairment in nerves or muscles in the past 6 months?	YES ☐ NO ☐

10. Do you have any other medical condition not listed above or do you have two or more medical conditions?

If you have other medical conditions, answer questions 10a-10c — If **NO** ☐ read the Page 4 recommendations

10a.	Have you experienced a blackout, fainted, or lost consciousness as a result of a head injury within the last 12 months **OR** have you had a diagnosed concussion within the last 12 months?	YES ☐ NO ☐
10b.	Do you have a medical condition that is not listed (such as epilepsy, neurological conditions, kidney problems)?	YES ☐ NO ☐
10c.	Do you currently live with two or more medical conditions?	YES ☐ NO ☐

PLEASE LIST YOUR MEDICAL CONDITION(S) AND ANY RELATED MEDICATIONS HERE: ______________________

GO to Page 4 for recommendations about your current medical condition(s) and sign the PARTICIPANT DECLARATION.

2021 PAR-Q+

If you answered NO to all of the FOLLOW-UP questions (pgs. 2-3) about your medical condition, you are ready to become more physically active - sign the PARTICIPANT DECLARATION below:

- It is advised that you consult a qualified exercise professional to help you develop a safe and effective physical activity plan to meet your health needs.
- You are encouraged to start slowly and build up gradually - 20 to 60 minutes of low to moderate intensity exercise, 3-5 days per week including aerobic and muscle strengthening exercises.
- As you progress, you should aim to accumulate 150 minutes or more of moderate intensity physical activity per week.
- If you are over the age of 45 yr and **NOT** accustomed to regular vigorous to maximal effort exercise, consult a qualified exercise professional before engaging in this intensity of exercise.

If you answered **YES** to **one or more of the follow-up questions** about your medical condition:
You should seek further information before becoming more physically active or engaging in a fitness appraisal. You should complete the specially designed online screening and exercise recommendations program - the **ePARmed-X+ at www.eparmedx.com** and/or visit a qualified exercise professional to work through the ePARmed-X+ and for further information.

Delay becoming more active if:

- You have a temporary illness such as a cold or fever; it is best to wait until you feel better.
- You are pregnant - talk to your health care practitioner, your physician, a qualified exercise professional, and/or complete the ePARmed-X+ **at www.eparmedx.com** before becoming more physically active.
- Your health changes - talk to your doctor or qualified exercise professional before continuing with any physical activity program.

- You are encouraged to photocopy the PAR-Q+. You must use the entire questionnaire and NO changes are permitted.
- The authors, the PAR-Q+ Collaboration, partner organizations, and their agents assume no liability for persons who undertake physical activity and/or make use of the PAR-Q+ or ePARmed-X+. If in doubt after completing the questionnaire, consult your doctor prior to physical activity.

PARTICIPANT DECLARATION

- All persons who have completed the PAR-Q+ please read and sign the declaration below.
- If you are less than the legal age required for consent or require the assent of a care provider, your parent, guardian or care provider must also sign this form.

I, the undersigned, have read, understood to my full satisfaction and completed this questionnaire. I acknowledge that this physical activity clearance is valid for a maximum of 12 months from the date it is completed and becomes invalid if my condition changes. I also acknowledge that the community/fitness center may retain a copy of this form for records. In these instances, it will maintain the confidentiality of the same, complying with applicable law.

NAME ______________________________ DATE ______________________________

SIGNATURE ______________________________ WITNESS ______________________________

SIGNATURE OF PARENT/GUARDIAN/CARE PROVIDER ______________________________

For more information, please contact
www.eparmedx.com
Email: eparmedx@gmail.com

Citation for PAR-Q+
Warburton DER, Jamnik VK, Bredin SSD, and Gledhill N on behalf of the PAR-Q+ Collaboration. The Physical Activity Readiness Questionnaire for Everyone (PAR-Q+) and Electronic Physical Activity Readiness Medical Examination (ePARmed-X+). Health & Fitness Journal of Canada 4(2):3-23, 2011.

The PAR-Q+ was created using the evidence-based AGREE process (1) by the PAR-Q+ Collaboration chaired by Dr. Darren E. R. Warburton with Dr. Norman Gledhill, Dr. Veronica Jamnik, and Dr. Donald C. McKenzie (2). Production of this document has been made possible through financial contributions from the Public Health Agency of Canada and the BC Ministry of Health Services. The views expressed herein do not necessarily represent the views of the Public Health Agency of Canada or the BC Ministry of Health Services.

Key References
1. Jamnik VK, Warburton DER, Makarski J, McKenzie DC, Shephard RJ, Stone J, and Gledhill N. Enhancing the effectiveness of clearance for physical activity participation; background and overall process. APNM 36(S1):S3-S13, 2011.
2. Warburton DER, Gledhill N, Jamnik VK, Bredin SSD, McKenzie DC, Stone J, Charlesworth S, and Shephard RJ. Evidence-based risk assessment and recommendations for physical activity clearance; Consensus Document. APNM 36(S1):S266-s298, 2011.
3. Chisholm DM, Collis ML, Kulak LL, Davenport W, and Gruber N. Physical activity readiness. British Columbia Medical Journal. 1975;17:375-378.
4. Thomas S, Reading J, and Shephard RJ. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). Canadian Journal of Sport Science 1992;17:4 338-345.

PRINT FORM RESET FORM

Appendix E. Research consent form

uOttawa

Université d'Ottawa
Faculté des sciences de la santé
École des sciences de l'activité physique

University of Ottawa
Faculty of Health Sciences
School of Human Kinetics

Research Consent Form

Research Project Title: **Development of an IMU-driven task identification and fatigue prediction framework**

Principal Investigator:
[redacted]
University of Ottawa
Faculty of Health Sciences
School of Human Kinetics
[redacted]
Ottawa, ON [redacted]

Student Researchers:
[redacted]
University of Ottawa
Faculty of Engineering
Department of Mechanical Engineering

[redacted]
University of Ottawa
Faculty of Health Sciences
School of Human Kinetics

COVID-19 Precautions

For the safety of both yourself and the researchers, precautions have been put in place to try and prevent the spread of COVID-19. Before arriving at the lab, you and the researchers will be screened for COVID-19 symptoms and will be asked to use hand sanitizer upon entering and exiting the lab. Both you and the researchers are required to wear a mask at all times during the data collection. All equipment that is touched by either a participant or researcher during the collection will be sanitized using disinfectant before and after the collection and all washable materials will be washed prior to each participants' session.

Background and Purpose of the Study

Fatigue in the workplace is associated with an increased risk for the occurrence of occupational accidents and the development of musculoskeletal disorders (MSDs). Typical fatigue assessment methods can detract individuals from the performance of their task (e.g., subjective perception scales, strength tests, etc.). Unobtrusively recognizing when workers may be excessively fatigued can provide data to encourage timely breaks, optimize work-rest ratios, and inform job rotation schedules, all of which can mitigate the risk for accidents and MSDs. If the most

125 University Private
Ottawa, ON
K1N 6N5
Canada

www.uOttawa.ca

relevant movement and muscle activation variables can be identified, then machine learning-based models can be developed to automatically estimate fatigue levels.

Thus, the purpose of this study is to use movement data and fatigue measurements to develop machine learning-based models capable of estimating fatigue levels. To accomplish this, movement and muscle activation variables associated with fatigue status during dynamic, simulated manual handling tasks will be identified. During the simulated manual handling tasks, your trunk and shoulder muscle activation, heart rate, and whole-body movement will be measured using surface electromyography (EMG), a heart rate monitor, and wearable inertial measurement units (IMUs), respectively.

Description of Study Procedures

You are invited to participate in a one-day motion analysis procedure for approximately 3 hours at the University of Ottawa Human Movement Biomechanics Laboratory [redacted]. The protocol consists of a fatiguing circuit where you are required to pull, lift, carry, and push a crate (8.1kg). One round requires you to pull the crate towards yourself, lift the crate from the floor, carry the crate (approximately 3 metres), release the crate on a table and push it forward, and walk back to the start. You are asked to follow a pace of 1 round per 15 seconds. One set is comprised of 20 rounds (5 minutes); between sets you will be asked for your subjective rating of fatigue (RPE) and to perform a maximal lift strength assessment, comprised of 3 maximal effort static deadlifts attempts. You are asked to perform as many sets as possible. The study ends when any one of these conditions are met: your RPE becomes ≥ 9 out of 10; your maximal lift strength becomes < 70% of your baseline value; you cannot maintain pace for 3 rounds in a row; or you wish to stop.

You will be asked to bring shoes and shorts that you are comfortable performing repetitive manual tasks in. Upon arrival, your height, mass, and various body measurements (e.g., shoulder height, arm length) will be measured. Then, you will be prepared to wear 7 EMG electrodes on the right side of your body, which includes sanitizing the attachment sites with alcohol wipes, and shaving to improve the electrode-skin interface (if necessary). The EMG electrodes will then be attached using medical-grade adhesive to your abdomen (external oblique), back (thoracic erector spinae, lumbar erector spinae, and latissimus dorsi on both sides), and shoulder (anterior and posterior deltoid). To normalize the readings from the EMG sensors, you will perform three trials of maximal voluntary contractions, comprised of a series of static pulling, sit-ups, and back extensions with your maximum strength. Once completed, a Samsung Galaxy Watch 4 will be worn on your left wrist to monitor heart rate. Then, smart pressure insoles will be worn inside of your shoes. Lastly, you will be outfitted with the inertial motion capture suit. Using provided sanitized neoprene bands and a custom shirt, IMUs will be placed on your lower limbs, upper limbs, trunk, pelvis, and head. To calibrate the IMU system, you will be asked to stand in a neutral posture (N-pose), walk for 5 seconds forward, turn around, and return to an N-pose at the starting location. After these calibration trials are completed, the experimental trials will begin.

Data Analysis

The data collected will be analyzed in two graduate students' projects: a master's engineering student (Di Kang) and a doctoral human kinetics candidate (Victor Chan). Complexity analysis of the EMG signals (e.g., sample entropy, a commonly-used measure of the irregularity of time-series signals) will be performed to determine their association with fatigue status during dynamic and repetitive manual handling tasks. Movement, heart rate, and EMG data will be used to develop machine learning-based models that use movement variables to directly estimate fatigue status. Data may also be used for future analyses.

Possible Risks and Discomforts

Fatigue will occur through this protocol due to the performance of repetitive manual handling tasks. However, any risk to you will be minimized by allowing you to stop at any point during the data collection. To do so, you can simply say "I wish to stop" or something alike. Furthermore, you must indicate that you are willing and capable of experiencing physical fatigue prior to participating. Shaving may be necessary to improve the contact between the EMG electrodes and skin; this may cause minor skin irritation. The tape used to attach the EMG electrodes may also cause minor skin irritation; similar to what is experienced with a bandage and typically fades within 2 to 3 days. Should you experience any major discomfort, please tell us immediately and seek primary care from a medical professional on campus [redacted] or a medical professional of your choosing.

Possible Benefits

You will not obtain any direct benefits from participating in this study, aside from a) performing fatiguing physical activity, and b) obtaining an educational opportunity to observe your own movement and muscle activation data on Xsens MVN Link software. However, the data collected will contribute to the development of machine learning-based models to estimate fatigue levels, which can help reduce MSD risk. Futhermore, the results of this study will contribute to the body of knowledge surrounding EMG analysis of fatigue during dynamic contractions.

Voluntary Participation

Sixteen male or female participants aged 16 years of age or older will be recruited for this study on a first come, first served basis. Participants must be pain- and injury-free in the past 12 months and be cleared to participate in fatiguing, physical activity. If the necessary number of participants has been met, you will be informed by a researcher via email that the study has completed. Your participation is voluntary and thus, you have no obligation to participate in this study. You may withdraw from the study at any time before or during the data collection session with no penalty or coercion. If chose to withdaw, all your data will be destroyed.

Confidentiality

All personal information is kept confidential. Information gained from this study will be stored electronically and will need a password to access, which will only be known to Dr. Ryan Graham and the research team. Paper study records are stored in a locked cabinet in a locked office and will be destroyed after 5 years post-publication; electronic records without personally identifying information will be stored indefinitely. You will not be identified by name in any reports/presentations of the completed study, and your face will never be presented in any reports/presentations. Your anonymity will be strictly maintained – you will never be identified by name; only by an independent study number (e.g., S01).

Compensation

Compensation is not provided for your participation in this study.

Funding Source

This study is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC; 153648)

Questions about the Study

You are free to ask questions at any time during the protocol and by contacting the investigators (on page 1). The ethical components of this research project have been approved by the University of Ottawa research ethics board. If you have any questions regarding the ethical conduct of this study, please contact:

> Protocol Officer for Ethics in Research at the University of Ottawa
> Tabaret Hall
> [redacted]

There are two copies of the consent form, one of which is yours to keep.

Research Project Title: **Development of an IMU-driven task identification and fatigue prediction framework**

Consent
I have read this consent form, and I agree to participate in the procedures of this study.

____
Printed Name of Participant

______________________________________ ______________________________
Signature of Participant Date

Investigator Statement (or Person Explaining the Consent)
I have carefully explained to the research participant the nature of the above research study. To the best of my knowledge, the research participant signing this consent form understands the nature, demands, risks and benefits involved in participating in this study. I acknowledge my responsibility for the care and well-being of the above research participant, to respect the rights and wishes of the research participant, and to conduct the study according to applicable Good Clinical Practice guidelines and regulations.

____
Name of Investigator/Delegate (printed)

______________________________________ ______________________________
Signature of Investigator/Delegate Date

Informed Consent to have Videos Taken
I consent to have video footage taken during the protocol. I understand that these videos are being used as part of the analysis and if any of these videos are used in a subsequent presentation or publication, that my face and any other identifiers will be blurred. You cannot participate in the research study without consenting to having video footage taken.

______________________________ ______________________________
Name Date

____
Signature

______________________________ ______________________________
Witness Name Witness Signature

Future Participation
- ☐ I am interested in being contacted to participate in future research performed by this laboratory (your email information will be saved in a password protected file).

Appendix F. Related projects completed in PhD

Chan, V. C., Beaudette, S. M., Smale, K. B., Beange, K. H., & Graham, R. B. (2020). A subject-specific approach to detect fatigue-related changes in spine motion using wearable sensors. *Sensors*, *20*(9), 2646.

Mavor, M. P., **Chan, V. C.**, Gruevski, K. M., Bossi, L. L., Karakolis, T., & Graham, R. B. (2023). Assessing the Soldier Survivability Tradespace Using a Single IMU. *IEEE Access*.

Kang, D., **Chan, V. C.**, & Graham, R. B. (unpublished). The relationship between trunk muscle EMG Sample Entropy and fatigue. Submitted to Journal of Electromyography and Kinesiology.

Appendix G. Methods to analyze EMG signals for studying fatigue

Table G.1. Examples of commonly used methods to analyze EMG signals for studying effects of fatigue.

Category	Method	Description
Amplitude-based parameters	Peak-to-peak amplitude	• Calculated as maximum minus minimum amplitude. (Kamen, 2014)
	Average rectified amplitude	• Average value of the rectified EMG signal. (Kamen, 2014) • Rectification: converting all voltages to positive values.
	Root mean square (RMS) amplitude	• Alternative to average rectified amplitude using square root of signal. (Kamen, 2014) • Rectification not required.
	Linear envelope	• Signal is rectified then low-pass filtered; computed as moving average of the resulting signal. (Kamen, 2014)
	Integrated EMG (iEMG)	• Signal is rectified then integrated over a selected time period. (Kamen, 2014)
Spectral features	Turning points	• Number of peaks and troughs in a selected time period. (Kamen, 2014)
	Zero crossings	• Number of times the signal crosses zero. (Kamen, 2014)
	Mean frequency (MNF)	• The sum of product of the EMG power spectrum and the frequency divided by the total sum of the power spectrum (Phinyomark et al., 2012).
	Median frequency (MDF)	• A frequency at which the total power is divided into two regions with equal amplitude (Karthick et al., 2018).
	Dimitrov's spectral fatigue index (FInsm5)	• A set of parameters based on the use of different special moments that are quantitative measurements of the shape of a signal (e.g., mean, skewness) (González-Izal et al., 2012). • Higher sensitivity to map fatigue compared to MDF (Dimitrov et al., 2006).
Time-frequency distributions	Instantaneous MNF (IMNF)	• A group of values obtained using an equation for the mean frequency at each instant of time (González-Izal et al., 2012). • An alternative frequency analysis method that is more suitable for non-stationary EMG signals (González-Izal et al., 2012).
	Wavelet spectral parameters	• Wavelet parameters obtained using stationary wavelet transform (SWT). • SWT is similar to discrete wavelet transform (DWT) without down-sampling. DWT acts like a bank of low- and high-pass filters that decompose a signal into numerous signal bands (Kilby & Hosseini, 2004). • Wavelet parameters are calculated as ratios of moments of different orders that are obtained for different wavelet bands, ratios between energies at different scales, or ratios between waveform lengths at different scales (González-Izal et al., 2012).
Non-linear parameters	Entropy parameters	• A non-linear measurement of the complexity of a signal used to estimate the randomness of the signal without any previous knowledge about the source generating the dataset (Delgado-Bonal & Marshak, 2019; González-Izal et al., 2012). • Approximate entropy (ApEn): measures the logarithmic probability that nearby pattern runs remain close in the next incremental comparison (Delgado-Bonal & Marshak, 2019). • Sample entropy (SampEn): The negative value of the logarithm of the conditional probability that two similar sequences of m points remain similar at the next point m+1, counting each vector over all the other vectors except on itself (Delgado-Bonal & Marshak, 2019).

	Fractal analysis	• Measures complexity using the fractal dimension (González-Izal et al., 2012).
	Recurrence quantification analysis (RQA)	• A method to transform single-channel EMG signal onto a bi-dimensional space. Used to detect deterministic structures in signals that repeat throughout the contraction (González-Izal et al., 2012).