Machine learning assisted human fatigue detection, monitoring, and recovery: A Review

Arsalan Lambay, Ying Liu, Phillip L. Morgan, Ze Ji

A R T I C L E  I N F O

Keywords:
Human physical fatigue
Occupational fatigue
Muscle fatigue index
Human–robot collaboration
Muscle fatigue model
Functional data analysis
Human performance modeling

A B S T R A C T

The use of knowledge-based information systems to improve human performance has been limited by a lack of comprehension of how an individual’s performance diminishes when fatigue accumulates, which might vary between individuals depending on their working environment. Although the rise in automation has been witnessed, there are still some physically demanding and exhausting jobs in the manufacturing environment that, if not appropriately managed, can result in long-term issues including musculoskeletal disorders and impairments to psychological well-being. To detect, comprehend and manage the development of solutions for fatigue detection, Machine Learning (ML) has been a useful tool. This paper presents a review of the use of ML techniques for the detection and monitoring of an operator’s work-related physical fatigue in repetitive work and Human–Robot Collaboration (HRC) settings. The novel review offers an overview of the detection complexity of human fatigue in manufacturing-related contexts. The review has three major components: First, the level of fatigue detection complexity with the help of ML, which presents only specific influencing factors generated in relation to human performance while operating under fatigue conditions are included — in the human worker and the detecting technology. Finally, the challenges and limitations of the complexity of holistic approaches in the monitoring/recovery of human fatigue in essence to the physical exertion of an individual are critically discussed.

1. Introduction

The paradigm of the industrial revolution has grown at an unprecedented speed. Industrial (I5.0) has brought advancements in intelligent agent systems, sensing devices and automation. Greater automation has increased the usage of robotic systems in manufacturing and warehousing activities and virtual assistance systems for job optimization. Despite the undeniable fact that automation has caused some job losses, this new era of I5.0 is distinguished by its reliance on highly trained workers that can benefit from technological improvements. The future of the industry will be judged on how successfully it manages its three main resources; By utilizing its three main resources to the fullest, I5.0 will depict the industry’s future, they are: a labor force, resources, and enabling technologies. The absence of trustworthy and tailored models that can measure the impact of work tasks on a worker’s performance is causing the inability to optimize the three resources mentioned above jointly. Expert systems are now utilized in some situations to automate or optimize decision-making processes by learning from human behavior; nevertheless, they must consider the impact of automation and supporting technology on human performance. Many reasons have been outlined by, Zahra et al. [6]; (1) human performance varies depending on an individual’s demographic characteristics, time of the day, and task complexity; (2) many studies on human performance in occupational settings have gathered data through surveys, with gaps in quantitative data on and understanding of how an individual’s performance changes over a day/work shift; and (3) there is a disconnect between predictive and prescriptive models (prescriptive analytics assists you in formulating precise suggestions for improvement, predictive analytics projects likely future results) that attempt to model workplace fatigue.

Inspired by human-in-loop advancement, automation has led to the development of collaborative robots (Collaborative Robots). Although many collaborative robots are highly automated, these advanced manufacturing jobs (jobs such as aerospace/medical and pharmaceutical equipment manufacturing) are still highly fatiguing. Fatigue has been identified as one of the leading causes of quality inefficiencies and accidents. Workplace fatigue is a multidimensional term that affects a worker’s productivity. Despite the fact that cobots are implemented to reduce human workload, repetitive and daily physical activities develop fatigue.
Furthermore, it is connected to impacting psychological, economic, and environmental factors [9]. Fatigue is recommended to be dealt with in consideration with occupational health and safety since it has serious short- and long-term repercussions. Occupational fatigue includes a variety of characteristics, according to the many works of literature, including mental and physical fatigue.

This paper focuses on physical fatigue as it has many detrimental effects on individuals, socially and economically. Physical fatigue reduces one’s ability to execute a physical task due to past physical exertion [10]. Physical fatigue can cause discomfort, reduced motor control, and a loss of physical capacity in the short term, making it especially harmful in production scenarios. The consequences could include declining performance and productivity, rising job quality concerns, an increase in accidents and human error frequency. Physical fatigue can lead to long-term health problems such as chronic fatigue syndrome and lowered immune function [11]. Several studies have suggested that fatigue is caused by numerous factors, including lack of sleep, which contributes to tiredness [12]; disorders such as depression, anxiety, and diabetes [13]; heavy physical activities – linked to whole-body fatigue [14]; and repetitive tasks – linked to localized muscle fatigue [15]. Fatigue impacts an individual’s social life and economic area. Employee fatigue, for example, in the US costs businesses $136.4 billion (about $420 per person in the U.S.) a year in lost productivity and other health-related expenses [16].

Despite plenty of research, there is yet to be a single unified definition and measurement of fatigue. This is largely due to disagreements over its nature and quantifying its dimensionality [17,18]. Our proposed review allows us to comprehend the study of the onset of fatigue modes. As mentioned above, fatigue is multidimensional and varies from person to person. Every occupation will have different causes for the onset of fatigue for different demographics. The current review — focused on workplace fatigue, is divided into two stages. The first stage aims to try and better understand the critical features used in ML for predicting fatigue. The second stage identifies the fundamental challenges in choosing the correct characteristics to predict fatigue.

The paper is organized in the following manner: Section 2; includes background and understanding of the concept of fatigue concerning its prediction of fatigue. Additionally, it includes a background understanding of demographic variables and human fatigue during HRC settings. Section 3 consists of the search framework employed to review the different fatigue papers. Section 4 involves a critical review of the literature addressing issues concerning the detection, monitoring and classification of fatigue and additionally identifying it in an HRC setting. Section 5 includes a discussion of the key outcomes of the critical review, followed by concluding remarks and future directions.

2. Studies

2.1. Fatigue

According to several studies that have been published, 20% of all working people report feeling fatigued [19–21]. Fatigue is comprehended as the sense of being exhausted or sleepy. However, the situation is different among professionals; there is an extensive discussion within and between the numerous related disciplines, but there is not yet a universal agreement on the definition [22]. Researchers use a wide range of supplemental measures and experimental techniques to evaluate and study fatigue because there is no universally accepted definition. We note these points because they are presented throughout the entire review of research on fatigue, making it challenging to address it in total capacity adequately. Although the advancement of technology, both in hardware and software, has opened the doors to new possibilities, the subjective nature of (often measuring as well as experiencing) fatigue makes it difficult to classify the root cause of it fully. However, and crucially, our ability to fully appreciate the concept of fatigue can aid in creating technology and intervention processes that will help employees manage 24/7 operations and lower hazards associated with fatigue [23].

The dichotomy between fatigue being a physical, mental, or both states of being is considered one of the most basic. However, the concept of fatigue as a physical condition phenomenon is the most researched in the field of occupational health. As a result, it is possible to precisely analyze the conditions of fatigue because of physical exhaustion. In essence, fatigue research has been concentrated towards physiological and visual data analysis such as blood pressure [24], heart rate [25], actigraphy (Actigraphy is a recognized technique for employing a noninvasive accelerometer to measure sleep characteristics over a span of days to weeks.) [26], and through using tests including e.g. the logical reasoning and numerical amplitude test [27] and the Psychometric Vigilance Test (PVT) [28]. According to early studies, sleep disorders are one of the main causes of sleeplessness; for instance, failing the PVT test is typically associated with an insufficient sleep cycle [29]. It is adverbious in benchmarking a standard of identifying and measuring fatigue, but it has a significant drawback. In addition to the assumption that other variables do not cause changes in the indicator, it also overlooks individual differences in how people respond to fatigue. For instance, research found that even during long work shifts, heart rate did not rise. Workers arguably reduced their production to meet the needs of the long shifts, which was the explanation for this, nonetheless. Moreover, performance changes might result from illness or adverse pharmaceutical side effects [19].

However, among the various definitions provided by researchers, according to Brown 1993 (p. 240), the definition of fatigue is subjectively experienced disinclination to continue executing the activity at hand because of simply reflecting on individual efficiency [30]. This approach has advantages in that it acknowledges the individual differences in how people experience fatigue, but it also has drawbacks. It raises the issue that fatigue is a psychological condition and offers no explanation for what would explain or induce disinclination. In this literature, the researchers view fatigue as a subjective phenomenon. The issue with this conclusion is that it needs to be clarified if fatigue refers to a mental or physical condition. Thus, making the distinction between the two is crucial. These variations are significant from a practical standpoint because they affect how quickly people recover and, consequently, how well they manage their fatigue [31].

Furthermore, the literature does not distinguish between acute and chronic forms of fatigue [19]. Various literature with varying degrees of success has combined subjective and objective metrics. Rosa (1991) revealed that 12-hour day shifts were associated with more significant fatigue, increased drowsiness, and lower sleep quality [27] and Baulk et al. observed that long work shifts decreased subjective and objective performance tasks evaluated through actigraphy [20]. The extended work week had no adverse effects on blood pressure, heart rate, or salivary cortisol levels. Van der Hulst (2003) concluded that subjective (as opposed to objective) measurements are more frequently associated with fatigue following an examination of extended working hours [32].

In the end, however, we want to emphasize how the study methodology affects the validity of published research. Most research examines correlations between specific variables using cross-sectional approaches [19]. It is crucial to establish a link between variables and features. Since longitudinal designs can rarely randomly assign participants to groups and have a high dropout rate over time, often resulting in small samples and underpowered studies [27]. The complex nature of the relationship between the independent and dependent variables is better highlighted by observational studies, but they also do not enable the determination of causal factors; the odds ratio reflects the factor most strongly correlated with the dependent variable but does not allow for theory testing [33]. We also want to stress that most research generally reflects the fundamental demographic factors of age and sex. Other demographic characteristics are not always investigated or, if they are, may not be considered as possible predictors of outcomes due to space restrictions in the investigators’ intrinsic emphasis. In
the other sections, we talk about a few of these factors. As they are not utilized directly to measure the dependent variable, demographic characteristics cannot be directly used to determine how they affect a manufacturing environment. Instead, they are often utilized as covariates in statistical analysis to divide the variation attributable to them and better ascertain the main effects of the research variables of primary interest under stated assumptions. Literature would considerably improve if all demographic factors were considered essential in more expansively framed hypotheses.

2.2. Worker performance and demographic variables

This section focuses on work practices that are less common than shift work but can nonetheless lead to fatigue. The duration of working time, shift length and schedule, and the number of off days are only a few factors affecting work schedules. Numerous fatigue-related effects are produced by interactions between these elements and the work undertaken, administrative or physical job control, sleep needs, individual traits, and e.g., non-work (e.g., domestic) circumstances [34]. Experts agree that combining at least three categories of factors – including time spent on a specific task, sleep patterns in physical and psychological state, and the quantity and quality of preceding sleep – causes work-related fatigue [35].

Several research papers have been reviewed in this paper to examine the role of fatigue concerning conventional demographic characteristics, as we believe that they can be a leading cause due to the subjective nature of fatigue. There has never been a research study that looked at the impact of a vast number of demographic variables simultaneously, age, gender, such as socioeconomic status, circadian cycle, etc. [36]. Although, the reviewed research indicates relationships between specific demographic traits and sleep duration, quality, and fatigue in the context of sporadic driving incidents, these relationships still require clarification as to how they affect when all other factors are taken into account. Typically, these studies do not investigate the impact and weighting of each demographic variable on the dependent variable(s) in the same research population. The study samples are clinical convenience samples that may need to be sufficiently represented [37]. As a result, literature frequently indicates methodological faults, limiting the ability to be conclusive.

Understanding the connection between sex and fatigue is difficult since men and women have different employment, levels of supervision, depths of training for work obligations, and specific socio-demographic factors. According to a survey of 1180 Swedish employees, women in low socio-economic levels were 1.4 times more likely to complain of fatigue than women in higher socio-economic positions. In men, there was no evidence of this relationship. In the same study, undertaken, administrative or physical job control, sleep needs, individual traits, and e.g., non-work (e.g., domestic) circumstances [34]. Experts agree that combining at least three categories of factors – including time spent on a specific task, sleep patterns in physical and psychological state, and the quantity and quality of preceding sleep – causes work-related fatigue [35].

Several research papers have been reviewed in this paper to examine the role of fatigue concerning conventional demographic characteristics, as we believe that they can be a leading cause due to the subjective nature of fatigue. There has never been a research study that looked at the impact of a vast number of demographic variables simultaneously, age, gender, such as socioeconomic status, circadian cycle, etc. [36]. Although, the reviewed research indicates relationships between specific demographic traits and sleep duration, quality, and fatigue in the context of sporadic driving incidents, these relationships still require clarification as to how they affect when all other factors are taken into account. Typically, these studies do not investigate the impact and weighting of each demographic variable on the dependent variable(s) in the same research population. The study samples are clinical convenience samples that may need to be sufficiently represented [37]. As a result, literature frequently indicates methodological faults, limiting the ability to be conclusive.

Understanding the connection between sex and fatigue is difficult since men and women have different employment, levels of supervision, depths of training for work obligations, and specific socio-demographic factors. According to a survey of 1180 Swedish employees, women in low socio-economic levels were 1.4 times more likely to complain of fatigue than women in higher socio-economic positions. In men, there was no evidence of this relationship. In the same study, taking care of children at home was not considered in female employees’ fatigue [38]. Sleep (lack of and/or poor quality) is one of the key reasons for people’s fatigue. The shift pattern, on the other hand, is impacted. This study compared gender inequalities in ‘total’ working hours. Despite working more weekly hours, women spent more time on domestic tasks than men. Although the study uses logistic regression for analysis, it could not distinguish between gender disparities in total work hours, it was still helpful in drawing patterns. There was, however, a connection between total weekly hours worked and reduced work capability in women. According to some studies, women may take longer to recuperate from shiftwork than men, although the difference is minimal [36]. However, considering that women are more likely than males to experience and acknowledge fatigue, it is unclear whether the gap is sex generated.

In many developed countries, age is also a highly studied demographic characteristic. Although fatigue affects employees of all ages, it seems to affect elderly individuals the most. There is no specific age for becoming exhausted, although it is thought that as one age, the human body becomes more fatigued. In most research, age is used as a sequential variable [19]. Job ability is defined as the dynamic interplay between an individual’s cognitive and physical resources (e.g., health, motivation), workplace (e.g., workload, work schedules), and social mores [36]. If the studies, methodological faults that make it impossible to draw firm conclusions regarding cause and effect, and various factors can influence the relationship between age and fatigue. Overall, how different demographics will impact the manufacturing industry is still to be determined. However, demographic characteristic variables are generally considered in a survey or questionnaire format; this results in delayed results in understanding the role of demographics in the onset of fatigue. However, a research study by Lambay et al. [39] showed the significance of individual demographic characteristics. It used the demographic characteristics variables in a machine learning model to predict fatigue. This helps determine the variables used to assess fatigue prediction. The paper helps to determine the significance of using the demographic characteristics of humans as a variable to comprehend the development of fatigue.

2.3. Human-robot collaboration

Although industrial revolution 5.0 has seen significant progress in the field of human-centric research, where human–robot collaboration has been successful, it is critical to understand the various applications of this collaboration. Section 2.3 introduces and highlights the numerous applications of human–robot collaboration, allowing readers to better grasp the various usage scenarios. The applications shown in this section have been chosen based on the repetitive movements involved in completing tasks, which are one of the primary causes of human physical fatigue. Additionally, the most significant degree of human–robot collaboration (HRC) involves both parties working together to complete tasks [40]. This may not be limited to physical interaction between the human and the robot. It is also considered to be a complex kind of contact to program and develop a seamless interaction [41].

HRC’s primary objective is to link humans’ intellect, flexibility, and adaptability with the accuracy, quickness, and adaptability of robots to gain the best from both worlds [42]. As a result, task planning and allocation concentrate on finding the best solution based on various factors. Tasks are distributed more than just according to resource availability. The presence of a human worker adds a further layer of uncertainty and unpredictability compared to purely automated systems [42]. This considers the different degrees of worker knowledge, their actual condition, fatigue, and their comfort and ergonomic needs [43]. Therefore, it is anticipated that these criteria would result in ongoing and dynamic adjustments to job allocations and task performance which can then be adapted by the robot accordingly. Overall, this emphasizes the necessity of realistic human models so that the collaborative robot can comprehend its counterpart depending upon their fatigue level.

Fig. 1 depicts examples of how collaborative robots are employed in four working areas where maximum repetitive movements are involved. All of these require human attention and physical exertion (perhaps not as much as without robot assistance) in repetitive tasks. As aforementioned in the above section, fatigue induced in an operator can hinder cognitive ability. For instance, from the Fig. 1 ‘A’, a study [48] shows how fatigue is induced due to repetitive motion which causes physical exertion eventually leading to fatigue during working on a assembly task. This could lead to accidents and quality deficiency in the manufacturing process. We observed that various research in the field of HRC had been dedicated to intention recognition, trust building, decision making, task recognition, etc. [49]. However, all these factors studied in HRC research can change when a person is fatigued, making the detection of fatigue is important.

In the paragraph mentioned above (in 2.3), the presumption that a person has a consistent degree of physical endurance is present in most HRC situations. This presumption can be regarded as accurate for simple jobs requiring little physical effort or time to complete. However, other complicated and unpredictable interaction situations can impact how well people function. In these mentioned scenarios
Fig. 1. Human–robot collaboration work settings. A — Assistance in Assembly work like positioning or supporting heavy components [44], B — Assistance in quality inspection [45], C — Collaborate with humans with manual material handling [46], D — Assistance in several machining processes such as drilling, polishing, etc. [47].

(Fig. 1), due to the nature of the procedure, the effort put in by the human was minimized; nonetheless, fatigue was not measured nor tracked. In contrast to human–robot collaboration, the human partner is susceptible to physical fatigue. This phenomenon can have an abrupt and unpredictable impact on their ability to engage physically [50]. In this situation, the robot should be able to detect human exhaustion and modify its behavior to provide extra physical assistance with the assigned work. Consequently, the human partner should exert less physical effort and monitor the work cognitively.

3. Review methodology

A systematic literature review is used to find, evaluate, and analyze the literature published between 2000 and 2024 to examine the study and implementation of machine learning for detecting, monitoring, and classifying fatigue. The year range is chosen as per the inclusion and exclusion criteria such as which considers ML in it. Fig. 2 depicts the general process in detail. First, a systematic literature review is used to find, choose, and evaluate pertinent papers. A systematic and transparent approach for locating, assessing, summarizing and documenting research work generated by researchers is known as a systematic literature review. The review process typically consists of many vital processes, such as defining the research questions, identifying the research, and choosing and evaluating the publications that have been gathered. The article selection and evaluation processes used a set of rules for the exclusion and inclusion criteria to assess each possible primary research. To understand the leading research themes and directions in machine learning applications regarding fatigue detection.

3.1. Research questions

Table 1 presents the research questions and their motivations for the systematic literature review.

3.2. Search strategy

3.2.1. Search term identification and resources

The search strategy used in this study covers keyword identification, searching resources, the search procedure, and article selection criteria to gather published articles that are qualified and pertinent to the topic. The keywords were as follows:

("Human Physical Fatigue", OR "Occupational fatigue" OR "Fatigue" OR "Work fatigue" OR "Muscle Fatigue Index" and "Human–Robot Collaboration" OR "HRC Settings" OR "Machine Learning" OR "Muscle Fatigue Model" "Reinforcement Learning" or "Supervised Learning and "Functional Data Analysis" OR "Human Performance Modeling" OR "Feature Selection")

Different databases, such as the IEEE Xplore digital library, Science Direct, ACM digital library Springer Link, Scopus, and Google Scholar, were used to search for pertinent publications. These databases contain a vast amount of material, including journal papers, conference proceedings, and books. They are the most representative databases of scientific research directly relevant to the subject of this review.

3.2.2. Article screening

The procedure for choosing articles is outlined in the above figure (Fig. 2). The initial step in the search procedure is to search for articles by the given keywords in the preset databases, which leads to finding 9524 published research works. Five thousand three hundred forty-three publications were left after these publications underwent an initial screening procedure based on the suggested exclusion criteria. Sixty-eight papers were included after the publications were chosen using the proposed inclusion criteria in Section 3.2.3. A manual search technique was also employed to find any new sources pertinent to the review not identified during the database search stage. The exclusion and inclusion criteria were also applied again. The manual search procedure resulted in the selection of 33 additional publications. The following criteria were used to evaluate the quality of the selected papers during the quality assessment step, and 210 publications were ultimately found.

3.2.3. Inclusion exclusion criteria

The inclusion and exclusion guideline for the article selection process is based on the research questions (shown in Table 1) and the keywords. While inclusion criteria were implemented on the whole text, exclusion criteria were applied to the publication’s title, abstract, and keyword list. The study’s exclusion criteria used are as follows:

- The article emphasizes various techniques for detecting fatigue, such as using only subjective measures.
- Articles published in a language other than English.
- Articles that utilized ML techniques were more concerned with other human factors other than physical fatigue.
The study's inclusion criteria used are as follows:
All the English-language papers published on ML techniques used to detect, monitor, and classify human physical fatigue.

Articles that present novel methods for enhancing the effectiveness of machine learning techniques used for detecting physical fatigue in HRC settings.

4. Machine learning for human fatigue

4.1. Detection of human fatigue

In Section 4.1, numerous approaches used by researchers to detect and monitor human physical fatigue are introduced and discussed. Section 4 provides an in-depth review of the many methodologies used by the research community for fatigue detection. The methodologies used by various researchers to assess and quantify human physical fatigue are revealed through in-depth investigation and analysis. By delving into the subtleties of these procedures, readers obtain a better understanding of the breadth and depth of fatigue detection and monitoring strategies. Furthermore, Section 4.1 acts as an introductory structure for the remainder of the sub-sections, preparing the way for a more in-depth examination of machine learning techniques and their application in fatigue detection in later sections. Many researchers over the decades have used three main types of procedures to detect fatigue. These are (1) physiological changes in the body, (2) gait and facial behavior and (3) questionnaire and interview type. The first two types for this study.

However, as mentioned above, in the exclusion criteria, we only include the first two types for this study.

Numerous models and dimensions are utilized in various evaluation procedures, to understand how fatigue manifests itself differently depending on the workplace and individual. A proper assessment of fatigue aids in preventing worker injuries, lowering associated costs, and designing appropriate shifts and work-rest schedules [57]. Many pieces of literature study/research the detection of fatigue. Much of the work is also focused on the transportation or sports industries. Fig. 3 shows the different physiological sensors used to detect fatigue. All these sensors shown are used the majority of time to detect fatigue in a manufacturing or industrial environment, except the pupillometry and muscle activity sensors, primarily used in driver fatigue prediction [56].

The data collected from these sensors are then utilized in a machine learning process to classify whether a person is fatigued or not. The main ML methods used for the detection and monitoring of human fatigue include supervised learning, unsupervised learning, neural networks, fusion or composite methods (which incorporate two or more processes to develop a model), and statistical analysis applied to the collected data. Fig. 4 provides an overview of the number of publications utilizing these methods for detection.

Machine learning modeling goes through various steps to create an effective and accurate system for detecting human physical fatigue.
Fig. 3. Physiological sensors used in the detection of fatigue [55,56].

Fig. 4. Graph for the number of studies showing the use of different ML methods for training and detection of fatigue.

Fig. 5 demonstrates an overview of the general method followed by researchers for detecting human physical fatigue in various scenarios including HRC. Initially, data is collected using a variety of sensors capable of detecting physiological, biomechanical, and behavioral signals assessing fatigue. These sensors may include heart rate monitors, EMG, EEG sensors, IMUs, and others, depending on the parameters being measured. Once collected, the data is preprocessed to reduce noise and extract essential features that represent various elements of fatigue. Methods for extracting features may include supervised/unsupervised form of learning, deep learning neural networks etc.

Following preprocessing, the data is separated into training and testing sets, which are used for model building and evaluation. During the training phase, different machine learning algorithms are applied to the labeled data to create predictive models. These algorithms can
range from basic statistical methods like logistic regression to more advanced techniques like support vector machines (SVMs), random forests, and deep learning neural networks; or in some cases statistical analysis [58]. The nature of the data, the complexity of the task, and the desired performance indicators all influence the algorithm selection process.

Once trained, the models are evaluated using the testing dataset to determine their accuracy in identifying fatigue. Accuracy, precision, recall, and F1-score are standard evaluation metrics used to quantify model effectiveness. Furthermore, procedures such as cross-validation may be used to confirm the model’s robustness and generalizability across multiple datasets. After analyzing the model’s performance, it may be further refined using approaches such as hyperparameter tuning or feature selection to improve its performance. Finally, the trained and validated model is used in real-world scenarios to continually monitor and detect indicators of fatigue in people participating in physical activities, especially in Human–Robot Collaboration (HRC) situations. Continuous monitoring enables prompt actions to minimize fatigue-related incidents while also improving workplace safety and productivity.

4.2. Fatigue detection through physiological changes in the body

The central nervous system’s electrical pulses are utilized by the physiological sensors for detecting physical fatigue. Electrodes connected to the body are then used to detect these electrical pulses. These electrodes transform the electrical pulse that is detected into signals that are examined for fatigue detection. The most used sensor is Electromyography (EMG). The other commonly used ones are force sensors, electrodermal activity (EDA), electroencephalograms (EEG) and electrocardiograms (ECG or EKG). The sensors are attached to the body part for which the fatigue is to be detected by detecting the changes in the electrical pulse generated in the body. The signal acquisition device records a signal produced by these sensors. A filter is then used to remove any unwanted noise or errors from the signal. The preprocessed signal’s quantifiable/functional features can then be extracted. Machine learning can then be used to classify human states, and extracted features from physiological sensors such as weariness, using the derived information. Table 2 shows the generally extracted features from physiological sensors.

Jasper built an electromyograph and utilized it to demonstrate novel research on epilepsy and neurology [59]. Since 1960, sEMG has often been employed in clinical studies. Basmajian and De Luca significantly impacted EMG research [60]. Electrodes are used in EMG to track electrical currents produced by contracting muscles [61]. Since the signal produced by EMG is usually complex and noisy, it typically needs to be filtered. The signal’s amplitude, and frequency, as a function of time, are some other parameters that can be used to describe it [62].

Electrodes can be applied to the skin’s surface to gather EMG signals non-invasively or invasively (using needle electrodes). Surface EMG (sEMG), the latter of which is non-invasive, is a popular technique for gathering signals from tired muscles during static and dynamic contractions [59]. Although sEMG can be captured from various body locations, electrode placement is crucial for obtaining accurate and consistent data [63]. Various one- and two-dimensional multi-electrodes’ propensity to offer the best selectivity and minimize crosstalk were researched by Dimitrov et al. [64]. They discovered that the new bi-transversal double differentiating electrodes provided the highest sensitivity and minimum crosstalk, especially when positioned above the end of the muscle section. The electrode’s location impacts the signal that is collected because it affects the contracting muscle electrical activity that is recorded.

To achieve a dependable and steady surface electrode signal, it is also crucial to position the electrodes to minimize crosstalk with the signal from other surrounding muscles [65]. A background noise level can be set, and the electrode positions and spacing can be optimized to reduce crosstalk. According to Gerdle et al. [66], the ideal spacing between electrodes is when the standard deviation of the noise in the signal should be less than the radius around the electrode with the maximum signal amplitude. One of the EMG’s drawbacks is the many noise sources that are picked up with the signal. Electrical circuits inside the EMG emit electromagnetic radiation, as do the surfaces of all, including artificial things and human beings. EMG electrodes transducer noise when they touch the skin [59]. Furthermore, several other factors hinder the quality of an EMG signal. The blood flow rate, diameter, location and depth of muscle fiber, the number of tissues between the muscle surface and place electrode, etc. [66]. These factors vary from person to person, increasing the complexity of detecting the signal.

Although the complexity mentioned earlier is present, EMG has been employed in several studies concerning the accumulation of localized muscle fatigue (LMF). Ollivier et al. [67] employed EMG to

Fig. 5. General process flow of human physical fatigue detection including HRC scenario.
understand the localized muscle fatigue accumulated during elbow flexion. The authors used root mean square (RMS) values to analyze and detect fatigue. The research is concentrated on different bipolar electrodes and the differences in the force values. Similarly, MLS Independent Component Analysis (ICA) techniques were applied to detect LMF when standard load exerts isometric contractions (I.C.) [68]. The study found that the motor unit’s conduction velocity is reduced compared to high-level contractions during a low bicep contraction. These studies were conducted in the lab, and in real-life scenarios, the factors affecting are different. Like Renberg et al. (2020) implemented the EMG analysis in cold conditions. They found that there were no significant changes in the activations of motor units in cold conditions [69]. However, when applied in the field research by Chowdhury et al. they found that dynamic isometric contractions are higher by 19.63% when compared with static contractions [70].

4.2.2. Fusion of sensor data for feature extraction

The use of only one sensor to measure localized muscle exhaustion has several restrictions, as was previously indicated. These restrictions have been attempted to be solved by combining two or more sensors to assess overall or localized muscle exhaustion. As an alternative, efforts to use other signal analysis methods have also been made. The features generated from various sensors can be combined and the best set of features can be then used to achieve higher accuracy of fatigue prediction. Techniques such as cross-validation and leave-one-out are widely used for achieving higher results by combining them with different supervised learning techniques [4,6].

Liu et al. [97] used an amalgamation of Force, EMG and MRI sensors to try and understand the patterns of the onset of fatigue during a handgrip task. They used image recognition techniques with statistical analysis to understand the significance of different patterns in the onset of fatigue. They proposed an Automated Image Registration (AIR) algorithm to train and test different patterns of the development of fatigue. It can be observed within Table 4 that similar attempts by several other authors are made using a force sensor and EMG together to examine human physical fatigue [89–95,98,100]. As an alternative to force sensors, the EMG sensor has also been coupled with other sensors. Jun Shi et al. (2007) used sonomyography with EMG signals to detect fatigue [99]. Utilizing ultrasonography, sonomyography (SMG) describes the physical and morphological alterations of skeletal muscles [107]. Furthermore, the force sensor and EMG with an EEG sensor were used to examine fatigue detection in lower limbs [104]. The additional part of this examination was that they used subjective analysis. They used Borg Test to understand the personal evaluation of fatigue. Equivalently, Fujisawa and the authors [101,102] used EMG and Ultrasound to predict fatigue accumulated in the isometric contractions of the knee at various angles. The same principles were used by Li et al. [102] to examine the isometric contraction of upper limbs at multiple angles.

4.3. Gait and behavioral characteristics used for fatigue detection

In 2000, Kazo Saito [108] tried to evaluate and measure industry fatigue using a portable fatigue meter. The meter would have an LED light indicator to display the fatigue state. It used the Critical Flicker Fusion frequency (CFF) combined with the Visual Reaction Test to evaluate the fatigue state. It had limitations, and a heart rate and voice recognition headset were used to overcome them. It was analyzed through an accurate speech recognition algorithm to evaluate fatigue [106]. Substitute to that using an optical motion capture system along with 13 markers and haptic gloves to track the movement of joints and fingers to evaluate whole-body fatigue [109]. The task was to evaluate fatigue during a manual material handling task. Additionally, virtual reality maps the person and generates fatigue mitigation. However, all these techniques are a decade old, and using them in an
Table 3  
Human physical fatigue detection techniques using physiological sensors data.

<table>
<thead>
<tr>
<th>ML Tech.</th>
<th>Task</th>
<th>Key Result @ Fatigue State</th>
<th>Sensor</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Changes in electrode analysis</td>
<td>Maximal elbow flexions</td>
<td>Increase in RMS value for Laplacian</td>
<td>EMG</td>
<td>[67]</td>
</tr>
<tr>
<td>ICA</td>
<td>Isometric Contraction</td>
<td>Reduced Motor units in high-level contractions</td>
<td>EMG</td>
<td>[68]</td>
</tr>
<tr>
<td>Discrete Wavelet Transform</td>
<td>Manual material handling</td>
<td>Spectral change for fatigue detection</td>
<td></td>
<td>[83]</td>
</tr>
<tr>
<td>Fatigue Index Model</td>
<td>Repetitive upper arm exertion</td>
<td>Fatigue detection by increase in ARV RMS &amp; PTP</td>
<td>EMG</td>
<td>[78]</td>
</tr>
<tr>
<td>STFT &amp; CWT</td>
<td>Detects by changes in Force signal</td>
<td></td>
<td></td>
<td>[79]</td>
</tr>
<tr>
<td>wavelet transform &amp; ANN</td>
<td></td>
<td>High correlation coefficients 0.93 avg</td>
<td></td>
<td>[80]</td>
</tr>
<tr>
<td>MFDMA &amp; SSM function; IG &amp; GA</td>
<td>Repetitive curl exercises</td>
<td>60% features statistically significant</td>
<td></td>
<td>[81]</td>
</tr>
<tr>
<td>FFI Algorithm</td>
<td>Fatigue detection through window sizes</td>
<td></td>
<td></td>
<td>[70]</td>
</tr>
<tr>
<td>Body Energy Model</td>
<td>Manual material handling</td>
<td>Limitations such as energy usage observed</td>
<td></td>
<td>[84]</td>
</tr>
<tr>
<td>Spectral Analysis</td>
<td>Increase in RMS &amp; No decrease in MVC</td>
<td></td>
<td></td>
<td>[69]</td>
</tr>
<tr>
<td>Linear regression</td>
<td>Mean Frequency decreases</td>
<td></td>
<td></td>
<td>[82]</td>
</tr>
<tr>
<td>ANNOVA</td>
<td>Manual exertion exercise task</td>
<td>Right postsures of forearms analysis</td>
<td>Potentiometer</td>
<td>[85]</td>
</tr>
<tr>
<td>SVM</td>
<td>Isometric Contraction task</td>
<td>Accuracy of 83.33%</td>
<td>EDA</td>
<td>[86]</td>
</tr>
<tr>
<td>Amplitude of signal Time domain analysis</td>
<td>Repetitive hand exercise</td>
<td>Significant result of compression test (ANNOVA)</td>
<td>EEG</td>
<td>[87]</td>
</tr>
<tr>
<td></td>
<td>Assembly line task</td>
<td>67% accuracy, time domain measure</td>
<td>Heart Rate</td>
<td>[88]</td>
</tr>
<tr>
<td>Back-propagation neural network</td>
<td>Hand Grip isometric contractions</td>
<td>Decrease in Handgrip Force after Fatigue</td>
<td></td>
<td>[95]</td>
</tr>
<tr>
<td>Signal amplitude analysis</td>
<td>Repetitive contractions exercise</td>
<td>Significant changes through ANOVA</td>
<td>EEG &amp; Force</td>
<td>[87]</td>
</tr>
<tr>
<td>Automated Image Registration</td>
<td>Patterns recognized for force loss</td>
<td></td>
<td>EMG &amp; MMG</td>
<td>[96]</td>
</tr>
<tr>
<td>Speckle tracking algorithm</td>
<td>Less force = fatigue = lower frequency</td>
<td></td>
<td></td>
<td>[97]</td>
</tr>
<tr>
<td>RMS &amp; MDF of EMG signal</td>
<td>Linear increase &amp; decrease in RMS &amp; MDF</td>
<td></td>
<td>EMG &amp; Ultrasound</td>
<td>[98]</td>
</tr>
<tr>
<td>MVC RMS and cw-NIRS</td>
<td>75% higher MVC for sEMG &amp; cw-NIRS</td>
<td></td>
<td></td>
<td>[99]</td>
</tr>
<tr>
<td>Spectral Power &amp; Peak Torque</td>
<td>Repetitive knee exercise</td>
<td>Higher sp &amp; pt in men as compared with MF</td>
<td></td>
<td>[100]</td>
</tr>
<tr>
<td>Ultrasound image entropy (USIE)</td>
<td>USIE showed clear changes in body muscle</td>
<td></td>
<td></td>
<td>[101]</td>
</tr>
<tr>
<td>Ultrasound signal estimation</td>
<td>Repetitive contractions exercise</td>
<td>70% accuracy</td>
<td></td>
<td>[102]</td>
</tr>
<tr>
<td>Cluster analysis</td>
<td>13% &amp; 25% decrease @ MVC &amp; RMS</td>
<td></td>
<td>EMG &amp; EEG</td>
<td>[103]</td>
</tr>
<tr>
<td>Penalized Logistic &amp; Regression</td>
<td>Manual material handling</td>
<td>Accuracy over 80% obtained</td>
<td>IMU &amp; Heart</td>
<td>[104]</td>
</tr>
<tr>
<td>Statistical, Ensemble &amp; Single Classifiers</td>
<td>Best model-Random Forest — accuracy 87%</td>
<td></td>
<td></td>
<td>[1]</td>
</tr>
<tr>
<td>Recurrent Neural Network</td>
<td>Accuracy of 63%</td>
<td></td>
<td></td>
<td>[105]</td>
</tr>
<tr>
<td>Accurate Speech recognition</td>
<td>Reading after fatigue exercise</td>
<td>82% in W and 72% in M accuracy</td>
<td>Heart &amp; Voice</td>
<td>[106]</td>
</tr>
</tbody>
</table>

Table 4  
Human physical fatigue detection techniques through different sensor fusion data.

<table>
<thead>
<tr>
<th>ML Tech.</th>
<th>Task</th>
<th>Key Result @ Fatigue State</th>
<th>Sensor</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>CWT</td>
<td>Repetitive arm exercise</td>
<td>No significant results</td>
<td></td>
<td>[89]</td>
</tr>
<tr>
<td>CWT &amp; Linear Regression</td>
<td>Increase in RMS value for fatigued state</td>
<td></td>
<td></td>
<td>[90]</td>
</tr>
<tr>
<td>ANNOVA statistical analysis comparison</td>
<td>Exercise as industry actions</td>
<td>Changes in MVC for detection</td>
<td>EMG &amp; Force</td>
<td>[91]</td>
</tr>
<tr>
<td>MDF &amp; MNF vs force</td>
<td>Manual material handling</td>
<td>Significant difference ANOVA test</td>
<td></td>
<td>[92]</td>
</tr>
<tr>
<td>Back Propagation NN</td>
<td>Screw driving task</td>
<td>force loss — GS, DT, and PF, 32.1%, 24.2%, 26.5%</td>
<td></td>
<td>[93]</td>
</tr>
<tr>
<td>Back-propagation neural network</td>
<td>Hand Grip isometric contractions</td>
<td>Decrease in Handgrip force after fatigue</td>
<td></td>
<td>[94]</td>
</tr>
<tr>
<td>Signal amplitude analysis</td>
<td>Repetitive contractions exercise</td>
<td>Significant changes through ANOVA</td>
<td>EEG &amp; Force</td>
<td>[87]</td>
</tr>
<tr>
<td>Automated Image Registration</td>
<td>Patterns recognized for force loss</td>
<td></td>
<td>EMG &amp; MMG</td>
<td>[96]</td>
</tr>
<tr>
<td>Speckle tracking algorithm</td>
<td>Less force = fatigue = lower frequency</td>
<td></td>
<td></td>
<td>[97]</td>
</tr>
<tr>
<td>RMS &amp; MDF of EMG signal</td>
<td>Linear increase &amp; decrease in RMS &amp; MDF</td>
<td></td>
<td>EMG &amp; Ultrasound</td>
<td>[98]</td>
</tr>
<tr>
<td>MVC RMS and cw-NIRS</td>
<td>75% higher MVC for sEMG &amp; cw-NIRS</td>
<td></td>
<td></td>
<td>[99]</td>
</tr>
<tr>
<td>Spectral Power &amp; Peak Torque</td>
<td>Repetitive knee exercise</td>
<td>Higher sp &amp; pt in men as compared with MF</td>
<td></td>
<td>[100]</td>
</tr>
<tr>
<td>Ultrasound image entropy (USIE)</td>
<td>USIE showed clear changes in body muscle</td>
<td></td>
<td></td>
<td>[101]</td>
</tr>
<tr>
<td>Ultrasound signal estimation</td>
<td>Repetitive contractions exercise</td>
<td>70% accuracy</td>
<td></td>
<td>[102]</td>
</tr>
<tr>
<td>Cluster analysis</td>
<td>13% &amp; 25% decrease @ MVC &amp; RMS</td>
<td></td>
<td>EMG &amp; EEG</td>
<td>[103]</td>
</tr>
<tr>
<td>Penalized Logistic &amp; Regression</td>
<td>Manual material handling</td>
<td>Accuracy over 80% obtained</td>
<td>IMU &amp; Heart</td>
<td>[104]</td>
</tr>
<tr>
<td>Statistical, Ensemble &amp; Single Classifiers</td>
<td>Best model-Random Forest — accuracy 87%</td>
<td></td>
<td></td>
<td>[1]</td>
</tr>
<tr>
<td>Recurrent Neural Network</td>
<td>Accuracy of 63%</td>
<td></td>
<td></td>
<td>[105]</td>
</tr>
<tr>
<td>Accurate Speech recognition</td>
<td>Reading after fatigue exercise</td>
<td>82% in W and 72% in M accuracy</td>
<td>Heart &amp; Voice</td>
<td>[106]</td>
</tr>
</tbody>
</table>

actual working environment would be costly and time-consuming. Also, it causes much discomfort to the operator. An alternative to that is to use an IMU and heart sensors. The inertial measurement unit (IMU) is the most used sensor for observing the gait changes in an individual. IMU comprises accelerometers, gyroscopes, and magnetometers, which observe gait behavior. These sensors are wireless and will cause less body discomfort. Maman et al. [110] used an IMU and heart rate sensor to evaluate fatigue. The sensors were attached at five locations to understand the localized and whole-body fatigue. They use Balloon Analogue Risk Task (BART), psychomotor vigilance task and Borg Test to subjectively evaluate. These evaluations were then combined with the features extracted from the data obtained. After combining, the data set was fed through Penalized logistic and regression ML models to classify fatigue and non-fatigued. They achieved an accuracy of over 80%. Similar hardware was used by Zahra Sedighi and authors [111] to understand which machine-learning models gave the highest accuracy. They evaluated three different types: statistical, ensemble and single classifier models. They observed that the Random Forest model achieved the highest accuracy of 87%. All other techniques are included in Table 5.

4.4. ML for monitoring/recovery after human fatigue detection

Even though fatigue has been studied for decades, a complete fatigue model still needs to be designed and produced accessible for a variety of real-life workplaces and other scenarios. Every major development in the field of fatigue research has been made in relation to sports and transportation, but there is still room for advancement in the industrial setting. Many efforts have been made to research the state of human fatigue in an industrial environment, but minimal effort has
been made to monitor it. It is the case that fatigue is subjective, and the onset can occur for different reasons that vary from person to person. For example, if a person works an 8-hour shift, they can be fatigued initially. Also, it can accumulate through the repetitive tasks performed throughout the day. This makes it essential to monitor human fatigue to avoid accidents or MSDs in the long run. Not only to monitor but also to create a recovery method. Previous methods included surveys and questionnaires to monitor. Nonetheless, this method is effective as it considers subjective feelings but can barely produce results in real time. Due to the intermittent recording style, the human can be in other states till the results are out. Physiological sensors can provide real-time monitoring of the development of fatigue in a repetitive manual material handling task. They used multivariate time series analysis to evaluate and understand features generated from an IMU and heart rate sensor to monitor.

A simple method was proposed by using an optical camera to record, track and monitor the fatigue levels of human operatives working on a computer table [120,121]. They employed Bayesian and Dynamic Bayesian Networks to follow the Facial cues such as eye and gaze monitoring. However, this system had limitations in tracking specific areas and specific tasks. An alternative solution was to employ EMG and dynamometer to monitor the maximum voluntary contractions by analyzing the MDF, MNF, and PSD generated from the signals [122]. Analysis of MNF, MDF, and PDF was observed to change over time, giving the estimation of fatigue level. Applying similar principles of the hardware to monitor, proposed a recovery system by using ARV and Fractal Dimension (F.D.). A recovery system named “light emitting diode therapy (LEDT) for recovery” was proposed [126]. In different approaches amalgamating EMG and force transducers were used to present a recovery after the fatigue state was detected by comparing the MVC of pre-and post-fatigue conditions [127]. The drawback of this system is that force estimation can vary from person to person and different genders.

Finding the proper force can be difficult and there are many false alarms. However, posture can say a lot about a person’s physical state. Applying this principle, a Rapid Entire Body Assessment (REBA) posture analysis was employed through a quadratic equation to monitor and propose a recovery plan [125]. This requires high computational time. Signals generated from an ultrasound transducer can be used to monitor the development of fatigue in a repetitive task [123]. The vibrations created can be used to monitor human fatigue. However, the system’s reliability could be better, as the vibrations can be due to any movement and can classify as a false alarm. Baghdadi and the authors [129] combine subjective and objective methods to monitor human fatigue in a manual material handling task. They used the Physical Activity Readiness Questionnaire and Borg test with the features generated from an IMU and heart rate sensor to monitor. They used multivariate time series analysis to evaluate and understand the development patterns of fatigue in a repetitive manual material handling task. Briefly outlined are all the monitoring and recovery research studies in Table 6.
4.5. HRC settings assisted by ML for human fatigue

These robots are designed and made keeping specific goals in mind to overcome the barriers of conventional robots with a safety net around them. The robots can work together and assess the human condition and adapt accordingly to human needs, i.e., a human-centric manufacturing environment. Although many advancements have been made for seamless human–robot interactions, they lack how humans adapt and operate. Especially, much research has been undertaken to create an understanding, recognition, and adaptation to human actions, some by human demonstration or some by robot learning. In these, EMG has played a crucial role [130]. Although these technologies have advanced, it is essential to detect the human fatigue state as seeing the previous section; if a person is fatigued, it can cause a loss in concentration, lower quality, etc. Note that there needs to be more work completed in this area of HRC.

Nevertheless, Peternel and co-authors [131–138] undertook fatigue detection, monitoring, and adaption. They employed an EMG sensor and force transducers to evaluate the joint angle torque while operating a physical exertion industrial task such as polishing, drilling, etc. They employed various machine learning techniques to understand the fatigue pattern and evaluate it accordingly for different tasks. They used the Gaussian Process Regression (GPR) model to classify fatigue levels for a predetermined human and robot polishing task [138]. They also employed ‘Dynamical Movement Primitives’ (DMP), ‘Locally Weighted Regression’ (LWR) and ‘Adaptive Frequency Oscillators’ (AFO) to detect and adapt according to the force generated by the human counterpart [132]. Applying similar principles, the model was improvised using the ‘Human force estimation adaption model’ for path planning and process following procedure for the robot [139].

Pramanick and authors [148] developed a ‘Defatigued’ model for the adaption of robots according to humans. They used only force sensors attached to the end of the effector to estimate the human working force. The major drawback of this method is that it does not consider the physiological changes in the muscle fibers that contribute to human fatigue. Wang et al. [140] used a similar working model by Peternel et al. Gaussian mixture models (GMMs) and updated it with the weighted system for adaption for an assembly task. Instead of only using an EMG sensor alone. Tsung-Chi et al. added the Vicon motion capture camera to track human movement when assisted by the robot in an assembly task [146]. A similar approach was to track the movement, but different sensors were used. They used an EMG and MVN Biomech suit to track humans by the robot [141]. The authors used MDF to evaluate fatigue. The fairly new research was undertaken by Kumar et al. by using the Riemann geometry features generated by a ‘myoelectrical’ armband [142]. They used SVM and Random Forest ML techniques to classify between fatigued and not fatigued. All the undertaken HRC studies for fatigue management are shown in Table 7. Furthermore, Fig. 6, indicates the number of studies involved in the different form of HRC setting in order to study fatigue detection and mitigation.

4.6. Summary for fatigue detection

Research conducted on human physical fatigue detection and in human–robot collaboration scenario was reviewed including different types of detection processes used, including physiological-based or behavioral-based. ML has played a crucial role in the detection of fatigue and in analyzing or generating features.

In the review, the three most representative fatigue detection processes are presented which are most often used to evaluate the onset of fatigue. Tables three to five are further segregated according to the type of detection processes used, including physiological-based or behavioral-based. ML has played a crucial role in the detection of fatigue and in analyzing or generating features.

In the review, the three most representative fatigue detection processes are presented which are most often used to evaluate the onset of fatigue. Tables three to five are further segregated according to the type of detection processes used, including physiological-based or behavioral-based. ML has played a crucial role in the detection of fatigue and in analyzing or generating features.
(CNN), which is well known for image recognition, has not been used to derive patterns for the onset of fatigue. Most of the researchers have chosen to evaluate the EMG signal through time series generated features or frequency generated features. In these features, the adaptation of statically analysis tests such as ANOVA has been extensively used to understand the significant difference between fatigue and non-fatigue state. In time and frequency analysis, the features generated are mean frequency (MFN), median of frequency (MDF), root mean square (RMS), maximum voluntary contractions (MVC), and spectral density (SD). The alternatives used by researchers are continuous wavelet transform (CWT) or Fast Fourier transform (FFT). Apart from these, only one researcher used multifractal detrended moving average algorithm (MFDMA) and multifractal singularity spectrum (SSM) function; information gain (IG) and genetic algorithm (GA) techniques.

Gait or behavioral characteristics comprise the second type of detection method which presents an alternative to physiological sensors. Most of the research in this section is conducted through an inertial measurement unit (IMU). IMU studies have used support vector machine (SVM) as their go-to ML algorithm. IMU has been employed with different classifiers such as ensemble to predict fatigue. The most generated feature is a jerk. The alternative to ML techniques is using time series analysis. However, in this sector, the use of optical motion cameras, force sensors and acceleration sensors are provided as an alternative to IMU.

Section 4.5 investigates the use of machine learning (ML) techniques in human–robot collaboration (HRC) contexts to detect and manage human physical fatigue. It emphasizes the significance of developing cobots that can analyze human circumstances and adjust accordingly, resulting in a human-centered manufacturing environment. While advances have been made in seamless human–robot interactions, there is still a lack of understanding about how humans adapt and operate in fatigue contexts. Several studies have investigated intention detection, gesture recognition, and adaptability to human behaviors, with electromyography (EMG) playing an important part. For instances, Paternel and co-authors conducted an impressive study on fatigue detection, monitoring, and adaptation utilizing EMG sensors and force transducers during strenuous jobs like polishing and drilling. They used machine learning techniques like Gaussian Process Regression (GPR), Dynamical Movement Primitives (DMP), Locally Weighted Regression (LWR), and Adaptive Frequency Oscillators (AFO) to analyze and adapt to fatigue patterns. Another study by Pramanick and colleagues created a ‘Defatigued’ model for robot adaptation using force sensor data. These studies demonstrate how various machine-learning techniques can be used in HRC situations to control fatigue, contributing to a better understanding of human–robot collaboration and fatigue detection.

5. Discussion

Although more ML technologies are being used in HRC systems and fatigue detection and monitoring, there are still several challenges to solve and opportunities to seize. Challenges are discussed in this section in light of the earlier literature review. Opportunities for future ML for human fatigue management and study are also addressed.

5.1. Challenges

5.1.1. Current machine learning techniques

Fatigue is very subjective and can result from numerous factors other than physical exertion. According to the reviewed articles, ML algorithms can be crucial in classifying or analyzing the signal for detecting different fatigue levels. However, the accuracy of any given ML algorithm depends on the data input for training and testing purposes. Several factors can impact data collection and feature generation. One of the examples is that profound knowledge of the domain is necessary, especially when handling sEMG data of ML detection. Many authors have used different signal analysis techniques, such as time–frequency analysis. These provide valuable, usable details for machine learning. However, there is no fixed or universally accepted method for selecting the signal’s sample size or input windows [116].

The use of ML for detection or monitoring requires a user to set a threshold or cut-off score for the features generated. This will enable a user to determine fatigue if the values are above that level. However, there is no standard for setting these scores/values and it is completely subjective to the person conducting the study. The lack of knowledge establishing the cut-off score of each physiological metric for physical forms of fatigue presents a significant obstacle to employing physiological metrics to quantify physical fatigue during physical exertion tasks. Little is known about the threshold scores regarding the physiological signals to signify extreme fatigue. Even though persistent physical fatigue is assumed to increase the risk of musculoskeletal problems and work-related injuries. Since each person responds differently to fatigue when performing different jobs in various surroundings as
mentioned earlier, this constitutes a significant issue when interpreting the physiological measurements collected through any given sensor for fatigue detection and monitoring.

5.1.2. Current data acquisition system and real-world implementations

The human physical fatigue modeling is a complex procedure to determine the human fatigue level at any given time. The modeling process involves many different features to be generated, which involves significant sensor data accuracy. However, the current research only involves some factors contributing to the development of fatigue. The sensors can detect physiological changes or gait behavior but do not consider the various external factors contributing to the onset of fatigue. As is known very well, fatigue can be caused by multiple factors, such as disturbed circadian rhythm, stress, etc. The sensors cannot detect these changes. Apart from the mentioned factors, some sensors lack the detection of changes in the human body, such as internal changes in the body, like muscle fibers. For example, although EMG can detect changes in muscle contraction, it lacks the identification of the changes in muscle fibers, which are one of the contributing factors of muscle fatigue. As observed from the reviewed papers, many researchers choose EMG sensors only. These sensors need to be prepared before any data can be collected. For example, they must be calibrated before use, and special care must be taken.

Furthermore, before any data can be collected, the human skin also needs preparations. For example, the human skin area needs to be cleaned with alcohol, and if any hair is present, that area needs to be shaved and cleansed. This preparation in an industrial environment raises the question of validity, usability and repeatability for every use. Apart from the trials, using an EMG sensor needs special care when placing the sensor at the right location, or it would generate much noise in the signal [61]. These limitations question usability in a manufacturing environment. Furthermore, the new wireless system of the EMG signal requires it to be stuck onto the skin of the body part for accurate electrode contact to generate low-noise data. This could cause discomfort to the human operator.

Additionally, in a working environment, especially an industrial setting, the EMG sensor can come loose and cause data loss for monitoring. On the other hand, hardware such as motion capture sensors or cameras are expensive and track a concise path for the human operator. An alternative solution used in the reviewed papers is an IMU sensor. It does give the edge over EMG for placement and preparations but comes with some limitations. The IMU measures the acceleration, angular movements, and magnetometer, which provides the gait behavior of an operator [4]. However, it fails to detect the localized muscle activity contributing to physical fatigue.

The reviewed paper employed force sensors for fatigue detection. Force sensors/transducers do give accurate results when coupled with other sensors. When used, change in human force is not necessarily due to fatigue, although it is one of the causes. For example, Human strength could reduce due to a lack of attention. Nonetheless, force estimation for a job can be subjective and vary from person to person or different genders. Lastly, ultrasounds sensors were used for fatigue estimation, which has a drawback. Firstly, they estimate the body’s vibrations, which could result from external factors in an industrial environment where lots of machinery is operating simultaneously. Secondly, accurate ultrasound sensors are big in size and cannot be feasible in a working environment, i.e., their application is limited to labs. To overcome these problems, data fusion from various sensors can give full and localized body fatigue.

Challenges are not only possessed during training or developing the model but also comes with real life implementation problems such as; To begin, data privacy and security concerns may arise among employees who are concerned about continual surveillance and control of their personal information. Addressing these issues necessitates the development of strict data protection standards, the encryption of sensitive information, and open communication with employees about data usage and safeguards. Second, guaranteeing user acceptance and comfort with wearable sensors presents a difficulty, as discomfort from wearing sensors for long periods of time, as well as skepticism or aversion to new technology, might impede adoption. Solutions include creating lightweight, inconspicuous wearable sensors and integrating workers in the design and implementation process to promote comfort and buy-in. Furthermore, individual diversity in physiological and behavioral reactions to fatigue provides a hurdle, making it impossible to develop a one-size-fits-all detection model. This difficulty can be addressed by personalizing fatigue detection models for people or using adaptive machine learning algorithms capable of learning and adapting to each worker’s patterns over time. Furthermore, ensuring system durability and reliability to reliably identify fatigue under a variety of situations and factory settings is critical. Diversifying training data and environments to create more resilient models, as well as building methods for continuous model evaluation and recalibration, are critical steps towards addressing this difficulty. Furthermore, any possible disturbance to existing workflows caused by the integration of new technologies must be addressed, with adaptive algorithms in cobots allowing for real-time response to identified weariness without requiring disruptive adaptation periods.

5.1.3. Working with data

The current data sets from the reviewed papers do not represent working populations in an industrial environment. The data collected are generally from students and managed in a lab environment. However, when data is contained in a realistic industrial environment, it affects the information differently compared to the lab collection. These limitations in the current datasets need to provide conclusive evidence about the working in the actual working environment and the sensors or the data collection repeatability. To the authors’ knowledge, in the current review, studies have examined less the application of physiological sensors in measuring fatigue in a substantial sample of manufacturing operators over an extended length of time (e.g., a few days or weeks). The only evidence found is in terms of interviews/surveys which provide intermittent results, which are not good for real-time assessment [19]. Furthermore, research comparing the use of physiological measurements to the industry standard of physical fatigue assessment or setting an industry standard is less explored.

ML algorithms play a crucial role in the classification of detection and monitoring of human physical fatigue, overcoming the limitations stated above. Thus, the amount of data used for training the ML algorithm affects how well it will work. For example, to detect the development of fatigue, statistical modeling is the prevailing algorithm to classify the levels of fatigue. It is capable of learning useful information from sparse datasets. It produces consistent results from the signals generated by examining useful information from the features extracted. However, the limitation is that it includes limited data. Human fatigue modeling requires enough data that represent different variables (demographic factors, muscle activations, etc.) for training processes to yield high-quality results. However, the current data are limited to the data extracted from the sensors only. Additionally, when using ML for fatigue predictions, different processing parameter combinations should be considered for higher results. Testing these variables (demographic and anthropometric) in a manufacturing setting requires considerable time. Therefore, gathering a lot of training data from experiments is not always feasible. Process simulations possess an alternative solution for training in a created virtual environment but lack repeatability in real-life scenarios. The limited data may lead to a high possibility of false predictions due to the lack of training. It is difficult for a well-trained ML model to be applied in a similar physical exertion process but with different movements where uncertainties affect the model performance [111]. As a result, building a solid and trustworthy ML model from a small set of training data is difficult.
5.1.4. Data fusion

The analysis of fatigue is a complex process, where several physiological, mental, and manufacturing setting factors are affected by various correlated factors. Thus, it makes it necessary and significant to fuse the data from multiple sensors or modalities to jointly analyze for enhancing knowledge discovery. One of the limitations is to fuse the heterogeneous data generated from the sensors as it normally has different types, signals, and features extracted. To tackle this issue several methods are adopted by researchers, such as extracting features from raw data to reduce data dimensions or time-series or frequency analysis techniques. As a result of the feature extraction processes, important underlying or pertinent features from the raw data are lost. Therefore, a crucial problem when using ML is how to combine the diverse data for modeling and analysis when acquired from different sensors. Additionally, much literature does not include the subjective analysis or any demographic variable’s characteristics of the person in the classification. It is one of the contributing factors and limitations in the current data sets for detecting fatigue. Deep learning machine learning modeling method can possess the alternative solution. However, the limited amount of data available in contemporary data sets makes it difficult to use deep learning.

5.2. Opportunities

5.2.1. Opportunities in working with data

Accurately training an ML model requires a large amount of data. The current data sets are arguably small and represent a limited population of working people in an actual industrial environment scenario. They are primarily students employed for data collection. One of the key opportunities is to use an actual working population that represents various demographic variables.

First, data is mainly collected from the age group of 18–45, and these limitations can be overcome by increasing the population size for different age groups. Second, most of the research is undertaken in a laboratory environment, and researchers can aim for a working environment like many examples from the construction industry. Lastly, if the first two points are not feasible, they could use the application of transfer learning. It is difficult for employers to implement practically in real-world industry applications due to its cost. Thus, transferring learning can help provide knowledge from a lab domain to a related domain, which is crucial and will help improve the models and reduce the training time. It is observed from the reviewed papers that more than one sensor is needed to detect whole-body fatigue. Thus, an amalgamation of various sensors would create the opportunity for better monitoring and detection of human fatigue. It could help in a specific application, and a more generic approach is needed. So that it could be repeatable and reliable.

5.2.2. Opportunities in machine learning modeling techniques and implementing in real life

As mentioned previously, increasing the dataset size could help employ deep learning, a data-hungry modeling process. It would further help in including various demographic variables, such as sleep cycle, etc., in one. Employing deep learning can further expand the knowledge about the development of fatigue by providing deeper insights and developing unknown patterns. Furthermore, increasing the chances of a seamless Human–Robot collaboration interaction over different applications. However, only a few pieces of literature are available on fatigue detection, monitoring, and adaptation. Fatigue can hinder one’s working process in many domains. This allows for detecting fatigue other than human operators’ estimated force.

However, in real-world applications, low latency is the main issue in real-time detection, monitoring and recovery. To achieve low latency in real-world applications, it is crucial to deploy analytical models on local devices. It is equally essential to build lightweight computing models for applying artificial intelligence to predictions. One of the promising methods is to improve further the use of data fusion techniques, which can help minimize the dimensionality of raw data. Furthermore, implementing machine learning technology for human physical fatigue monitoring in the workplace, particularly in Human–Robot Collaboration (HRC) contexts, provides several benefits. First, fatigue detection can improve safety and productivity by reducing work-related injuries and accidents. In addition, automatic job adjustment and robot aid can help maintain high levels of production even when fatigue sets in. Moreover, early detection of fatigue helps prevent chronic health problems, improving worker health and well-being. Also, by incorporating machine learning technology into human resource management settings, businesses can increase overall operational efficiency and market competitiveness. The capacity to recognize and mitigate fatigue efficiently can lead to more efficient job allocation and resource utilization, resulting in higher production output and less downtime.

6. Conclusions

The focus of this paper has been on the research and application of various ML technologies for the detection and monitoring of human fatigue in a manufacturing setting with and without an HRC setting. Fatigue has been researched for decades. It is very subjective in nature. Despite many studies, no universal or accepted definition is present. Furthermore, detection of the onset of fatigue is crucial and methods to detect this are needed. If fatigue develops, it can cause attention loss which further hinders not only the quality of the process but also affects the socio-economic status of a person. Run over the longer term, prolonged instances of fatigue can cause musculoskeletal disorders (MSDs).

The prominence of ML for the prediction of human fatigue challenges served as inspiration for the paper. A significant work of current literature advocates the use of physiological sensors to detect and monitor the accumulation of fatigue. The paper followed a systematic literature review process for examining and selecting the articles for review. It has been observed that there are two general methods adopted by researchers to detect fatigue.

The two methods are analytical modeling and signal-processing analysis techniques for physiological sensors. In analytical modeling, the use of three main categories of modeling techniques is used; they are statistical, ensemble and single classifiers are used. Whereas, in signal analysis; the use of time and frequency domain analysis is the most employed. Apart from the mentioned the use of motion tracking is also used. However, these methods have their advantages and disadvantages. Although, there are some limitations in the current detection method, which could be overcome by further research into data fusion by various sensor use, transfer learning and deep learning. Furthermore, the use of various sensor fusions is recommended, and different combinations should be tried to overcome the current challenges.

CRediT authorship contribution statement

Arsalan Lambay: Writing – original draft, Methodology, Formal analysis, Conceptualization. Ying Liu: Writing – review & editing, Supervision, Methodology, Conceptualization. Phillip L. Morgan: Writing – review & editing, Supervision, Methodology. Ze Ji: Writing – review & editing, Supervision.

Declaration of competing interest

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

Data availability

No data was used for the research described in the article.
References


L. Willard, ṁ2. REPORT DATE DATE 12/21/03 4. TITLE AND SUBTITLE AGENCY USE ONLY (Leave blank)


L. Willard, ṁ2. REPORT DATE DATE 12/21/03 4. TITLE AND SUBTITLE AGENCY USE ONLY (Leave blank)


L. Willard, ṁ2. REPORT DATE DATE 12/21/03 4. TITLE AND SUBTITLE AGENCY USE ONLY (Leave blank)


