Wearable Technology and Machine Learning for Assessing Physical Fatigue in Industry 4.0

Human Factors and Wearable Technologies, Vol. 141, 2024, 28–36
https://doi.org/10.54941/ahfe1005052

Wearable Technology and Machine Learning for Assessing Physical Fatigue in Industry 4.0

Carlos Albarrán Morillo[1], Micaela Demichela[1], Devesh Jawla[2], and John D. Kelleher[3]

[1]Department of Applied Science and Technology, Politecnico di Torino, Italy
[2]Technological University Dublin, Ireland
[3]School of Computer Science and Statistics, Trinity College Dublin, Ireland

ABSTRACT

Industry 4.0 is a shift towards automation and data integration in manufacturing and process sectors. However, manual material handling and repetitive operations still cause significant physical strain on operators, leading to fatigue and exhaustion. This fatigue not only hampers performance but also compromises production quality and efficiency, potentially leading to human errors and accidents. Prolonged exposure to physical fatigue can lead to conditions like chronic fatigue syndrome (CFS) and work-related musculoskeletal disorders (WMSDs). Given these implications, safeguarding occupational health and safety necessitates a proactive approach to managing operator physical fatigue. This study uses wearable devices and health information to propose a real-time measurement and monitoring solution for operator physical fatigue in operational environments. The Empatica EmbracePlus smartwatch was used to quantify fatigue during simulated industrial tasks. Participants engaged in repetitive tasks, while the device monitored vital indicators like heart rate, electrodermal activity, and skin temperature. Self-reported fatigue levels were assessed using the Borg scale to provide ground truth labels for the collected data. The acquired dataset served as input for machine learning models to classify physical fatigue into discrete levels, ranging from 2 to 5 distinct categories. The results highlight the efficacy of the XGBoost algorithm in accurately classifying physical fatigue, demonstrating a classification accuracy of 94.1% for five levels and 99.4% for three levels and the pulse rate as the more reliable indicator of fatigue levels. Additionally, a Bayesian Neural Network model, while producing similar results to the XGBoost algorithm, offers the added advantage of providing credible intervals for its predictions. This research lays the foundation for future deployments of the developed human performance model in real-world industrial environments.

Keywords: Machine learning, Human performance modelling, Industry 4.0, Physical fatigue, Physiological parameters, Wearable sensors

INTRODUCTION

Industry 4.0 is a disruptive trend that incorporates sophisticated technologies such as IoT, AI, and robotics into industrial settings to improve automation, analysis, and maintenance efficiency (Javaid et al., 2021; Ghobakhloo, 2020).

Nonetheless, despite the growing use of advanced technology and automation, operators are still critical components of modern industrial systems (Romero et al., 2020). Unfortunately, the industrial setting has historically placed a greater emphasis on process/product quality, thereby overlooking the critical role of human operators (Bondarouk et al., 2020). Empirical research has shed light on the importance of human factors (HF) in manufacturing processes, finding links between quality deficiencies and negative human consequences such as workload-induced weariness and injury-related risk factors (Reiman et al., 2021). Consequently, there is a rising realization of the need to focus and improve human elements in operations systems to boost overall system performance and maintain higher quality standards (Neumann et al., 2021). Workers in a variety of industries typically experience physical weariness because of doing physically demanding and repeated jobs (Albarrán Morillo and Demichela, 2023). Lifting, pushing, tugging, and carrying big things require a lot of effort, which can lead to eventual tiredness. The repetitious nature of these actions causes additional strain on the body, resulting in physical weariness (Albarrán Morillo and Demichela, 2023). Physical weariness considerably raises the likelihood of human mistake and job mishaps (Yeow et al., 2014). Fatigue impairs workers' cognitive capacities and motor capabilities, making them more prone to making mistakes and losing focus (Valentina et al., 2018). This decreased vigilance and coordination can lead to mishaps such as slips, trips, and falls, which are a major safety risk in a variety of work settings. Additionally, physical weariness has a substantial impact on total work performance. Workers that are fatigued are more likely to experience diminished productivity and efficiency (Mahmud Akter and Ahmad, 2011). Physical fatigue's effects might worsen with time, leading to more serious health consequences (Balachander et al., 2014). These effects include chronic fatigue syndrome (CFS), work-related musculoskeletal diseases (WMSD), and a reduction in immune function. CFS is distinguished by persistent and profound fatigue that does not improve with rest and is accompanied by a variety of physical and psychological symptoms (Lee et al., 2023). WMSD, on the other hand, refers to a range of disorders affecting muscles, tendons, and sensitive tissues caused by repetitive or intense physical activity, resulting in pain, limited mobility, and other symptoms (Lee et al., 2023). As a result, monitoring physical exhaustion has become an essential component for early detection (Meeus et al., 2007).

Even though exhaustion is a subjective sense, and the level of fatigue experienced by an individual varies based on factors such as overall health and well-being, job demands, and circumstances, there is currently no perfect way for quantifying physical fatigue. As a result, the most precise approach of assessing physical exhaustion levels is currently being developed. Conducting questionnaire-based interviews is one way for assessing personal physical weariness (Kumar, 2001). This approach is based on subjective evaluations and may be prone to biases caused by an individual's mood or willingness to provide an accurate account of their exhaustion. To improve reliability, objective measures are required. Physical weariness is intimately associated with the sympathetic nervous system (SNS), which can be assessed via physiological signs (Okawa et al., 2019). Physical weariness can be assessed using a variety of physiological measures, including heart activity, blood activity, and skin reaction. Monitoring personal physiological signals with wireless sensors allows for continuous tracking of physical exhaustion level.

In this study, we present a novel model for classifying physical fatigue in different levels, integrating both objective and subjective measures.

The remainder of this work is arranged as follows. The next part describes our human performance modeling technique, the experimental procedure, and the data gathering and elaboration process. The experimental results are reported in the next section. Finally, the essay discusses the findings and conclusions, summarizes the limitations, and suggests future research paths.

HUMAN FATIGUE PROPOSED FRAMEWORK

The proposed methodology evaluates tiredness by simulating industrial duties in a fitness setting (see Figure 1). Participants do physically demanding activities such as pushing, tugging, picking up, bending, and lifting, which are frequent in the factory environment. These activities need repetitive movements and the use of physical force, mimicking real-world labor conditions. The data for tiredness prediction is derived from the Empatica Embrace-Plus bracelet, which offers real-time physiological signals depending on the wearer's physical activity level. The bracelet continually measures three key components: pulse rate (PR), electrodermal activity (EDA), and temperature. During the data collecting procedure in the fitness setup, the participants' exhaustion levels are labelled with the Borg scale or Borg test (Borg, 1982). This method allows individuals to subjectively rate their level of physical exertion on a scale ranging from 6 (no exertion) to 20 (maximum exertion).

For a more in-depth exploration of the proposed framework, readers are encouraged to refer to our previous article (Albarrán Morillo and Demichela, 2023).

Participants and Data Collection

The study included 33 healthy volunteers (21 men and 13 females) ranging in age from 21 to 41 years, with a mean of 25.6 ± 4.4 years. With this sample size, we have nearly reached the desired statistical power of 80% (34 individuals), which is critical for assuring the reliability and validity of the experimental findings (Serdar et al., 2021). People with current or previous injuries, pain, discomfort, medical conditions, or those taking medications were excluded from the initial screening questions (see Figure 1). The study followed the Declaration of Helsinki standards. Prior to data collection, each participant was given an informed consent form that contained extensive information about the study, such as its nature, potential benefits, hazards, and alternatives.

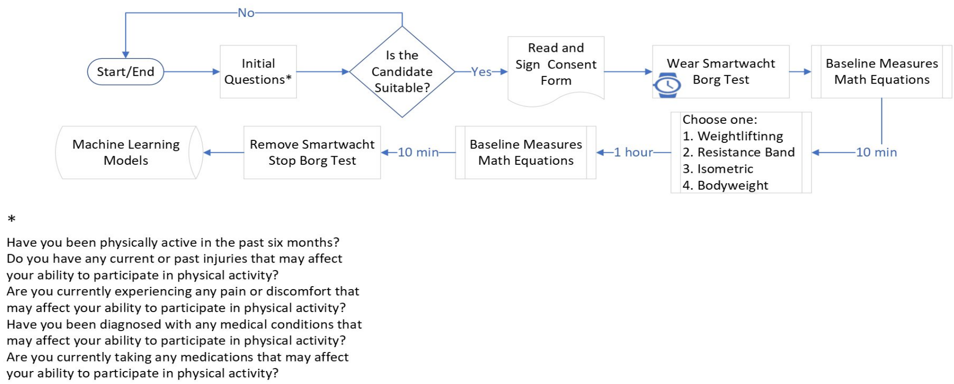

Figure 1: Data collection procedure in the fitness setup.

Physical Fatigue Classification Models

A typical and useful strategy is to select a few algorithms that could be appropriate for the task at hand. Each algorithm is then trained, and the model with the highest predicted accuracy is chosen as the final candidate. According to prior research, ensemble learning approaches such as Random Forest (RF) and the boosting algorithm XGBoost are well-suited to addressing the time-series structure of physiological data, resulting in more accurate and reliable fatigue classification results (Anwer et al., 2022; Bustos et al., 2022). Along with RF, we used J48 and logistic model tree (LMT), which are tree-based algorithms comparable to RF but with different methods and features. Furthermore, our literature analysis revealed that Naive Bayes (NB) is widely employed in tiredness classification (Purnomo et al., 2020). To give a fuller comparison, we incorporated the Tree-Augmented Naive Bayes (TAN) method in our analysis. TAN is a modification of the classic NB method that considers feature dependencies, perhaps making it better at capturing complex correlations in physiological data. Finally, this research introduces a unique aspect: the use of Bayesian Neural Networks (BNNs) trained utilizing Markov Chain Monte Carlo (Papamarkou et al., 2022). This technique gives not only predictions but also the likelihood of correctness, as well as reasonable intervals (confidence scores), allowing a robust and complete study of the parameter landscape during the training process. As a result, this approach can improve human interpretability, making the fatigue categorization process more useful and insightful.

EXPERIMENTAL RESULTS

The datasets were analysed using various machine learning algorithms to classify the data based on different levels of fatigue. The fatigue levels were categorized using the scores from the Borg scale into two levels: low fatigue (6-12) and high fatigue (13-20), three levels: low (6-12), moderate (13-16), and high (17-20), and four levels: low (6-10), moderate (11-14), high (15-17), and very high (18-20). The collected dataset was divided into 70% for training and 30% for testing. A mixed dataset approach was adopted, allowing samples from the same individual to be used for both training and testing. The Weka software was utilized to conduct 10-fold cross-validation using three classification algorithms: J48, LMT, and RF. k-fold cross-validation involves dividing the dataset into k subsets of approximately equal size. The model is then trained on k-1 of these subsets and tested on the remaining subset. This process is repeated k times, each time using a different subset as the test set. For the TAN algorithm, Hugin Expert software was utilized. Finally, the XGBoost algorithm and BBNs were implemented using Julia software. Normalization was used for each feature, ensuring a thorough evaluation of the model's performance across diverse participants.

Table 1 shows the accuracy of the algorithms employing 3 digital biomarkers, EDA, pulse rate and temperature.

Table 1. Algorithms and classification accuracy using PR, EDA and temperature for 2, 3 and 4 physical fatigue levels.

Algorithm	**Classification Accuracy (%)**		
Levels	**2**	**3**	**4**
J48	85.9	81.6	78
LMT	85.7	80.6	76.2
RF	85.5	82.6	82.1
TAN	94.3	79.6	63.5
XGBoost	99.9	99.4	97.4

The classification accuracy for the J48 algorithm across 2, 3, and 4 physical fatigue levels is 85.9%, 81.6%, and 78%, respectively. For the LMT algorithm, the accuracy achieved for 2, 3, and 4 physical fatigue levels is 85.7%, 80.6%, and 76.2%, respectively. The RF algorithm achieved classification accuracies of 85.5%, 82.6%, and 82.1% for 2, 3, and 4 physical fatigue classification levels. The TAN algorithm obtained accuracies of 94.3%, 79.6%, and 63.5% for 2, 3, and 4 levels, respectively. Finally, the XGBoost algorithm performed exceptionally well, achieving classification accuracies of 99.9%, 99.4%, and 97.4% for 2, 3, and 4 fatigue levels, respectively. The XGBoost algorithm demonstrated the highest accuracy among all the tested algorithms, achieving exceptional results in multi-level fatigue classification. By categorizing fatigue into multiple levels, we sought to gain a deeper understanding of individuals' fatigue states during different activities and work tasks. The XGBoost algorithm was chosen as the preferred method for the subsequent data analysis steps.

Additionally, we evaluated whether combining multiple parameters improved the overall classification accuracy (see Table 2). The fatigue classification was conducted with three and five levels, as these score ranges demonstrated statistical significance, while the remaining levels did not offer meaningful distinctions. For the five-level classification, we categorized the Borg test scores as follows: low-fatigued (6-8), low-moderate (9-11), moderate (12-14), high (14-17), and very high (18-20).

Table 2. Classification accuracy (%) using XGBoost and inputs PR, EDA and skin temperature individually and their combinations.

Input	**Classification Accuracy (%)**	
Levels	3	5
EDA	53.5	50.4
Temperature	53.6	49.1
Pulse rate	99	92.1
EDA + Pulse rate	99.1	92.7
EDA + Temperature	53.8	50.2
Temperature + Pulse rate	99.1	91.2
All together	99.9	94.1

The results demonstrate that the combination of three sensors, EDA, pulse rate, and temperature, achieved the best performance with an accuracy of 99.9% for the three levels classification and 94.1% for the five levels classification. When using a combination of two sensors, EDA and pulse rate, the model achieved an accuracy of 92.7% for the five levels classification and 99.1% for the three levels classification. Similarly, combining temperature and pulse rate resulted in an accuracy of 91.2% for the five levels classification and 99.1% for the three levels classification. The using only the pulse rate sensor showed excellent performance, achieving 99% accuracy for the three levels classification and 92.1% for the five levels classification, even outperforming the combination of pulse rate with temperature for the five levels classification (91.2%). The robust performance of the pulse rate sensor, even when used alone, highlights its potential as a reliable indicator of physical fatigue. In contrast, the performance was significantly lower when using only one sensor for EDA or temperature, or the combination of both, achieving less than 54% accuracy for the three levels classification and less than 51% accuracy for the five levels classification.

The authors introduce a novel algorithm using Bayesian Neural Networks (BNNs) trained with Markov Chain Monte Carlo (MCMC). The results show impressive performance, especially for 3-level classification (see Table 3). While the accuracy for the 5-level classification is slightly lower compared to XGBoost, achieving 86.7% accuracy is still noteworthy. The method also offers advantages in human interpretability, such as obtaining credible intervals, which provide a measure of uncertainty for the model's predictions.

Table 3. XGBoost and BBNs classification accuracy.

Input	**Classification accuracy (%)**	
Levels	3	5
XGBoost	99.9	94.1
BBNs	99.9	86.7

DISCUSSION

The study reveals that the XGBoost algorithm has demonstrated exceptional performance in classifying physical fatigue levels, with accuracy rates of 99.9% for three levels and 94.1% for five levels when combined with three physiological parameters: EDA, pulse rate, and skin temperature. This outperforms previous studies, which achieved classification accuracy of 96.5% for four levels and 93.5% for the same classification (Anwer et al., 2022; Bustos et al., 2022; Pinto-Bernal et al., 2021). The XGBoost algorithm has demonstrated superior performance compared to RF and other tree-based approaches in the realm of fatigue classification.

Pulse rate is a key finding in this study, as it is a more reliable indicator of fatigue and less influenced by external factors. It is directly related to heart activity and responsiveness to changes in exertion and stress levels, making it a valuable tool for assessing fatigue. The study also found that using pulse rate alone is sufficient for accurate fatigue prediction, and there is no statistical significance in incorporating additional sensors alongside pulse rate. This has practical implications for various industries, sports training, healthcare, and other fields where fatigue management is essential.

Lastly, to improve human interpretability for future real-world applications where fatigue predictions are crucial, we used a Bayesian Neural Network (BBN) model trained with Markov Chain Monte Carlo (MCMC) (see Table 3). The BBN-MCMC technique provides various benefits for understanding the model's predictions and decision-making process. Unlike some complicated machine learning algorithms, the BBN-MCMC model generates credible intervals that serve as confidence scores for each prediction. With credible intervals, users and stakeholders may have more faith in the model's outputs and make more educated decisions depending on the amount of certainty offered. Furthermore, the BBN-MCMC approach ensures a solution for the model's parameter space. This is especially useful since it guarantees that the model's predictions are not locked in local optima, so helping to a more reliable and robust classification performance.

CONCLUSION AND FUTURE WORK

The proposed physical fatigue monitoring system is a significant advancement in supporting everyday physical fatigue assessment. Utilizing a non-intrusive smartwatch device, the system achieves high classification accuracy of >99% for 3 levels and >86% for 5 levels. This research sets the stage for practical applications in various industries, enabling real-time, non-intrusive fatigue detection for improved health and safety outcomes. A potential use for the physical exhaustion model may be to monitor assembly line workers' physical exertion to forecast and avoid injuries. Testing the physical fatigue detection model on assembly lines is highly relevant due to repetitive actions and heavy lifting, common in sectors like automotive, electronics, and pharmaceuticals. The application of the model involves assessing data and estimating injury risks for each operator. To enhance workplace safety, the model aligns operators with workstations based on their projected injury risk. Those with lower risks are assigned to physically demanding tasks, optimizing their capabilities. Conversely, operators at higher risk are strategically placed in less demanding positions, minimizing the potential for injuries. Future research could incorporate external factors influencing operators' fatigue levels, such as humidity, temperature, noise levels, lighting conditions, ergonomics, shift timing and demographic factors. This could involve evaluating the accuracy of classification, assessing computational capabilities, and analysing the importance (gain) of each feature within the model.

The proposed physical fatigue monitoring system has promising results, but it needs to address several limitations to ensure its practicality and reliability in real-world applications. Data privacy and ethics are crucial in industrial settings, as collecting physiological data from individuals raises concerns about data protection, informed consent, and potential misuse. The study used a specific dataset for training and testing, but the sample size may be limited, affecting the generalizability of the models to broader populations or different environments. Consequently, it is imperative to assess the model's generalizability or consider collecting a more extensive dataset. Obtaining real-time performance in the model is critical for practical applications. Real-time performance is crucial for practical applications, particularly in promptly detecting fatigue. Enhancements in algorithms for faster processing and reduced latency are vital for achieving instantaneous tiredness detection. However, it is worth mentioning that the processing power required for the BBNs method is ten times greater than that of other algorithms, posing a challenge that must be addressed for effective real-time implementation.

Despite these limitations, this study is a significant step forward in the development of physical tiredness monitoring systems.

ACKNOWLEDGMENT

This work is funded by CISC project, which is part of the grant H2020-MSCA-ITN-2020 grant agreement no. 955901.

REFERENCES

Albarrán Morillo C., Demichela M. (2023) "Exploring the Impact of Repetitive Exercise on Physical Fatigue: A Study of Industrial Task Simulation in a Controlled Fitness Setting", Chemical Engineering Transactions, 99, 163–168.

Anwer, S., Li, H., Umer, W., Antwi-Afari, M. F., Mehmood, I., Yu, Y., Haas, C., et al. (2022) "Identification and Classification of Physical Fatigue in Construction Workers Using Linear and Nonlinear Heart Rate Variability Measurements", SSRN Electronic Journal, Elsevier BV.

Balachander S., Rao P., Sarkar S., Singh S. (2014) "Chronic fatigue syndrome: A review", Medical Journal of Dr. D. Y. Patil University, 7(4), pp. 415–422.

Bondarouk, T., Iqbal, Q., Vrchota, J., Cunha, L., Silva, D., & Maggioli, S. (2020) "Exploring the status of the human operator in Industry: A systematic review", Frontiers in Psychology, 13.

Borg, G. (1982). Psychophysical bases of perceived exertion. Medicine and Science in Sports and Exercise, 14(5), 377–381.

Bustos, D., Cardoso, F., Rios, M., Vaz, M., Guedes, J., Torres Costa, J., Santos Baptista, J., et al. (2022) "Machine Learning Approach to Model Physical Fatigue during Incremental Exercise among Firefighters", Sensors, MDPI AG, Vol. 23, No. 1, p. 194.

Kumar, S. (2001). Theories of musculoskeletal injury causation. Ergonomics, 44(1), 17–47.
Ghobakhloo, M. (2020) "Industry 4.0, digitization, and opportunities for sustainability", Journal of Cleaner Production, Elsevier BV, Vol. 252, p. 119869.
Javaid, M., Haleem, A., Singh, R. P. and Suman, R. (2021) "Artificial Intelligence Applications for Industry 4.0: A Literature-Based Study", Journal of Industrial Integration and Management, World Scientific Pub Co Pte Ltd, Vol. 07, No. 01, pp. 83–111.
Lee, Y. C., Hong, X., & Man, S. S. (2023) "Prevalence and Associated Factors of Work-Related Musculoskeletal Disorders Symptoms among Construction Workers: A Cross-Sectional Study in South China", International Journal of Environmental Research and Public Health, 20(5).
Mahmud Akter, M., & Ahmad, N. (2011) "Study the Impact of Fatigue and Optimizing Productivity of an Assembly Line of Garment Industry", Article in International Journal of Scientific and Engineering Research, 2(11).
Meeus, M., Nijs, J. and De Meirleir, K. (2007) "Chronic musculoskeletal pain in patients with the chronic fatigue syndrome: A systematic review", European Journal of Pain, Wiley, Vol. 11, No. 4, pp. 377–386.
Neumann, W. P., Winkelhaus, S., Grosse, E. H., & Glock, C. H. (2021) "Industry 4.0 and the human factor – A systems framework and analysis methodology for successful development", International Journal of Production Economics, 233.
Okawa, N., Kuratsune, D., Koizumi, J., Mizuno, K., Kataoka, Y. and Kuratsune, H. (2019) "Application of autonomic nervous function evaluation to job stress screening", Heliyon, Elsevier BV, Vol. 5, No. 2, p. e01194.
Papamarkou, T., Hinkle, J., Young, M. T. and Womble, D. (2022) "Challenges in Markov Chain Monte Carlo for Bayesian Neural Networks", Statistical Science, Institute of Mathematical Statistics, Vol. 37, No. 3.
Pinto-Bernal, M. J., Cifuentes, C. A., Perdomo, O., Rincón-Roncancio, M. and Múnera, M. (2021) "A Data-Driven Approach to Physical Fatigue Management Using Wearable Sensors to Classify Four Diagnostic Fatigue States", Sensors, MDPI AG, Vol. 21, No. 19, p. 6401.
Purnomo, A., Barata, M. A., Soeleman, M. A. and Alzami, F. (2020) "Adding feature selection on Naïve Bayes to increase accuracy on classification heart attack disease", Journal of Physics: Conference Series, IOP Publishing, Vol. 1511, No. 1, p. 012001.
Reiman, A., Kaivo-oja, J., Parviainen, E., Takala, E.-P. and Lauraeus, T. (2021) "Human factors and ergonomics in manufacturing in the industry 4.0 context – A scoping review", Technology in Society, Elsevier BV, Vol. 65, p. 101572.
Romero, D., Stahre, J. and Taisch, M. (2020) "The Operator 4.0: Towards socially sustainable factories of the future", In Computers and Industrial Engineering, Vol. 139, Elsevier Ltd.
Serdar, C. C., Cihan, M., Yücel, D. and Serdar, M. A. (2021) "Sample size, power and effect size revisited: Simplified and practical approaches in pre-clinical, clinical and laboratory studies", Biochemia Medica, Croatian Society for Medical Biochemistry and Laboratory Medicine, Vol. 31 No. 1, pp. 27–53.
Valentina V., Fabio S., Martina C. and Alessandro P. (2018) "Fatigue accumulation in the assignment of manual material handling activities to operators", IFAC-PapersOnLine, 51(11), pp. 826–831.
Yeow, J. A., Ng, P. K., Tan, K. S., Chin, T. S., & Lim, W. Y. (2014) "Effects of Stress, Repetition, Fatigue and Work Environment on Human Error in Manufacturing Industries", Journal of Applied Sciences, 14(24), 3464–3471.